Composition du jury

Pr. J.F. Gohy (Président du jury)
Université catholique de Louvain

Pr. P. Soumillion (Promoteur)
Université catholique de Louvain

Pr. B. Hallet
Université catholique de Louvain

Pr. P. Morsomme
Université catholique de Louvain

Pr. J.F. Collet
Université catholique de Louvain

Pr. B. Joris
Université de Liège

Remerciements

La thèse, c'est comme un marathon ou un triathlon longue distance ou encore un Ironman !

C'est une aventure de longue haleine dont l'objectif à atteindre est hors de portée visuelle pratiquement jusqu'à la fin du parcours… C'est un défi que l'on se donne à soi-même. Mais c'est surtout une formidable histoire de relations, de rencontres et d'amitié.

Pour pouvoir m'aligner sur la ligne de départ, j'ai eu la chance de rencontrer sur ma route les professeurs Jacques Fastrez et Patrice Soumillion qui m'ont accordé leur confiance et m'ont accueillie dans leur laboratoire. J'aimerais leur dire à quel point j'ai apprécié découvrir le monde de la recherche à leurs côtés. Je remercie particulièrement Patrice de m'avoir proposé ce projet passionnant et d'avoir consacré de nombreuses heures à diriger ma recherche avec un optimisme sans faille.

Pendant la course, les chemins empruntés nous ont menés à collaborer avec d'autres laboratoires et équipes de chercheurs. J'aimerais remercier, tout d'abord, le professeur Bernard Hallet d'avoir suivi ce travail en partageant ses connaissances et d'avoir mis à notre disposition de nombreux outils de biologie moléculaire. Merci à André Tonon (Institut Ludwig, UCL) pour les expériences de cytométrie en flux : j'ai pu compter sur son expertise, sa patience et sa bonne humeur lors des longues après-midis de tri des bactéries. Merci à Hervé Degand et au professeur Pierre Morsomme pour la qualité et la rapidité des analyses de spectrométrie de masse réalisées lorsque les résultats étaient pressants. Merci à l'équipe des PDR et plus particulièrement à Amandine Baijot et Joseph Nader, pour l'ambiance accueillante et pour le temps qu'ils m'ont consacrés lors de mon court séjour dans leur laboratoire. Finalement, merci aux membres du jury qui ont accepté d'évaluer ce travail et qui m'ont permis d'apporter des améliorations à ce manuscrit grâce à leurs remarques et questions.

Durant tout le parcours a régné, dans l'aile B du Lavoisier d'abord et ensuite dans la tour C4 du Carnoy, une ambiance chaleureuse, où il est agréable de travailler. Merci à vous les collègues… Avec le temps, certains de vous sont devenus des amis. J'aimerais vous dire combien j'ai apprécié votre présence à mes côtés tout au long de ces journées. Merci à Véro, toujours à l'écoute et disponible pour les démarches administratives.

Merci à Laurie, miss boulette, qui a toujours une anecdote de film ou autre bien marrante à raconter. Merci à Vanina qui m'a guidée avec gentillesse et efficacité au début de mon parcours. Merci, Julie, pour ton écoute et tes bons conseils. Merci à JM, le fantôme qui hantait par sa bonne humeur les couloirs du labo. Merci pour l'aide apportée dans nos projets qui s'entrecroisaient. Merci à Valou, ex-kokoteur et compagnon d'année, pour toutes les activités où tu as finalement décidé de venir. Merci à Gabi, miss-people, pour tous les moments de complicité qu'on a partagés, pour nos conversations de filles dans un bureau avec une majorité de gars mais aussi pour nos discussions scientifiques enrichissantes. Egalement, je te remercie d'avoir si bien relu ma thèse. Merci à Brouno, qui après l'Ironman me trouve un nouveau défi à relever comme l'UTMB. Tu viens avec moi ? Merci à Heykel, Superman, qui m'a sauvée d'une mauvaise aventure. Merci à Anast, Mary et Stéph qui viennent d'arriver au labo et qui relèvent la balance vers une majorité de filles. Merci pour la surprise organisée après ma pré-défense, ça m'a fait chaud au cœur.

Pendant toute la durée de la course, des supporters attentifs se sont placés en nombre tout le long du parcours et leurs encouragements n'ont jamais cessé. Je pense à mes ami(e)s et plus particulièrement à Adrienne, Anne et Elodie. Merci pour notre complicité dans les entraînements sportifs programmés un peu trop longs, pour nos soirées cinéma, nos longues conversations téléphoniques…Je pense également à mes parents qui m'ont soutenue et encouragée depuis le début de mes études et merci à papa qui a même eu le courage de relire ma thèse. Après presque 9 ans, vous allez pouvoir vous reposer ou passer à d'autres activités pour soutenir votre petite fille. Bien sûr, je pense à mes frères, sœur, beaux-frères et belles-sœurs et en particulier à Stéphanie et Vincent. Merci pour tous les messages, les surprises et les soirées passées en votre compagnie. Des petits supporters bien motivés sont arrivés en cours de route, Tom, Aurian et Nathan. Merci à eux pour les moments magiques de tendresse et de fous-rires. Et finalement, le plus important, Maxence, je ne vois qu'un seul moyen de te remercier pour tout ce que je te dois : t'aimer encore et toujours.

Sans le soutien de vous tous, je n'aurais pas pu franchir le ruban de l'arrivée,

A toutes et à tous, merci.

Résumé

Mots clés : peptides cycliques, expression périplasmique, sélection "anti-darwinienne".

Avec l'augmentation de la résistance aux antibiotiques, il existe une pression croissante visant à développer des nouvelles classes de composés antimicrobiens pour lutter contre les maladies infectieuses. Les peptides représentent une classe intéressante d'agents thérapeutiques grâce à leur haute activité et spécificité. Cependant, une des principales problématiques à leur utilisation est leur instabilité vis-à-vis des protéases. Des méthodes traditionnelles existent pour augmenter leur résistance qui sont, par exemple, basées sur la modification des extrémités N- et C-terminales par acétylation et amidation respectivement, et /ou le remplacement d'acides aminés par des résidus non naturels.

Les peptides cycliques sont quant à eux intrinsèquement résistants au clivage protéolytique et sont donc d'un intérêt tout particulier pour la recherche de molécules thérapeutiques. Le potentiel de ces composés a été rapidement compris et des stratégies pour leur synthèse ont été développées. Une approche élégante utilise la technologie SICLOPPS (Split Intein mediated Circular Ligation Of Proteins and PeptideS)) et a permis de générer *in vivo* des collections de peptides cycliques qui atteignent les 10^8 composés différents. La taille de ces collections est plusieurs ordres de grandeur supérieure à celle obtenue par la synthèse chimique. De plus, comme les peptides sont biosynthétisés et génétiquement encodés, des stratégies de criblage ou de sélection *in vivo* peuvent être envisagées pour identifier rapidement des composés actifs.

Ce travail de thèse a été consacré à la recherche de peptides cycliques ayant une activité antibiotique causant la mort de leur bactérie hôte. Pour atteindre cet objectif, nous avons, dans un premier temps, créé et évalué une stratégie permettant une répression totale de l'expression des peptides cycliques. En effet, sans ce système de répression, on risquerait de perdre rapidement les peptides les plus intéressants par sélection *in vivo*. La fonctionnalité de la stratégie, basée sur une reconstitution de gène par double recombinaison homologue et site spécifique, a été prouvée. Cependant, sa faible efficacité ne nous a pas permis de l'utiliser pour la régulation de l'expression des banques de peptides cycliques.

Une application intéressante pourrait tout de même être envisagée pour l'assemblage d'opérons de grande taille *in vivo*.

Ensuite, nous avons créé une méthode permettant de produire des collections de peptides cycliques dans le périplasme d'*E. coli* qui seront combinées aux banques cytoplasmiques déjà créées pour augmenter les chances d'identifier des peptides ayant une activité antibiotique. Cette méthode combine la technologie SICLOPPS avec un transport co-traductionnel et a permis de générer des collections d'hexa-, hepta-, et octa-peptides cycliques avec une diversité totale d'environ 17 millions de composés. Les analyses au spectromètre de masse des extraits périplasmiques ont permis de détecter les peptides cycliques avec un rapport signal/bruit relativement haut. L'efficacité d'épissage apparait comme importante bien que nous ayons identifié des précurseurs inactivés par glutathionylation de la cystéine catalytique, une modification post-traductionnelle qui n'avait jamais été détectée auparavant dans le périplasme.

Finalement, nous avons réalisé les premières mises au point d'une stratégie de sélection génétique "anti-darwinienne" qui devrait permettre d'isoler des peptides ayant une activité bactéricide. La stratégie est d'induire l'expression des peptides dans les cellules et d'ensuite récupérer par FACS les bactéries qui "se suicident" à cause de l'effet antibiotique de leur peptide. Nous avons démontré la faisabilité du tri des bactéries mortes et de la récupération de leur ADN plasmidique. Cependant, de nombreuses étapes de la stratégie doivent encore être optimisées. Deux clones intéressant ont été identifiés après deux tours de sélection sur des banques de peptides cycliques cytoplasmiques et des repiquages. Bien que l'expression des peptides ait une influence négative sur la croissance des bactéries, l'effet bactéricide n'a pas été observé.

Abréviations et symboles

°C	Degré Celsius
A	Ampère
aa	Acides aminé
ADN	Acide désoxyribonucléique
amp	Ampicilline
ara	Arabinose
ARNr	Acide ribonucléique ribosomique
ARNt	Acide ribonucléique de transfert
Bla	β-lactamase TEM-1
BSA	Albumine de sérum bovin
CFU	Colony forming unit
cm	Chloramphénicol
Da	Dalton
$DnaBI_C$	Fragment carboxy-terminal de Ssp DnaB
$DnaBI_N$	Fragment amino-terminal de Ssp DnaB
DO	Densité Optique
DTT	Dithiothréitol
E. coli	Escherichia coli
EDTA	Acide éthylène diamine tétra-acétique
EPL	Expressed Protein Ligation
ery	Erythromycine
FACS	Fluorescence Activated Cell Sorter
glu	Glucose
Int_C	Fragment carboxy-terminal de l'intéine
Int_N	Fragment amino-terminal de l'intéine
IP	Iodure de Propidium
IPTG	Isopropyl- β-thiogalactoside
IRS	Séquence de reconnaissance interne
kan	Kanamycine
kDa	Kilo Dalton
MALDI-TOF	Matrix-assisted laser desorption ionization-time of flight
MIC	Minimum Inhibitory Concentration
min	Minute
NBF	Nucleotide-Binding Fold
ORF	Open Reading Frame
PAGE	PolyAcrylamide Gel Electrophoresis
pb	Paire de base
PBS	Phosphate Buffered Saline
PEG	PolyEthylène Glycol

PCR	Polymerase Chain Reaction
pH	Potentiel hydrogène
RH	Recombinaison Homologue
rpm	Rotations par minute
RSS	Recombinaison Site Spécifique
RT	Room Temperature
s	Seconde
SICLOPPS	Split Intein-mediated Circular Ligation of Peptides and ProteinS
SDS	Sodium Dodecyl Sulfate
SRP	Signal Recognition Particule
Ssp DnaB	Intéine du gène *dna*B de *Synechocystis* sp. PCC6803
TAE	Tris Acetate EDTA
TBS	Tris Buffered Saline
Tet	Tétracycline

Table des matières

I. INTRODUCTION ..... - 1 -

INTRODUCTION GÉNÉRALE ET OBJECTIFS DU TRAVAIL ..... - 3 -

LE CONTRÔLE DE L'EXPRESSION DES GÈNES TOXIQUES DANS *ESCHERICHIA COLI* ..... - 5 -

Introduction ..... *- 5 -*

Systèmes basés sur le vecteur d'expression pET ..... *- 6 -*

Systèmes basés sur le vecteur d'expression pBAD ..... *- 8 -*

La stratégie de l'ARN anti sens ..... *- 10 -*

L'utilisation de promoteurs induits par inversion ..... *- 11 -*

L'utilisation de promoteurs compétitifs ..... *- 11 -*

La stratégie du promoteur hybride ..... *- 12 -*

LA RECOMBINAISON GÉNÉTIQUE ..... - 13 -

Introduction ..... *- 13 -*

La recombinaison homologue ..... *- 13 -*

Mécanisme de la recombinaison homologue RecA dépendante ..... - 13 -

Recombinaison homologue entre deux plasmides ..... - 15 -

Recombinaison homologue impliquant de l'ADN phagique ..... - 16 -

La recombinaison site spécifique ..... *- 18 -*

Le processus de la recombinaison site spécifique ..... - 19 -

La recombinaison site spécifique catalysée par le recombinase à tyrosine TnpI ..... - 20 -

LES STRATÉGIES DE BIOSYNTHÈSE DE PEPTIDES CYCLIQUES ..... - 23 -

Introduction ..... *- 23 -*

La voie de biosynthèse non ribosomique ..... *- 23 -*

La voie de biosynthèse ribosomique ..... *- 26 -*

La biosynthèse de peptides cycliques via les intéines ..... *- 27 -*

Introduction sur les intéines ..... - 27 -

Nomenclature ..... - 28 -

Découverte et distribution ..... - 28 -

Structure ..... - 29 -

Mécanisme d'épissage ..... - 31 -

Applications en ingénierie protéique ..... - 33 -

La technologie de l'autoclivage : la purification protéique ..... - 33 -

La technologie de l'Expressed Protein Ligation (EPL) ..... - 34 -

La technologie de l'épissage en trans ..... - 37 -

La technologie SICLOPPS ..... - 38 -

LES VOIES DE TRANSPORT DES PROTÉINES DANS LE PÉRIPLASME D'*E. COLI* ..... - 41 -
Introduction ..... *- 41 -*
La machinerie de translocation Sec ..... *- 42 -*
Modèle co-traductionnel ..... - 43 -
Modèle post-traductionnel ..... - 44 -
Sélection du modèle de translocation ..... - 46 -
La machinerie de translocation Tat ..... *- 46 -*
Comparaison des séquences-signal N-terminales ..... *- 47 -*

II. RÉSULTATS ET DISCUSSION ..... - 49 -

RÉGULATION DE L'EXPRESSION BASALE DES GÈNES TOXIQUES ..... - 51 -
Introduction ..... *- 51 -*
Objectif ..... *- 53 -*
Résultats ..... *- 54 -*
Création du système de reconstitution de gène par recombinaison double ..... - 54 -
Stratégie pour la construction du plasmide p15I$_C$PX ..... - 56 -
Stratégie pour la construction du phagemide pColYI$_N$ ..... - 58 -
Vérification de la sensibilité au chloramphénicol des cellules TG1- pColYI$_N$ .. - 60 -
Modification des vecteurs pour déplacer la séquence d'homologie au niveau du gène de résistance au chloramphénicol ..... - 60 -
Evaluation du taux de recombinaison ..... - 63 -
Discussion ..... *- 65 -*
BIOSYNTHÈSE DE COLLECTIONS DE PEPTIDES CYCLIQUES DANS LE PÉRIPLASME D'*ESCHERICHIA COLI* ..... - 71 -
Introduction ..... *- 71 -*
Objectif ..... *- 71 -*
Résultats ..... *- 72 -*
Production d'une β-lactamase cyclique dans le périplasme ..... - 72 -
Stratégie de construction ..... - 74 -
Caractérisation des constructions pCBla et pPBla ..... - 76 -
Production de collections d'hexa-, hepta- et octa-peptides cycliques dans le périplasme ..... - 79 -
Stratégie de construction des collections ..... - 79 -
Caractérisation des collections avec le peptide signal de Bla ..... - 79 -
Stratégie de construction des collections avec le peptide signal de DsbA ..... - 83 -
Caractérisation des collections avec le peptide signal de DsbA ..... - 83 -
Discussion ..... *- 94 -*
MISE AU POINT D'UNE APPROCHE DE SÉLECTION ANTI-DARWINIENNE POUR IDENTIFIER DES PEPTIDES TOXIQUES POTENTIELLEMENT ANTIBIOTIQUES ..... - 101 -
Introduction ..... *- 101 -*

Objectifs ..... *- 101 -*
Résultats ..... *- 103 -*
Mise au point de la stratégie sur un système modèle ..... - 103 -
Test de l'effet de l'antibiotique ..... - 103 -
Test du marquage à l'IP ..... - 104 -
Test d'analyses au FACS ..... - 105 -
Test de tri au FACS ..... - 107 -
Test de mise au point de la récupération des plasmides après le tri ..... - 110 -
Application de la stratégie de sélection sur 3 collections de peptides cycliques avec expression cytoplasmique ..... - 111 -
Discussion ..... *- 116 -*

III. CONCLUSION ET PERSPECTIVES ..... - 119 -

IV. MATÉRIEL ET MÉTHODES ..... - 127 -

SOUCHES BACTÉRIENNES ET MILIEUX DE CULTURE ..... - 129 -
Génotypes des souches d'E. coli ..... *- 129 -*
Milieux de culture ..... *- 129 -*
Préparation de cellules électrocompétentes ..... *- 130 -*
Transformation des cellules par électroporation ..... *- 131 -*
LES VECTEURS DE CLONAGE ..... - 132 -
Les plasmides : pGIV012, pUC18Ery, pAP02 et pBADTet ..... *- 132 -*
Le phagemide : pHDiexKanOmpt ..... *- 133 -*
CULTURE, PURIFICATION ET TITRAGE DES PHAGEMIDES ..... - 134 -
Encapsidation et culture des phagemides ..... *- 134 -*
Purification des phagemides ..... *- 134 -*
Titrage des phagemides ..... *- 135 -*
TECHNIQUES DE MANIPULATION DE L'ADN ..... - 136 -
Préparation d'ADN plasmidique d'E. coli ..... *- 136 -*
Electrophorèse sur gel d'agarose ..... *- 136 -*
Extraction d'un fragment d'ADN sur gel d'agarose ..... *- 137 -*
Séquencage ..... *- 137 -*
PCR (Polymerase Chain Reaction) ..... *- 137 -*
Restriction ..... *- 138 -*
Déphosphorylation ..... *- 139 -*
Polymérisation ..... *- 139 -*
Ligation ..... *- 140 -*
Purification d'ADN en solution ..... *- 140 -*
TA-cloning ..... *- 140 -*
Mutagenèse dirigée ..... *- 141 -*

ANALYSE DES PROTÉINES ..... - 142 -
Préparation d'extraits périplasmiques ..... *- 142 -*
Préparation d'extraits cytoplasmiques ..... *- 142 -*
Préparation des fractions membranaires ..... *- 143 -*
Dosage de protéines (méthode de Bradford) ..... *- 143 -*
Analyses SDS-PAGE ..... *- 144 -*
Western Blot ..... *- 146 -*
Analyses par spectrométrie de masse (MALDI-TOF-TOF) ..... *- 148 -*
ANALYSE DES BACTÉRIES MARQUÉES AU MICROSCOPE À ÉPIFLUORESENCE. ..... - 149 -
CYTOMÉTRIE EN FLUX : FACS ..... - 151 -

V. RÉFÉRENCES BIBLIOGRAPHIQUES ..... - 155 -

VI. ANNEXE ..... - 177 -

I. Introduction

Introduction générale et objectifs du travail

Un des challenges actuels des scientifiques est de découvrir de nouvelles technologies performantes afin de faire évoluer des petites molécules en composés biologiquement actifs avec un potentiel thérapeutique[1]. La révolution de la biologie moléculaire et les avancées techniques ont permis de relancer un intérêt pour l'étude et l'exploitation des peptides. Ils se révèlent être des médicaments de choix grâce à leur haute activité et spécificité qui éliminent les risques d'action sur des structures moléculaires autres que la cible prévue, à leur faible accumulation dans les tissus et à leur toxicité souvent limitée[2,3]. Ils ont cependant quelques inconvénients, à savoir leur faible stabilité, leur difficulté de délivrance et de perméation à travers les membranes biologiques, leur synthèse chimique coûteuse et difficile, et leur élimination rapide par l'organisme[4].

Pour améliorer les propriétés biochimiques et biophysiques des peptides, la cyclisation est couramment employée en chimie médicinale[5-8]. En particulier, la cyclisation entre les extrémités N- et C-terminales qui permet de rigidifier la structure et d'augmenter la stabilité. Tout d'abord, des peptides cycliques ont été créés par des méthodes de synthèse chimique[9-12]. Ensuite, grâce aux progrès dans le domaine de la biologie moléculaire et l'ingénierie protéique, la biosynthèse de peptides cycliques naturel "non naturel" a été rendue possible[13]. Deux stratégies principales ont été décrites : l'une est basée sur l'ingénierie génétique des synthases de polypeptides non ribosomiaux (NRPS)[14,15], l'autre sur l'expression de polypeptides recombinants fusionnés à une intéine modifiée (technologie SICLOPPS, Split Intein mediated Circular Ligation Of Proteins and PeptideS)[16-19]. La biosynthèse offre des avantages par rapport aux méthodes purement synthétiques. En effet, en utilisant les outils de la biologie moléculaire, de grandes collections combinatoires de peptides cycliques peuvent être potentiellement générées et criblées *in vivo*.

Dans le cadre de la thèse, la possibilité de réaliser la cyclisation polypeptidique catalysée par une intéine dans le périplasme d'*Escherichia coli* a été évaluée dans le but d'étendre les applications de la technologie SICLOPPS. Des collections de peptides cycliques cytoplasmiques (travail de thèse de J-M Léonard) et périplasmiques ont été créées et combinées à une stratégie de sélection génétique "anti-darwinienne" afin d'isoler des composés ayant une activité antibiotique provoquant la mort de leur bactérie hôte. Dans le contexte de sélection "anti-

darwinienne", l'expression des peptides doit être hautement régulée pour permettre l'identification des substances les plus toxiques en évitant la mort bactérienne précoce. Un système d'expression totalement réprimable basé sur une double recombinaison homologue et site spécifique a donc été mis en œuvre et évalué.

Dans cette introduction, les deux premiers chapitres seront consacrés à la bibliographie sur les méthodes permettant l'expression des gènes toxiques dans *E. coli* et sur la recombinaison génétique en général. Le chapitre suivant présentera les stratégies de biosynthèse de peptides cycliques et en particulier la technologie basée sur les intéines. Finalement, dans le dernier chapitre, seront décrites les différentes voies permettant le transport des polypeptides dans le périplasme d'*E. coli*.

Le contrôle de l'expression des gènes toxiques dans *Escherichia coli*

Introduction

Escherichia coli est l'un des hôtes le plus utilisé pour la production de protéines recombinantes[20-22]. Son succès s'explique par le fait que c'est une des bactéries qui a été la plus étudiée et pour laquelle le plus grand nombre d'outils biotechnologiques ont été développés (mutants, virus, plasmides, …). De plus, cette bactérie présente des propriétés intéressantes comme une haute efficacité de transformation et une culture simple, rapide et efficace. Malheureusement, un grand nombre de séquences clonées (provenant d'organismes variés) interfèrent sévèrement avec la survie des cellules. Cela résulte en des rendements d'expression fortement diminués et parfois même abolis. En fait, certains gènes sont tellement toxiques qu'*E. coli* n'est pas capable de maintenir leur vecteur d'expression durant la phase de croissance car une expression basale à partir d'un promoteur même fortement réprimé suffit à tuer la cellule. Par conséquent, ces gènes toxiques nécessitent un contrôle total de leur expression basale pour permettre leur expression avec une bonne efficacité.

La quantité de transcription basale varie avec les différents promoteurs d'*E. coli* et est due essentiellement à la combinaison de différents facteurs[23] : la stabilité et la cinétique de formation du complexe répresseur-opérateur[24], la position de l'opérateur dans le promoteur, la cinétique de formation du complexe ARN polymérase-promoteur[25], le nombre de copies du plasmide qui détermine le nombre d'opérateurs à contrôler, la présence d'un autre promoteur sur le plasmide à partir duquel la transcription du gène pourrait démarrer[26,27]. Différents systèmes ont été développés afin de permettre à *E. coli* de tolérer le gène toxique et d'avoir une croissance normale jusqu'au moment de l'induction du gène. De plus, après l'induction, ces systèmes ont une expression efficace qui va assurer la production rapide et quantitative de la protéine hautement toxique avant la mort cellulaire. Au niveau de l'efficacité de la répression, il est difficile d'évaluer et de comparer les systèmes développés pour deux raisons. Premièrement, la mesure d'un taux basal d'expression est très difficile si celui-ci est très faible. Deuxièmement, la toxicité est une notion relative. Le nombre de molécules nécessaires pour provoquer la mort cellulaire est variable d'une protéine toxique à l'autre. Pour pouvoir obtenir des résultats comparables, le

niveau basal de chaque système devrait donc être évalué en utilisant une même protéine.

Dans la suite de ce chapitre seront décrits six systèmes qui permettent de diminuer l'expression basale dans *E. coli* en contrôlant étroitement la transcription et la traduction des gènes pour diminuer la quantité d'ARN messagers durant la phase de croissance. Le plus souvent, il s'agit de systèmes artificiels créés en utilisant un promoteur répressible naturel soumis à un niveau additionnel de contrôle. Il est important de remarquer que ces systèmes ont été très souvent combinés pour permettre l'expression quantitative de gènes hautement toxiques.

Systèmes basés sur le vecteur d'expression pET

Un des systèmes les plus puissants développés pour le clonage et l'expression des protéines recombinantes dans *E. coli* est le vecteur d'expression pET commercialisé par Novagen (Figure 1). La particularité de ce vecteur est de placer les gènes cibles sous le contrôle de signaux de transcription et traduction forts du bactériophage T7[28]. Puisque l'ARN polymérase d'*E. coli* ne reconnait pas le promoteur T7, une source d'ARN polymérase T7 doit être introduite dans la souche pour permettre la transcription des gènes cibles. Des souches hôtes lysogénéisées par le phage λDE3 encodant la polymérase sont disponibles. Dans ces souches, le gène de la polymérase est lui-même sous le contrôle du promoteur L8-UV5 *lac*, un dérivé du promoteur *lac*[29,30]. En théorie, l'expression du gène d'intérêt ne devrait se produire qu'après induction avec l'IPTG. Cependant, le contrôle n'est pas absolu et il existe une faible activité de base qui permet un certain niveau de transcription dans l'état de non induction. Ce très faible niveau d'ARN polymérase T7 présent dans les cellules non induites peut empêcher l'établissement de plasmides portant des gènes cibles toxiques ou rendre le plasmide instable.

Trois stratégies peuvent être utilisées pour diminuer ce niveau d'expression basal. La première méthode est d'incorporer du glucose 1% dans la culture. En effet, en présence de glucose, la concentration d'AMP_C bactérien chute, produisant une désactivation de la protéine activatrice du catabolisme (CAP) et une inhibition de l'expression de l'opéron *lac*. La diminution en AMP_C peut être due à l'effet du système de la phosphotransférase des sucres (PTS) sur l'activité

de l'adénylate cyclase, l'enzyme de la synthèse de l'AMP_C. L'enzyme III du PTS donne un phosphate au glucose au cours de son transport. Cependant, la forme phosphorylée de l'enzyme III active également l'adénylate cyclase. Si le transfert du glucose par le PTS est rapide, la quantité d'enzyme III phosphorylée est faible et l'adénylate cyclase est moins active, ce qui diminue la concentration en AMP_C[29]. La seconde stratégie est d'introduire dans la souche les plasmides pLysS et pLysE qui produisent de manière constitutive le lysozyme T7, un inhibiteur naturel de l'ARN polymérase T7[31,32] (Figure 1). Le niveau basal d'expression en présence de pLysS est presque 10 fois plus faible mais un bruit de fond résiduel est toujours présent. La dernière stratégie est de modifier le promoteur T7 sur le plasmide pET en ajoutant l'opérateur *lac* pour créer un promoteur hybride T7/lac (Figure 1). Le répresseur *lacI*, qui réprime le promoteur *Lac*UV5, va pouvoir également se lier sur le promoteur T7/lac pour réduire efficacement la transcription par l'ARN polymérase T7. Comme LacI est un gène faiblement exprimé, une répression 10 fois supérieure peut être obtenue en utilisant un promoteur muté qui surexprime lacI[33].

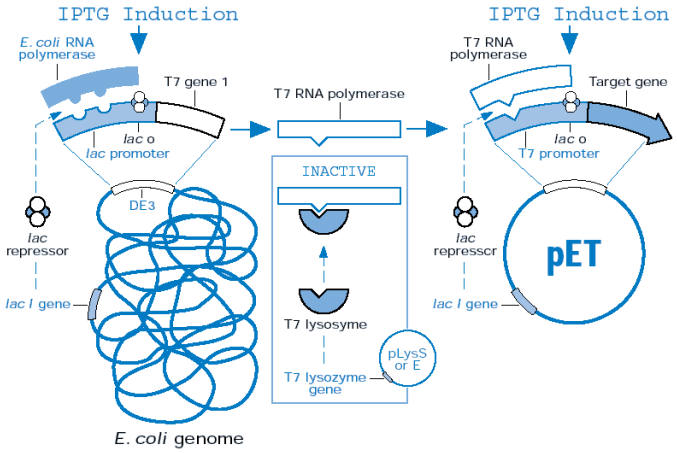


Figure 1 : Les éléments de contrôle du système d'expression pET (repris du site de Novagen).

Malgré le développement de ces stratégies, les plasmides portant des gènes hautement toxiques ne peuvent être maintenus dans ces souches hôtes. Dans certains cas, et plus particulièrement pour la DNaseI pancréatique bovine, la transformation dans une souche d'*E. coli* ne contenant pas le gène codant pour l'ARN polymérase T7 a permis de contourner le problème de la toxicité[34]. L'expression du gène hautement toxique est induite lorsque la culture à atteint la phase exponentielle de croissance en fournissant la polymérase T7 à travers

l'infection par un phage λ contenant le gène codant pour l'ARN polymérase T7. Quoi qu'il en soit, malgré les nombreux moyens de contrôle de l'expression du gène cible dans le vecteur pET, une répression totale ne peut être obtenue. Un exemple d'échec de clonage dans le vecteur pET dû à une faible expression de protéine toxique est rapporté dans la littérature[35]. Il s'agit du clonage du gène sauvage ndd, encodant une protéine toxique Ndd provoquant la désorganisation de l'architecture du nucléoïde. Dans ce cas, aucun clone n'a pu être obtenu, même dans un hôte ne portant pas le gène de l'ARN polymérase T7. Le problème du clonage de ndd dans *E. coli* a été résolu grâce à la découverte d'un allèle thermosensible ndd2[35,36].

Systèmes basés sur le vecteur d'expression pBAD

Les plasmides d'expression basés sur le promoteur P_{BAD} sont commercialisés par Invitrogen et sont appelés vecteur pBAD (Figure 2). Ces vecteurs possèdent, en plus de la région promotrice $P_{BAD,}$ une origine de réplication ColE1, un gène conférant la résistance à l'ampicilline, une région codant pour la protéine araC. En clonant le gène d'intérêt en aval du promoteur P_{BAD}, un contrôle étroit de l'expression basale et un haut niveau de production de la protéine cible en présence de L-arabinose peut être obtenu.

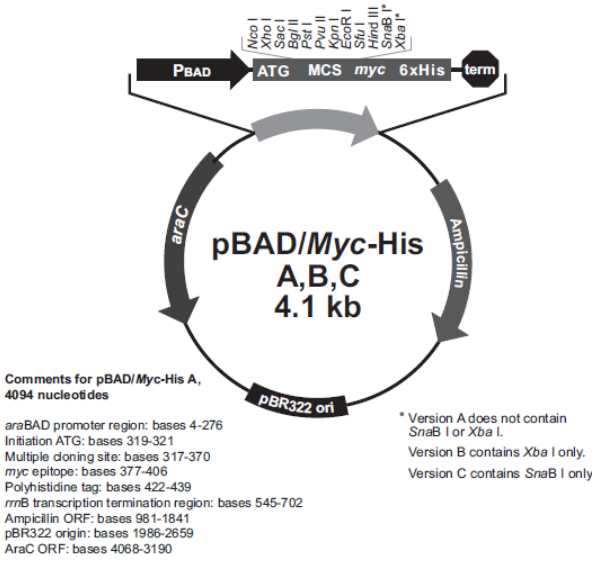


Figure 2 : Représentation schématique des caractéristiques du vecteur pBAD/Myc-His (repris du site de Invitrogen).

L'expression du gène est contrôlée par la protéine AraC (Figure 3). Elle joue le rôle de régulateur de transcription positif et négatif dépendant de la présence ou non d'arabinose[37-39]. En absence d'arabinose, la protéine AraC, sous une forme dimérique, met en contact le site O_2 et la moitié du site I_1du promoteur, formant une boucle dans l'ADN de 210 pb et empêchant la transcription. Une fois ajouté dans le milieu, l'arabinose se lie à la protéine AraC provoquant un changement de conformation qui libère lu site O_2 de la liaison au site I_2. La boucle d'ADN est alors relâchée et la transcription peut commencer. Pour augmenter l'activité de transcription, l'AMPc s'associe à la protéine activatrice CAP pour former un complexe CAP-AMPc se liant à l'ADN et stimulant la liaison de AraC aux sites I_1 et I_2. Le niveau d'expression basal est faible dans ce système mais peut être réduit à un niveau très bas en présence de glucose[20,40]. Le glucose agit en effet en diminuant les niveaux d'AMPc, qui à son tour diminue la liaison de la protéine activatrice CAP à l'ADN.

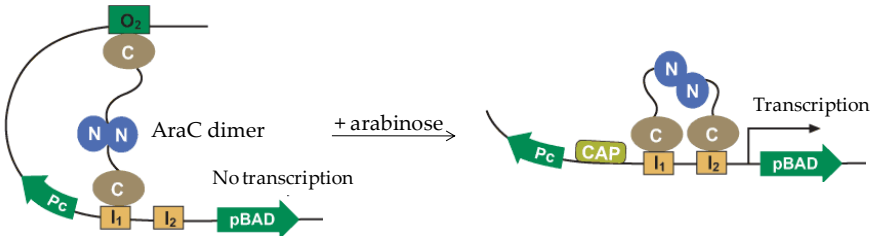


Figure 3 : Schéma de la régulation du promoteur P_{BAD} (repris du site Invitrogen)

Différents systèmes ont été construits avec des vecteurs pBAD et fournissent un ensemble de rapports induction/répression allant de 200 à 1200[41]. Ce rapport constitue une excellente gamme pour la répression de la plupart des gènes. Deux cas particuliers illustrent l'efficacité du système pBAD. Dans le premier cas, le système a permis de maintenir le niveau d'expression de la protéine CcdB en dessous de son seuil de toxicité en condition de répression[42]. Dans le second cas, le système a permis l'expression de la Rnase P archéale hautement toxique en quantité suffisante pour les études structurales[43,44]. Récemment, les avantages du promoteur P_{BAD} hautement régulé et l'efficacité du système d'expression T7 ont été combinés dans la construction d'une souche d'*E.coli* dans laquelle la copie chromosomique de l'ARN polymérase T7 a été mise sous le contrôle du promoteur P_{BAD} (souche BL21-AI, Invitrogen). Comparé à la souche avec le plasmide pLysS, la nouvelle souche à un niveau d'expression basal 4 fois plus faible et un niveau d'expression après induction similaire[45].

Le problème du promoteur P_{BAD} est qu'il est sujet au phénomène d'expression avec induction tout ou rien ou "all-or-none"[46]. En effet, en présence de concentrations intermédiaires en arabinose, certaines cellules de la population sont totalement induites tandis que d'autres restent non induites[47,48]. Ce phénomène se produit car l'expression du gène du transporteur de l'arabinose est contrôlée par l'arabinose lui-même (Figure 4). Différentes stratégies existent pour réduire le phénomène et obtenir une expression homogène dans toutes les cellules de la population. La première est de placer le gène du transporteur sous le contrôle d'un promoteur qui n'est pas régulé à l'arabinose par exemple sous le contrôle du promoteur P_{LAC} inductible à l'IPTG[49]. La seconde est d'utiliser une souche mutante de la protéine lacY qui permet une diffusion facilitée de l'arabinose[50,51].

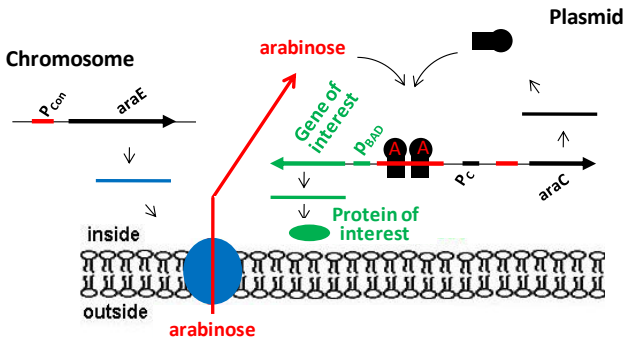


Figure 4 : Représentation schématique du système d'expression basé sur le promoteur P_{BAD} régulé avec la protéine araC et l'arabinose qui rentre dans la cellule grâce au transporter araE.

La stratégie de l'ARN anti sens

Une autre méthode efficace qui permet de diminuer le niveau d'expression basale des gènes, utilise la stratégie de l'ARN anti sens. Elle est basée sur la synthèse d'un ARN complémentaire à l'ARN messager du gène pour empêcher l'expression[52,53]. La stratégie de l'ARN anti sens a été utilisée avec succès pour exprimer l'endonucléase EcoRI dans *E. coli* après inactivation de sa méthylase[54] et la DnaseI pancréatique bovine[55]. Dans ces systèmes, l'expression du gène est sous le contrôle du promoteur λpL qui permet une bonne répression en présence du répresseur C1. Cependant, l'expression est encore réduite par la présence d'un promoteur anti sens *lac* en aval du gène. En présence de son inducteur (l'IPTG), le promoteur *lac* dirige la synthèse d'ARN messagers anti sens qui vont

former des duplexes d'ARN qui sont incapables d'être traduits par les ribosomes. De plus, la transcription dans cette direction interfère avec la transcription à partir du promoteur λpL à cause des collisions des ARN polymérases qui se dirigent dans des directions opposées[56]. Après la dérépression du promoteur fort λpL, des quantités suffisantes de protéines sont obtenues malgré l'inactivation d'une fraction des ARN messagers. L'efficacité du système pourrait sans doute être optimisée en augmentant le niveau d'induction du promoteur anti sens jusqu'au niveau requis pour neutraliser l'expression basale des gènes hautement toxiques.

L'utilisation de promoteurs induits par inversion

La stratégie de répression basée sur les promoteurs induits par inversion a été initialement décrite par A. J. Podhajska[57] et N. Hansan[58]. Dans ce système, le promoteur, entouré par deux sites d'attachement λ (att), est cloné dans une direction opposée au gène et est inversé par recombinaison site spécifique en utilisant une intégrase λ. Le niveau d'expression basale est donc contrôlé par la production de l'intégrase et peut être virtuellement de zéro dans une souche ne contenant pas de gène encodant l'intégrase. Dans la souche d'expression, le gène de l'intégrase est cloné dans un vecteur inductible à la chaleur sous le haut contrôle du promoteur λP_L/cI857 et du promoteur anti sens *ompA*. De cette façon, la production d'intégrase est réduite au maximum en condition de répression. Le système a été utilisé avec succès pour exprimer des domaines variables du récepteur des cellules T[59]. Une répression forte était nécessaire car le récepteur dans sa forme soluble et non dégradée est toxique. Par rapport aux stratégies décrites précédemment, ce système à l'avantage de permettre le choix du promoteur qui contrôle l'expression du gène hautement toxique.

L'utilisation de promoteurs compétitifs

La méthode basée sur l'utilisation de promoteurs compétitifs a été développée pour diminuer le niveau d'expression basale du promoteur T7. Dans ce système, un vecteur contenant un promoteur T7 supplémentaire est introduit dans la souche d'*E. coli* et va entrer en compétition avec le promoteur T7 du vecteur d'expression pour la liaison de la polymérase T7[60]. La diminution de l'expression basale obtenue grâce à ce système a permis de cloner et exprimer la protéase

HIV-1 dans *E. coli*[61]. Dans cette méthode, la différence du nombre de copies entre le vecteur compétitif et le vecteur d'expression dans la cellule hôte affecte l'efficacité de la compétition et donc la diminution du niveau d'expression basal obtenu. Pour les gènes hautement toxiques, ce rapport doit être optimisé.

La stratégie du promoteur hybride

Cette stratégie permet de créer de nouvelles séquences promotrices en combinant les éléments de régulation des opérons *lac*, *ara* et *tet*. Le point crucial dans la création de ces nouvelles séquences est de réaliser la bonne combinaison promoteur-opérateur. En effet, les paramètres cinétiques de l'interaction promoteur-ARN polymérase ainsi que la position de l'opérateur dans la séquence du promoteur, affectent l'occupation de l'opérateur par son répresseur et définissent l'efficacité de la répression[62]. Deux promoteurs ont été créés et nommés, $P_{LtetO-1}$ et $P_{lac/ara-1}$[63]. Dans le promoteur $P_{LtetO-1,}$ le site de liaison du répresseur CI a été remplacé par une séquence encodant l'opérateur 2 (*tet*O2) de l'opéron de résistance à la tétracycline du transposon Tn10[64]. Le promoteur hybride est hautement réprimé par le répresseur Tet et induit en ajoutant de l'anhydrotétracycline[63]. Quand le promoteur est combiné aux origines de réplication ColE1 (50-70 copies), p15A (20-30 copies) et pSC101 (3-4 copies), le rapport induction/répression est respectivement de 2500, 3500 et 5000[63]. Dans le promoteur $P_{lac/ara-1}$, le promoteur P_{lac} sauvage a été combiné avec le site de liaison du régulateur *ara*C. Le nouveau promoteur hybride est réprimé en ajoutant du glucose et induit en présence d'IPTG et d'arabinose. Un rapport induction/répression de 1800 a été obtenu quand le promoteur est combiné avec une origine de réplication pSC101. Le vecteur résultant a été utilisé avec succès pour exprimer l'endonucléase Cfr9I très toxique en l'absence de sa méthyltransférase.

La recombinaison génétique

Introduction

La recombinaison génétique joue un rôle clé dans la vie des organismes, elle permet entre autre la génération de diversité génétique, le maintien et la propagation des génomes, le contrôle de l'expression des gènes et le mouvement des particules virales dans et hors des chromosomes hôtes. La recombinaison génétique est divisée en trois grandes familles[65] qui se distinguent par les mécanismes mis en œuvre, les protéines impliquées et la nature de l'ADN recombiné. Premièrement, la recombinaison découverte par Morgan au début des années 1890 et qui fait intervenir des régions de molécules d'ADN homologues[66]. Deuxièmement, la recombinaison transpositionnelle décrite pour la première fois par McClintock dans le maïs et qui permet aux éléments transposables de bouger d'un endroit à un autre sur le chromosome[67]. Et finalement, la recombinaison site spécifique mise en évidence en premier par Campbell et qui permet des arrangements au niveau de sites hautement spécifiques sur le chromosome[68].

La recombinaison homologue

Dans *E. coli*, la recombinaison homologue est un mécanisme majeur pour le réarrangement de l'ADN, bien que son rôle biologique primaire soit la réparation des lésions associées à la réplication de l'ADN endommagé[69]. Comme elle permet l'insertion, la délétion et la modification de fragments d'ADN au niveau de sites homologues, elle a été largement utilisée comme outil génétique et biotechnologique[70].

Mécanisme de la recombinaison homologue RecA dépendante

Chez les bactéries, toutes les voies de recombinaison homologue dépendent de la protéine la RecA qui initie la réaction en liant l'ADN simple brin et en catalysant les réactions d'invasion, de couplage et de transfert du brin. L'approvisionnement en RecA est facilité par les interactions spécifiques avec la nucléase RecBCD[71,72] ou par l'action du complexe RecFOR[73-75]. Etant donné que la nucléase RecBCD possède une haute affinité pour les terminaisons d'ADN

double brin, la recombinaison qui implique des cassures double brin dans l'ADN emploie RecBCD dans l'étape d'initiation. Par contre, la réparation des brèches d'ADN simple brin semble être initiée par la protéine RecFOR qui augmente l'affinité de RecA pour l'ADN simple brin[76].

Les différentes étapes de la recombinaison homologue dans *E. coli* médiées par RecA et RecBCD sont explicitées ci-dessous (Figure 5).

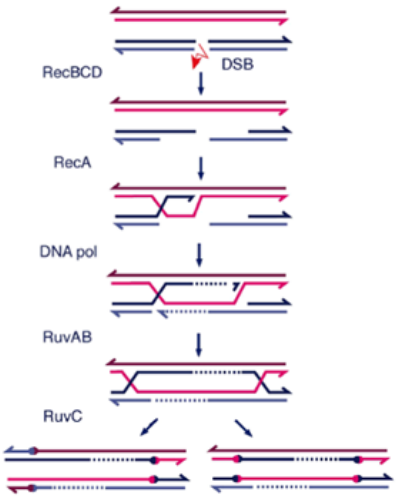


Figure 5 : Les étapes de la recombinaison homologue pour la réparation des cassures double brin dans *E. coli*. La description des étapes est détaillée dans le texte ci-dessous. (Figure reçue du Prof. B. Hallet).

Dans un premier temps, RecBCD qui possède des activités d'hélicase et de nucléase vient se lier à la molécule d'ADN au niveau de la cassure double brin. En utilisant l'énergie d'hydrolyse de l'ATP, RecBCD déroule la double hélice et dégrade la molécule d'ADN pour former une région d'ADN simple brin avec une extrémité 3' terminale. C'est alors qu'intervient RecA qui va se fixer sur le simple brin nouvellement formé, conduire l'extrémité 3' vers la chromatide homologue et déplacer le brin de même polarité. En utilisant le brin complémentaire sur le duplex homologue comme template et l'extrémité 3' du brin envahisseur comme amorce, la polymérase va synthétiser la région d'ADN détruite. Ensuite, les brins vont être scellés par une ligase pour former les structures en croix appelées jonctions de Holliday. La protéine RuvA va reconnaitre et se lier à ces jonctions et va ensuite recruter RuvB. Ces protéines vont fournir l'énergie nécessaire pour échanger les paires de bases et permettre

la migration des brins d'ADN. L'étape finale de la recombinaison implique la résolution de ces jonctions de Holliday par l'endonucléase RuvC. Les jonctions peuvent être résolues selon deux orientations alternatives : suivant l'orientation de résolution, les produits seront associés ou non à un crossing-over[77].

La fréquence de recombinaison dans *E. coli* est linéairement dépendante de la longueur de la séquence d'ADN homologue. En effet, cette séquence n'est pas seulement nécessaire pour les interactions initiales lors de la formation de l'hétéroduplex mais est également essentielle pour le processus de migration des brins car les sites de jonction ne traversent pas les régions d'ADN non homologues[78,79]. Il y a une longueur minimale en dessous de laquelle la recombinaison devient inefficace. Cette longueur est comprise entre 23-27 pb pour la voie dépendante de RecBCD et entre 44-90 pb pour la voie dépendante de RecFOR[80].

Recombinaison homologue entre deux plasmides

La recombinaison homologue intermoléculaire entre deux plasmides circulaires dans *E. coli* a été étudiée dans les années 80 par R. Kucherlapati[81] et plus récemment par S. T. Lovett[82]. Des plasmides à faible nombre de copies ont été utilisés par les deux chercheurs. Dans le cadre des expériences de S. T. Lovett, il s'agit d'un dérivé de pBR322 (15-20 copies) et d'un dérivé de pACYC (10-12 copies). R. Kucherlapati a travaillé avec deux plasmides dérivés de pBR322 (15-20 copies).

Dans ces deux études, la recombinaison a été mesurée par la fréquence de reconstruction d'un gène intact, le gène de la néomycine phosphotransférase (*neo)* et le gène qui confère la résistance à la tétracycline (*tet)* respectivement. Des plasmides ont été créés pour détecter la recombinaison homologue entre des séquences d'homologie de longueur croissante (25-501 pb). Le premier plasmide porte une séquence fixe correspondant à la partie 5′ de l'allèle *neo* ou *tet*A. L'autre plasmide porte une séquence de longueur variable correspondant à la partie 3' de l'allèle *neo* ou *tetA*. La recombinaison homologue, au niveau de la séquence commune fournit un plasmide dimérique et restitue le gène *neo* ou *tetA* sous une forme fonctionnelle.

Dans la souche RecA$^+$, le taux de recombinaison pour une séquence d'homologie de 25 pb est plutôt faible (10^{-8}) et augmente fortement avec l'accroissement de la longueur d'homologie (Figure 6). Le taux atteint un plateau de 2. 10^{-5} pour une séquence d'homologie supérieure à 200 pb[82]. Dans la souche RecA$^-$, les taux de recombinaison sont toujours inférieurs à ceux observés dans la souche RecA$^+$. La différence est peu marquée pour les séquences d'homologie faible mais atteint un facteur 100 pour les séquences supérieures à 200 pb.

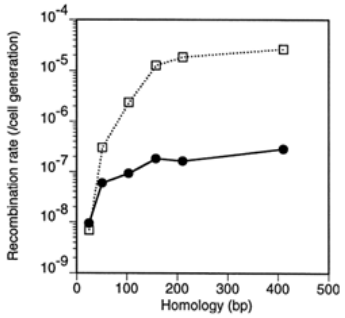


Figure 6 : Taux de recombinaison dépendant de RecA en fonction de la longueur d'homologie dans une souche RecA+ (carrés blancs) et RecA- (ronds noirs) (Repris de l'article de S. T. Lovett, 2002[82])

Recombinaison homologue impliquant de l'ADN phagique

De manière surprenante, la recombinaison homologue utilisant l'ADN simple brin des bactériophages (M13, fl, fd) comme substrat n'a été que faiblement étudiée. Cependant, quelques techniques l'utilisant ont été développées dans les années 80 pour le remplacement, la réparation et la mutagenèse de gènes.

Le système d'échange d'allèle développé par P. Blum constitue un premier exemple qui implique la recombinaison homologue entre un bactériophage M13 et le chromosome[83]. Il permet le remplacement d'une séquence d'ADN chromosomique par une séquence provenant d'un vecteur phagique. Pour pouvoir sélectionner uniquement les souches dans lesquelles le vecteur s'est intégré dans le chromosome, une méthode empêchant la réplication autonome du phage a été développée en utilisant la série de vecteurs M13mp. En effet, ceux-ci possèdent une mutation non-sens dans leur gène II empêchant leur réplication autonome dans des hôtes non suppresseurs[83]. Le produit du gène II, la "DNA nickase", est en effet une enzyme essentielle pour la réplication du

génome viral du phage M13. Dans les souches non suppressives, le vecteur ne pouvant se répliquer est donc perdu lors de la division cellulaire, sauf s'il se recombine avec le chromosome bactérien. Pour tester l'efficacité des phages M13mp à s'intégrer dans le génome par recombinaison homologue, les phénotypes de l'opéron histidine ont été utilisés. Des phages M13mp9 contenant l'opéron histidine ont donc été construits et ont ensuite été utilisés pour infecter des cellules non suppressives et auxotrophes à l'histidine.

Les conditions optimales pour l'intégration du vecteur phagique dans le chromosome ont été déterminées[84]. Deux variables ont été optimisées, le rapport de phages sur les cellules (multiplicité d'infection) et la durée d'absorption des phages. Les chercheurs ont observé que la fréquence de lysogénisation augmente rapidement en fonction du logarithme de la multiplicité d'infection (Figure 7). A une très haute multiplicité d'infection (10^5), et une séquence d'homologie de 3,1 kpb, l'intégration du bactériophage par recombinaison homologue se produit avec une fréquence proche de l'unité.

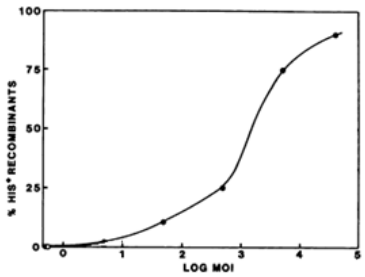


Figure 7 : Fréquence d'insertion du phage M13mp9:his2 par recombinaison homologue en fonction du logarithme de la multiplicité d'infection. Le temps d'absorption est de 60 min pour des multiplicités d'infection indiquées et une densité cellulaire de 10^6/ml (repris de l'article de S. Artz, 1989[84])

En utilisant la même approche et une séquence d'homologie de 1,5 kpb, des fréquences de lysogénisation de 10^{-2} ont été reportées avec une multiplicité d'infection de 1-100[85]. La durée d'absorption des phages a été optimisée en utilisant une multiplicité d'infection constante. Une recombinaison maximale nécessite un temps d'incubation minimal de 20 minutes pour des cellules en phase exponentielle et de 60 minutes pour des cellules en phase stationnaire. L'augmentation du temps d'incubation n'augmente pas la fréquence de recombinaison.

Un second exemple illustrant l'utilisation de phages est la méthode basée sur une double recombinaison permettant de récupérer des mutations dans le gène *lacI* sur un plasmide F' de 100 kpb grâce au vecteur phagique M13[86]. Le transfert de mutations a été observé en analysant la progéniture du phage après infection et, avec une séquence d'homologie de 2,2 kpb, le taux de recombinaison homologue reporté est de 7.10^{-4}.

La recombinaison site spécifique

La recombinaison site spécifique a été décrite pour la première fois par Campbell en 1962 après l'observation de l'intégration du chromosome d'un virus bactérien λ dans le chromosome de son hôte[68]. Ensuite, la liste des fonctions réalisées par les systèmes de recombinaison site spécifique n'a cessé de s'accroitre. D'autres exemples de fonctions biologiques catalysées par les recombinases sont la résolution des cointégrats générés par le saut des transposons apparentés à Tn3[87,88], l'inversion de la séquence G dans le phage Mu pour contrôler son infectivité et sa gamme d'hôte[89,90] et la résolution des dimères de molécules par la protéine Cre pour le maintien stable du plasmide du phage P1[91].

La recombinaison site spécifique fait intervenir une recombinase qui va reconnaitre et se lier à des sites spécifiques pour ensuite couper, échanger et religuer l'ADN. Le méchanisme de coupure et de religation de l'ADN intervient sans gain ou perte de nucléotides et avec une conservation stricte de l'énergie du lien phosphodiester. Les enzymes qui réalisent ce processus sont classées en deux familles distinctes en fonction du résidu qui joue le rôle de nucléophile dans la réaction de clivage d'ADN[92]. Il s'agit des recombinases à serine ou à tyrosine. En fonction de l'arrangement initial des séquences spécifiques, la recombinaison mène à trois résultats : l'intégration, l'excision (résolution) ou l'inversion (Figure 8).

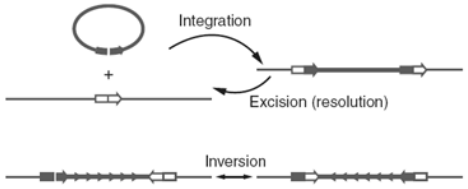


Figure 8 : Les trois résultats de la recombinaison site spécifique (Repris de l'article de P. Rice, 2006[93]).

Les composants essentiels du système de recombinaison site spécifique sont la recombinase et les deux sites spécifiques. Les sites spécifiques les plus simples sont des segments courts de duplex d'ADN, de ~30 pb en longueur, qui contiennent une paire inversée de séquences de reconnaissance parfois appelés RBE (recombinase binding element). Le clivage et la ligation de l'ADN ont lieu entre les deux séquences de reconnaissance. Dans la nature, les sites de recombinaison sont souvent plus complexes et contiennent des séquences additionnelles qui peuvent avoir une longueur de plus de 100 pb. Ces séquences peuvent jouer un rôle structurel, de régulation, ou les deux[94]. Elles peuvent initier ou stabiliser l'assemblage des sites de recombinaison, inhiber les assemblages inappropriés, guider les domaines catalytiques des recombinases vers les sites de recombinaison, activer les recombinases et déterminer la sélectivité de la recombinaison (par exemple, promouvoir la délétion et empêcher l'intégration et vice versa).

Le processus de la recombinaison site spécifique

Le processus de recombinaison site spécifique peut être divisé en une série d'étapes simples (Figure 9). Tout d'abord, les molécules de recombinase se lient aux deux sites spécifiques de recombinaison. Les sites s'assemblent alors pour former le complexe synaptique. C'est dans ce complexe que les molécules de recombinases catalysent l'échange de brins, en clivant et religant l'ADN. Le mécanisme est similaire à celui utilisé pas les topoisomérases. Les brins d'ADN ne sont pas clivés par hydrolyse mais par attaque nucléophile du lien phosphodiester par le groupe hydroxyle de la sérine ou la tyrosine présente sur la chaine latérale de la recombinase. D'un coté de la coupure, un lien covalent recombinase-ADN est formé et de l'autre coté un groupe hydroxyle est libéré. La ligation des brins d'ADN est réalisée par le processus inverse. L'hydroxyle libéré attaque le lien phosphodiester entre la recombinase et l'ADN de l'autre partenaire, libère la recombinase et donne le produit recombinant.

Les détails du processus diffèrent en fonction de l'utilisation d'une tyrosine ou d'une sérine pour l'attaque nucléophile. Les recombinases à sérine introduisent des cassures double brin au niveau des deux sites de recombinaison. Les quatres brins d'ADN sont clivés avant que l'échange ne soit initié[88]. Au contraire, les recombinases à tyrosine sont contrôlées par un mécanisme allostérique qui

empêche deux molécules adjacentes d'être actives au même moment sur le tétramère[95,96]. Un brin à la fois est clivé sur chaque duplexe et ce brin doit être ligué à son partenaire avant que le second brin ne soit clivé. Cela produit un intermédiaire appelé jonction de Holliday[97,98].

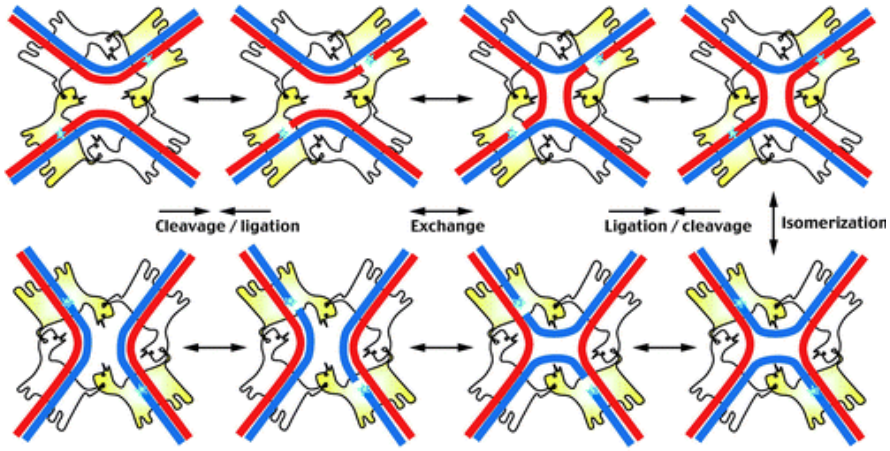


Figure 9 : Représentation schématique de l'échange de brin catalysé par le recombinase à tyrosine. Le complexe synaptique comprend deux duplex d'ADN liés à quatre recombinases. Les étoiles bleues représentent les sites actifs. Un brin de chaque duplexe est clivé, échangé et ligué pour former la jonction de Holliday (Repris de l'article de P.A. Rice, 2006[93]).

La recombinaison site spécifique catalysée par le recombinase à tyrosine TnpI

La recombinase à tyrosine TnpI est une enzyme encodée avec la transposase TnpA sur l'élément génétique mobile Tn4430 de *bacillus thuringiensis* (Figure 10). Le rôle de la TnpI intervient après celui de la TnpA. En effet, dans un premier temps, la transposase TnpA coopère avec la machinerie de réplication de l'hôte pour former un cointégrat dans lequel les plasmides donneurs et cibles sont fusionnés avec deux copies du transposon dans des orientations directes à leurs jonctions. Ensuite, la recombinase TnpI intervient pour résoudre le cointégrat en deux plasmides en catalysant la réaction de recombinaison site spécifique entre les deux copies dupliquées du transposon. Le transposon Tn4430 est délimité par deux séquences terminales répétées inverses de 38 pb (TIR), homologues à celles de Tn3, Tn21 ou Tn917, qui sont nécessaires pour la transposition réplicative. Tn4430 appartient donc à la famille du transposon Tn3. Cependant, contrairement aux autres transposons de cette famille, qui sont connus pour résoudre les cointégrats avec une recombinase de la famille des recombinase à

sérine[99], le transposon Tn4430 utilise la recombinase TnpI qui fait partie de la famille des recombinase à tyrosine[100,101], comme par exemple la protéine Int du bactériophage λ, Cre du bactériophage P1 ou XerC et XerD d'*E.coli*.

La TnpI se lie à des motifs fonctionnels dans le site de résolution interne (IRS, 116 pb) du transposon Tn4430 pour catalyser la réaction de résolution (Figure 10). Deux motifs composés de séquences répétées inverses (IR1 et IR2) forment le centre du site de recombinaison ou "site-cœur" au niveau desquels vont se réaliser le clivage et l'échange des brins. De plus, deux motifs additionnels (DR1 et DR2) adjacents au site-cœur sont nécessaires pour une activité de recombinaison maximale et imposent une sélectivité de résolution à la réaction[102]. En effet, la liaison de la TnpI aux motifs DR1 et DR2 résulte en la formation d'un complexe synaptique spécifique dans lequel l'orientation directe des sites de recombinaison est favorisée. En absence de DR1 et DR2, le complexe synaptique est relâché et la réaction donne lieu à une variété de produits par collisions aléatoires.

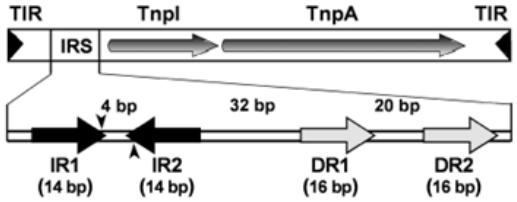


Figure 10 : Organisation du transposon Tn4430. Les séquences terminales répétées inverse de 38 pb (TIR) sont représentées par des triangles. Les flèches représentent les régions codantes de la recombinase TnpI et de la transposase TnpA. La position du site interne de recombinaison (IRS, 116 pb) est indiquée et son organisation est détaillée en dessous. Les séquences répétées inverses de 14 pb (IR1 et IR2) et les séquences répétées en tandem de 16 pb sont représentées par des flèches noires et grises respectivement. Les distances qui séparent les différents motifs sont indiquées (repris de l'article de Vanhooff et al., 2006[102]).

La TnpI agit donc à la fois comme composant catalytique du complexe de recombinaison et comme élément de régulation pour assurer l'interaction sélective entre les sites de recombinaison. La TnpI peut fonctionner avec une grande variété de substrats *in vitro*, et également avec différents types de cellules *in vivo*, incluant les cellules mammifères. Cette recombinase peut donc être exploitée pour créer de nouveaux outils pour la biologie moléculaire et l'ingénierie génétique[103].

Dans le cadre de ce travail, cette TnpI, reçue du Prof. B. Hallet, a été utilisée pour résoudre le cointégrat formé après recombinaison homologue entre un plasmide

et un phagemide. Un site interne de recombinaison (IRS) contenant les sites accessoires DR1 et DR2 ont été clonés sur chacun des vecteurs pour favoriser les réactions de délétion de façon irréversible.

Les stratégies de biosynthèse de peptides cycliques

Introduction

Plus des deux tiers des antibiotiques utilisés de nos jours pour le traitement des maladies humaines sont des produits naturels, c'est-à-dire des petites molécules issues d'organismes vivants, ou des dérivés semi-synthétiques de ces molécules[104-106]. Parmi ces produits naturels se trouve une catégorie de composés particulièrement intéressants que sont les peptides cycliques. De nos jours, deux voies principales utilisées par les organismes vivants pour leur synthèse ont été décrites, l'une est dépendante et l'autre indépendante des ribosomes[107].

La voie de biosynthèse non ribosomique

La voie de synthèse de peptides cycliques indépendante des ribosomes a été découverte dans les années 1970[108]. Cependant, les détails des mécanismes réactionnels n'ont été élucidés que 20 ans plus tard dans les années 1990[109]. Dans cette voie, les peptides sont synthétisés par de grands complexes enzymatiques appelés synthétases ou NRPS (Non Ribosomal Peptide Synthetase)[110]. Les synthétases sont organisées en modules et chaque module est responsable de l'incorporation d'un acide aminé spécifique dans la chaîne peptidique en formation. Les modules, composés de plusieurs domaines, possèdent une organisation commune leur permettant d'activer un acide aminé fixé et de le lier par une liaison à un autre acide aminé, mais également une certaine spécificité pour leur permettre la fixation d'un acide aminé plutôt qu'un autre. La spécificité de liaison au substrat des séries de modules reflète l'ordre des acides aminés dans les produits finaux. Il a été prouvé qu'un module contient 3 domaines (Figure 11) :

- Adénylation (A) : domaine central dans l'action des peptides synthétases. Il permet la fixation d'un acide aminé spécifique et son activation grâce à une réaction d'adénylation sur ce dernier.
- Thiolation (T ou PCP pour Peptidyl Carrier Protein) : permet au peptide en cours de synthèse de rester fixé à la synthétase tout au long du processus d'élongation via une liaison thioester.
- Condensation (C) : permet la catalyse des différentes réactions et la formation de la liaison peptidique.

Un domaine spécifique intervient pour la libération du peptide : le domaine Te (thioestérase). Ce domaine va couper la liaison thioester qui relie le peptide synthétisé au dernier domaine T.

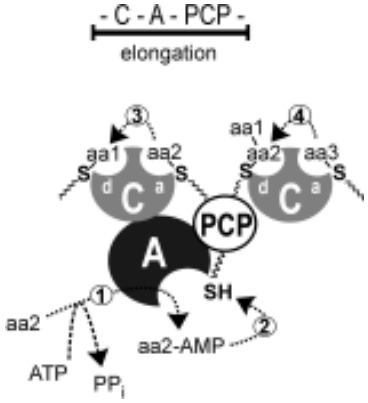
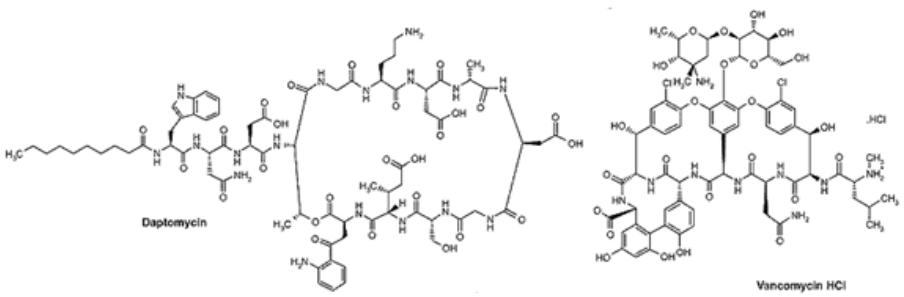


Figure 11 : Représentation schématique de l'organisation des domaines d'un module de NRPS. Un module contient au minimum un domaine C, A et PCP. La description des étapes du méchanisme de synthèse est détaillée dans le texte ci-dessus (repris de l'article de Marahiel, 2002[111])

Les acides aminés utilisés dans cette voie ne sont pas uniquement les vingt acides aminés protéogéniques. En effet, plusieurs centaines d'acides aminés différents peuvent être incorporés au sein du peptide et en plus, dans la plupart des cas, des modifications ont lieu durant ou après l'élongation de la chaine pour encore complexifier la structure[112]. Ces modifications peuvent être des méthylations, des épimérisations, des hydroxylations, des hétéro-cyclisations, des formylations… La seule limite de cette voie de biosynthèse concerne la longueur de la chaîne peptidique formée qui ne peut excéder 48 résidus. La vancomycine[113], la daptomycine[114] et la cyclosporine sont des exemples de peptides cycliques non ribosomiques (Figure 12).



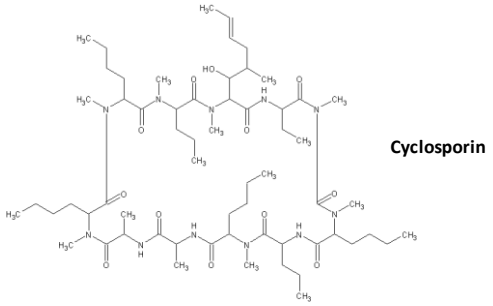


Figure 12 : Structure de trois peptides cycliques non ribosomiques synthétisés par les NRPS (repris de l'article de P. A. Hunter, 2001[115])

Par cette voie, la nature a pu créer une grande diversité de composés naturels en réalisant des combinaisons diverses des différents modules et en jouant sur les modifications intervenant durant et après la synthèse. Cependant, la variabilité disponible dans de tels systèmes n'a certainement pas été entièrement exploitée. Par l'ingénierie génétique, il est donc possible de créer de nouveaux peptides naturels "non naturels" qui peuvent être aussi structurellement divers et aussi précieux au niveau thérapeutique que les produits naturels existants[116,117]. Le challenge dans ce cas est de réussir à produire *in vivo* de nouvelles synthétases avec des interactions aussi stables entre les différents modules. Des groupes de chercheurs ont déjà développé différentes stratégies et de nouvelles molécules ont déjà été créées[118-121].

Un exemple intéressant est la biosynthèse d'analogues de la daptomycine (Figure 12)[122], un composé qui possède une activité bactéricide contre les pathogènes à gram positif. Trois gènes sont impliqués dans sa biosynthèse, *dptA*, *dptBC* et *dptD* et codent pour trois sous-unités NRPS nécessaire pour l'incorporation des 13 acides aminés (Figure 13). Les nouveaux composés dérivés de la daptomycine ont été créés en échangeant un ou des module(s) dans la sous-unité de DptBC au niveau de sites spécifiques par la stratégie de recombinaison médiée par la recombinase λ-Red décrite par B. Wanner[123] (Figure 14). Ces échanges génèrent des modifications des acides aminés présents dans la structure cœur de l'antibiotique.

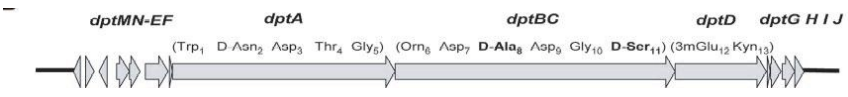


Figure 13 : Représentation schématique de l'organisation des sous-unités et des domaines du gène codant pour le grand complexe enzymatique responsable de la formation de la daptomycine (repris de l'article de K. Nguyen, 2006[124]).

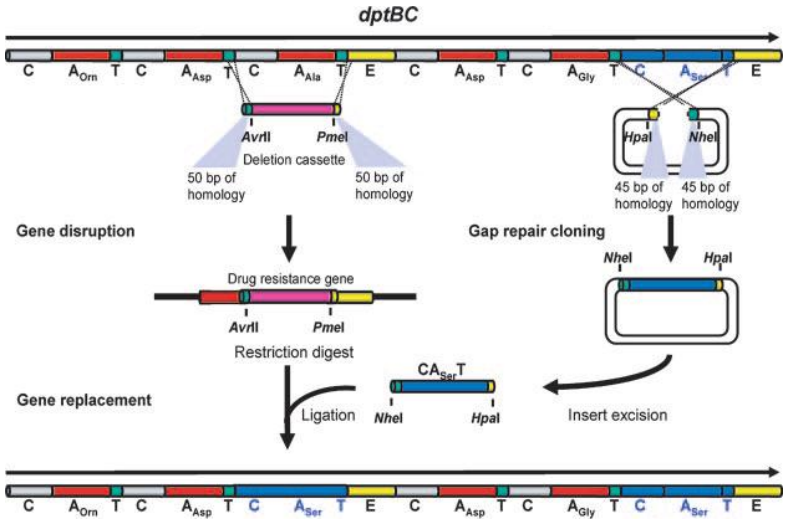


Figure 14 : Stratégie pour la construction d'un gène hybride dptBC qui utilise la recombinaison médiée par la recombinase λ-Red. La région codante des domaines du module D-Ala$_8$ ont été remplacé par une cassette de résistance à un antibiotique terminant par des sites de restriction rares. La cassette a ensuite été excisée et remplacée par un fragment codant pour d'autres domaines grâce à des sites de restriction compatibles (repris de l'article de B. Wanner, 2000[123]).

Malgré le succès des premières stratégies, des grandes banques combinatoires de composés n'ont pas encore été réalisées et cela s'avère difficile car les méthodes d'ingénierie basées sur la dégénérescence du code génétique ne peuvent pas être envisagées.

La voie de biosynthèse ribosomique

La voie de biosynthèse des peptides cycliques dépendante des ribosomes a été découverte à la fin des années 1980 mais, à nouveau, les avancées dans la compréhension des mécanismes ont seulement été réalisées récemment, en 2008, par deux groupes de recherche différents dirigés par E. W. Schmidt[125] et C. Hertwewk/E. Dittmann[126]. Dans cette voie, le gène encodant le précurseur du peptide est associé avec les gènes des enzymes responsables de sa maturation dans un opéron. Le précurseur biologiquement inactif est tout d'abord synthétisé à partir d'un ARNm et subit ensuite des modifications enzymatiques post-traductionnelles spécifiques pour aboutir au peptide actif[127]. Le précurseur contient, en plus de la séquence peptidique complète du produit, un peptide leader N-terminal (LP) qui, très probablement dirige les modifications avant de libérer le produit final[128]. Les enzymes responsables des modifications peuvent

être des déshydratases, des cyclo-déshydratases ou cyclo-déshydrogénases, des méthyle-transférases, des protéases, des désaminases, des mono-oxygénases,… On peut citer dans les exemples de peptides cycliques ribosomiques, les thiocillines[129], les trunkamides[125] et les patéllamides[130].

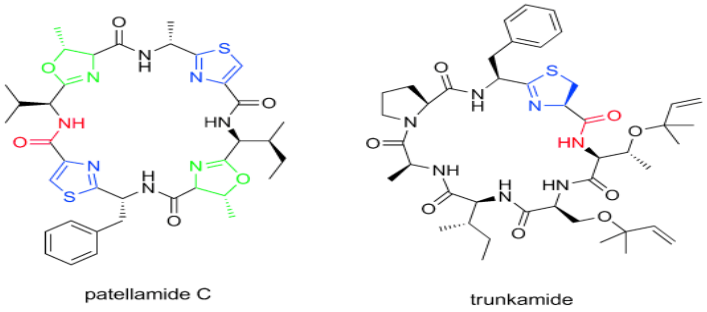


Figure 15 : Structure de deux peptides cycliques ribosomiques (repris de l'article de B. S. Moore, 2008[130])

La découverte de cette voie de biosynthèse a permis d'élargir les connaissances de base sur les modifications post-traductionnelles et d'envisager la possibilité de créer par ingénierie de nouveaux peptides naturels "non naturels" en combinant les gènes modificateurs de différents peptides. Des premières expériences ont révélé que la reprogrammation de la synthèse des peptides ribosomiaux peut être réalisée[131,132]. Il sera donc intéressant de suivre si ces recherches mèneront à de nouveaux médicaments.

En 1999, une stratégie appelée SICLOPPS basée sur l'ingénierie des intéines et permettant la biosynthèse de peptides cycliques naturels "non naturels" a été développée par S. J. Benkovic[133]. La suite du chapitre présentera les intéines et l'évolution des progrès dans leurs applications biotechnologiques qui ont finalement mené au développement de la technologie SICLOPPS.

La biosynthèse de peptides cycliques via les intéines

Introduction sur les intéines

Les intéines sont des éléments génétiques similaires aux introns, insérés en phase dans des régions codantes hautement conservées de protéines essentielles[134,135] (Figure 16). Cependant, contrairement aux introns, les intéines sont transcrites et traduites avec leur protéine hôte. C'est seulement au niveau

protéique que l'intéine s'excise de manière auto-catalytique de sa protéine. Au cours de la réaction d'épissage protéique, l'intéine catalyse la ligation peptidique des deux séquences protéiques adjacentes, appelées extéines. La ligation des extéines différencie l'épissage protéique de l'auto-protéolyse effectuée par certaines protéines[136]. Les intéines conservent leur activité quand elles sont insérées dans une nouvelle protéine hôte ou exprimées dans un contexte non natif, ce qui les rend très utiles pour les applications d'ingénierie protéique.

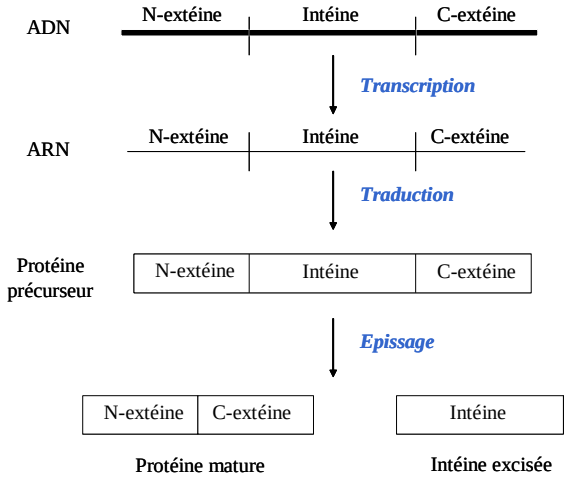


Figure 16 : Illustration schématique de l'épissage protéique catalysé par les intéines.

Nomenclature

Les intéines sont nommées en fonction de l'organisme et de la protéine hôte dans lesquels elles résident[134,137]. On utilise la première lettre du genre de l'organisme hôte, suivie des deux premières lettres de l'espèce et finalement trois ou quatre lettres du nom de la protéine hôte. S'il existe plus d'une intéine dans une même protéine, les différentes intéines sont numérotées. Les intéines qui occupent un site homologue dans une protéine hôte dans des organismes différents sont nommées intéines allèles.

Découverte et distribution

La première intéine a été découverte au début des années 1990 dans le gène TFP1 qui code la sous-unité catalytique de l'ATPase vacuolaire de *Saccharomyces cerevisiae*[138-140]. Depuis, un grand nombre d'intéines ont été identifiées, elles sont

présentes principalement chez les archaea mais également chez les bactéries, les levures, certaines algues et même dans les virus. Aujourd'hui, plus de 400 intéines sont connues dans plus de 30 différents types de protéines[137,138]. Les intéines sont généralement localisées dans des gènes essentiels à la vie de l'organisme, principalement dans des gènes codant des protéines du métabolisme de l'ADN (ADN polymérase, hélicase, gyrase, recombinase, ribonucléotide réductase)[141]. Le plus souvent, il n'y a qu'une seule intéine par protéine. Chez les archae, on connaît des gènes possédant deux, voire trois intéines. Dans la protéine hôte, l'intéine est située préférentiellement au niveau de régions importantes pour l'activité de la protéine : site actif, poche de fixation du substrat ou poche de fixation du co-facteur[142,143].

Structure

Les intéines ont une taille comprise entre 134 et 1650 acides aminés[144]. La majorité d'entre elles sont bifonctionnelles et constituées d'un domaine d'épissage avec des régions N- et C- terminales entrecoupées par un domaine endonucléase (Figure 17). Les deux régions d'épissage forment un domaine fonctionnel unique, au niveau de la structure tri-dimensionnelle de l'intéine. Le domaine endonucléase central ne participe pas au mécanisme d'épissage protéique[145] mais assure la transmission horizontale du gène de l'intéine[146-149]. Comme ce domaine central n'est pas essentiel pour l'épissage, des mini-intéines, c'est-à-dire des intéines sans ce domaine, ont été créées par ingénierie[150,151]. Quelques mini-intéines existent également à l'état naturel, elles représentent environ 20% des intéines totales[137,152]. Une troisième catégorie d'intéines, sont les intéines dimériques naturelles[153] ou générées par ingénierie[154-156] capables de réaliser des réactions d'épissage intermoléculaire ou en *trans*.

Les intéines ont une faible identité de séquence, excepté les intéines allèles trouvées dans un même site d'insertion dans des protéines homologues d'organismes différents. Cependant, différents motifs de courtes séquences ont un degré d'identité plus élevé parmi les intéines (Figure 17). Les motifs A et B dans la région d'épissage N-terminale, les motifs F et G dans la région d'épissage C-terminale, les motifs N2 et N4 (nettement moins conservés) dans la région d'épissage N-terminale et finalement les motifs C, D, E et H des endonucléases (dans le cas des intéines à endonucléases). Les intéines partagent des résidus hautement conservés au niveau des jonctions d'épissage. Le premier résidu de

l'intéine est presque toujours une cystéine (parfois, une sérine). Les deux derniers résidus de l'intéine sont une histidine-asparagine. Le premier résidu de la C-extéine (résidu +1) est une cystéine, une sérine ou une thréonine. A l'intérieur de l'intéine, quelques autres résidus sont également conservés, par exemple, une thréonine et une histidine dans le motif B.

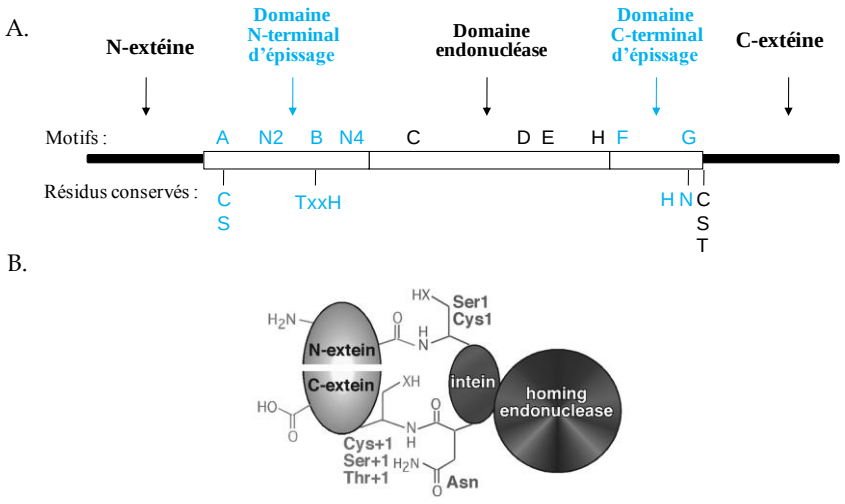


Figure 17 : Partie A. Organisation d'un précurseur d'épissage avec un domaine endonucléase. Les motifs de l'intéine sont montrés au dessus du précurseur et les résidus conservés sont en dessous du précurseur. Partie B. Folding du précurseur pour former le site actif d'épissage (repris de l'article de B. Perler 2006[135]).

L'intéine utilisée dans le cadre de la thèse est la mini-intéine Ssp dnaB constituée de 154 acides aminés. Cette mini-intéine dérive d'une intéine, composée de 429 acides aminés, encodée dans le gène de la "replicase DNA helicase" de la cyanobactérie *Synechocystis sp.* PCC6803. La mini-intéine a été obtenue en délétant le domaine central endonucléase de l'intéine, constitué de 275 acides aminés, résidus 107 à 382. La structure cristallographique de la mini-intéine Ssp dnaB a été résolue en 2002 par l'équipe de Y. Ding grâce à la méthode de la diffraction des rayons X[157]. La protéine utilisée pour déterminer la structure contient 4 résidus de l'extéine situés en début et fin de séquence (les résidus Ser-2, Gly-1, Ser+155 et Ile+156). De plus, deux résidus (Cys+1 et Asn+154) ont été substitués en alanine pour capturer le précurseur en un état de pré-splicing. Pour modéliser la structure de la mini-intéine sauvage, les deux alanines ont été changées en leur résidu d'origine à l'aide du programme "Swiss-PdbViewer". La représentation a été réalisée à l'aide du programme "MSViewer" (Figure 18).

La structure, constituée de 12 brins β, a une forme de fer à cheval assurée par deux brins β antiparallèles. Ce motif en forme de fer à cheval est conservé parmi les intéines. Les deux régions N- et C- terminales de l'intéine sont fortement associées, avec de nombreux ponts hydrogènes et des interactions hydrophobes. Les jonctions d'épissage sont localisées au centre de la structure. Tous les résidus importants pour l'épissage sont très proches les uns des autres. Dans le domaine N-terminal (en rouge), on peut voir les résidus Thr70 et His73 ainsi que la Cys1. Dans le domaine C-terminal (en bleu), on retrouve l'His153, l'Asn154 et la Ser155 (de la C-extéine).

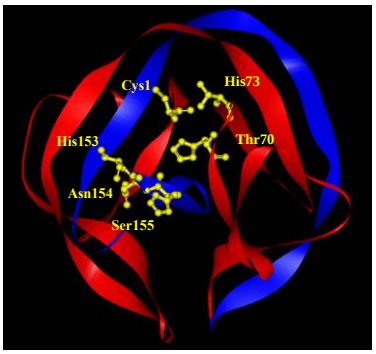


Figure 18 : Modèle de la structure tridimensionnelle de la mini-intéine Ssp dnaB obtenu avec le programme "MSViewer" après substitution des deux résidus Ala à l'aide du programme "Swiss-PdbViewer". La région d'épissage N-terminale est représentée en rouge et la région C-terminale en bleu.

Mécanisme d'épissage

Les connaissances actuelles des réactions chimiques qui sont à la base de l'épissage protéique proviennent pour la plupart des études réalisées par F.B. Perler[158,159] et H. Paulus[160]. Dans un premier temps, des études par mutagenèse dirigée ont permis d'identifier les résidus essentiels pour l'épissage[161]. Ensuite, la découverte d'une intéine thermostable dans l'ADN polymérase de *Pyrococcus sp.*[162,163] a permis d'élucider la nature générale du processus et les réactions chimiques impliquées. Par des observations *in vitro*, ils ont déterminé que l'épissage est entièrement catalysé par des résidus présents dans l'intéine, est un processus intramoléculaire[164] et ne nécessite aucun cofacteurs ou sources d'énergie métabolique. Le mécanisme implique des réarrangements de liaisons chimiques plutôt qu'un clivage de deux liaisons suivi d'une synthèse d'une nouvelle liaison peptidique.

L'épissage protéique s'effectue en quatre étapes (Figure 19) :

- **Etape 1** → Attaque nucléophile par la chaîne latérale du premier résidu de l'intéine (Ser ou Cys) sur le groupement carbonyle précédent
- **Etape 2** → Attaque nucléophile par le groupement hydroxyle ou thiol du premier résidu de la C-extéine (Ser, Thr ou Cys) sur le lien (thio)ester
- **Etape 3** → Clivage du lien amide au niveau de l'extrémité C-terminale de l'intéine par cyclisation de l'Asn
- **Etape 4** → Réarrangement spontané du lien (thio)ester entre les extéines liées pour obtenir la forme finale la plus stable

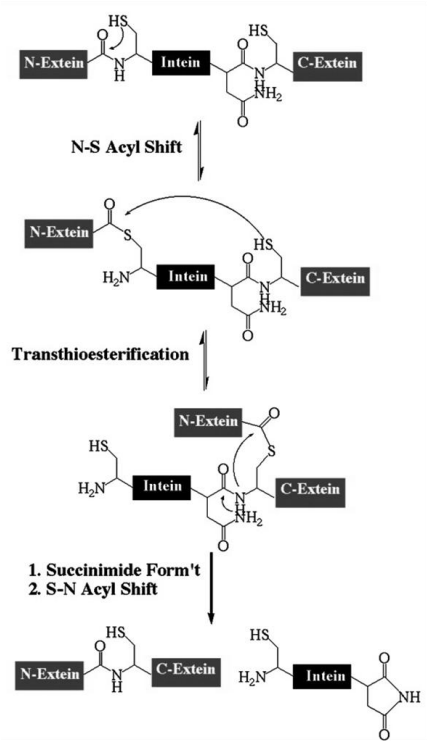


Figure 19 : Mécanisme d'épissage protéique d'une intéine à cystéines Les chaînes latérales du premier (Cys ou Ser) et du dernier (Asn) résidu de l'intéine sont représentées ainsi que celle du premier résidu de la C-extéine (Cys ou Ser). Le mécanisme est détaillé dans le texte (repris de l'article de T. W. Muir, 2004[165]).

La dernière étape est une réaction rapide et spontanée mais les trois premières nécessitent d'être catalysées par l'intéine. Dans ce sens, l'intéine peut être vue comme une enzyme à turnover unique dont les substrats sont les jonctions intéine-extéine qui se trouvent dans un seul centre catalytique.

Applications en ingénierie protéique

L'utilité des intéines en ingénierie protéique a été comprise après la résolution du mécanisme d'épissage. En effet, la compréhension des activités de clivage et de ligation a mené à l'utilisation des réactions catalysées par l'intéine dans de multiples applications, comme par exemple, la purification protéique, l'addition de marqueurs d'affinité, l'analyse des interactions protéine-protéine, la semi-synthèse et la cyclisation protéique (Tableau 1). Toutes ces applications sont basées principalement sur quatre technologies : l'épissage, l'autoclivage, la ligation de protéines exprimées (EPL) et l'épissage en trans (T-S).

Application	Method
Protein purification	Cleavage
Specific phosphorylation	EPL
Protein tagging – fluorescence	EPL
Protein tagging – binding to solid supports	EPL, T-S
Protein tagging – nucleic acids	EPL
Incorporation of biophysical probes or unnatural aa	EPL, T-S
Protein targeting	EPL, T-S
Partial isotopic labeling	EPL, T-S
Backbone cyclization in vitro	pTwin (EPL)
Backbone cyclization in vivo - cyclic peptide libraries	SICLOPPS
Small molecule molecular switches	Splicing, T-S
Temperature dependent splicing	Splicing
Semisynthesis of cytotoxic proteins	EPL, T-S
Semisynthesis in vivo	EPL,T-S
Enzyme mechanism studies	EPL,T-S
Protein interactions	EPL, T-S
Transgenic plants	T-S, splicing

Tableau 1 : Exemples des applications utilisant les intéines (repris de l'article de F. B. Perler, 2006[166]).

La technologie de l'autoclivage : la purification protéique

La première application des intéines est apparue en 1997 avec le développement d'un système permettant la purification de protéine en une seule étape chromatographique par M.Q. Xu[167,168]. Ce système utilise une intéine modifiée en conjugaison avec un domaine de liaison à la chitine (CBD) comme marqueur d'affinité (Figure 20). Le concept est basé sur l'observation que l'intéine mutée au niveau des résidus catalytiques de la jonction C-terminale peut être induite pour réaliser une réaction de clivage autocatalytique au niveau de son extrémité N-terminale par ajout d'un thiol[136]. La protéine d'intérêt est clonée en phase au

niveau de l'extrémité N-terminale de la fusion intéine-CBD et est purifiée par absorption sur une colonne de chitine. L'induction du clivage autocatalytique libère la protéine d'intérêt tandis que la fusion intéine-CBD reste liée à la colonne. Un grand avantage de cette technologie est que le clivage est réalisé sans l'ajout d'une protéase exogène qui risquerait de dégrader la protéine d'intérêt et nécessiterait une étape de purification supplémentaire.

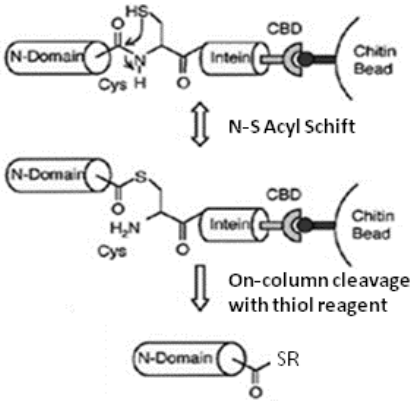


Figure 20 : Représentation schématique du système de purification protéique utilisant la technologie de l'autoclivage des intéines. Le précurseur exprimé contient une intéine modifiée en C-terminal, contenant un domaine de liaison à la chitine et ne permettant que le clivage en N-terminal de l'intéine (repris et modifié de l'article de H. Yu, 1999[169]).

La technologie de l'Expressed Protein Ligation (EPL)

Cette première application a ouvert la voie vers le développement d'une méthode appelée "Expressed Protein Ligation" (EPL) qui est une variante de la "Native Chemical Ligation" (NCL)[170,171] qui permet la ligation d'un peptide présentant un thioesther C-terminal avec un peptide muni d'un cystéine N-terminale. En effet, en 1998, en étudiant la réaction d'autoclivage de l'intéine modifiée utilisée pour la purification, P.A. Cole[172] a observé qu'il était possible d'intercepter le thioesther généré dans la protéine d'intérêt avec un peptide naturel ou synthétique contenant une cystéine N-terminale pour former une protéine semi-synthétique (Figure 21). Ensuite, la puissance et la flexibilité de la technologie ont mené à de nombreuses applications qui sont reprises dans la revue de R. M. Hofmann[173]. On peut citer en particulier l'incorporation d'acides aminés non codés dans une séquence protéique[174,175], la synthèse de protéines

cytotoxiques[176], le marquage de protéines pour les analyses RMN[177] et l'addition de sondes fluorescentes pour la création de biosenseurs[178,179].

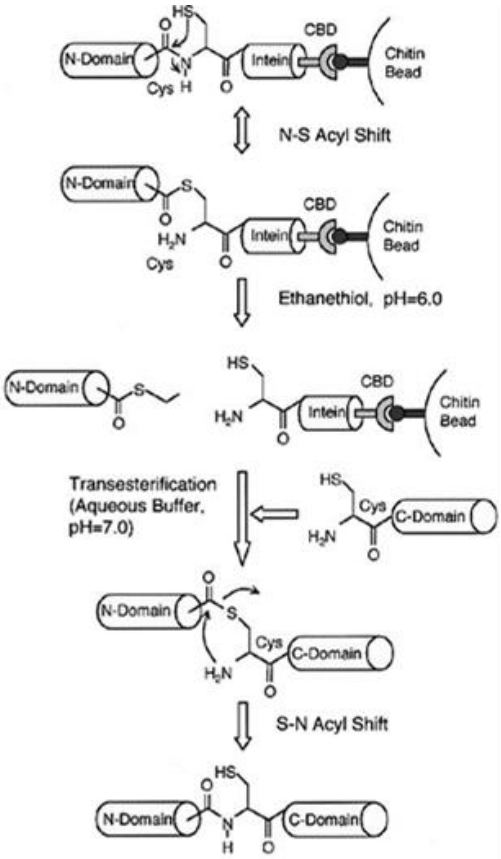


Figure 21 : Représentation schématique du système de semi-synthèse protéique utilisant la technologie de l'"Expressed Protein Ligation". Le système implique la ligation de deux peptides, l'un porteur d'une cystéine en N-terminal et l'autre d'un α-thioester en C-terminal. La réaction entre ces deux peptides engendre la formation d'un lien peptidique. L'α-thioester du fragment N-terminal est généré par un clivage sur colonne d'affinité d'un précurseur synthétisé *in vivo* et contenant une intéine modifiée (repris de l'article de H. Yu, 1999[169]).

De plus, en 1999, deux groupes distincts dirigés par M.-Q. Xu et T.-W. Muir ont développé simultanément une variante intramoléculaire de l'EPL qui permet de cycliser une protéine d'intérêt *in vitro* en générant des groupes réactifs complémentaires sur la même protéine. Dans la stratégie de M.-Q. Xu[180], appelée système TWIN (Two Inteins), la protéine d'intérêt est clonée entre deux intéines modifiées fusionnées à des marqueurs d'affinité, l'une ayant une activité de clivage C-terminal et l'autre une activité de clivage N-terminal (Figure 22). Dans celle de T.-W. Muir[181], le thioester à l'extrémité C-terminale de la protéine

d'intérêt est généré grâce à la fusion avec une intéine modifiée et la cystéine à l'extrémité N-terminale est libérée par clivage d'une protéase spécifique.

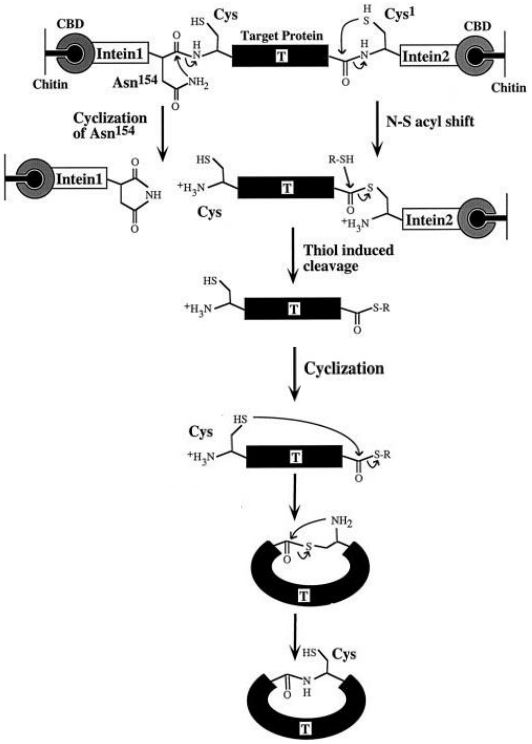


Figure 22 : Système TWIN. Le gène de la protéine à cycliser est bordé par deux gènes d'intéines modifiées et fusionnées à un domaine de liaison à la chitine. Une fois exprimé, le précurseur se lie par ses deux extrémités à la colonne et l'épissage est induit par le pH et par un thiol. La protéine résultante peut ensuite polymériser ou cycliser en fonction de ses propriétés (repris de l'article de M.Q Xu, 1999[18]).

Un peu plus tard, en 2001, T.-W. Muir a démontré que sa variante intramoléculaire de l'EPL peut également être utilisée *in vivo* et a prouvé que la réaction de ligation est remarquablement propre et efficace pour la cyclisation de protéines[16]. Pour les peptides, l'efficacité de la stratégie n'a pas encore été prouvée et il se pourrait qu'elle soit moins efficace car le clivage par la protéase est influencé par les résidus en aval de la cystéine. Cependant, la cystéine N-terminale peut être introduite en utilisant une autre méthode comme une intéine modifiée[180] ou une stratégie de fusion à l'ubiquitine[182]. En effet, pour la seconde méthode, l'acide aminé désiré peut être produit à l'extrémité N-terminale après clivage par une protéase spécifique de l'ubiquitine.

La technologie de l'épissage en *trans*

La technologie de l'épissage en *trans* a été découverte par différents groupes indépendants en 1998 grâce à l'observation qu'une intéine, coupée artificiellement au niveau du domaine endonucléase par ingénierie, est capable de s'associer pour donner une intéine fonctionnelle[183-186] (Figure 23B). L'assemblage se produit spontanément *in vivo* quand les fragments sont exprimés dans la même cellule. Cependant, *in vitro*, l'épissage en *trans* nécessite souvent la dénaturation et la renaturation des fragments du précurseur. Peu après ces observations, H. Wu[187] a rapporté l'existence d'une intéine naturellement fragmentée dans la protéine DnaE de *Synechocystis* sp. PCC6803. Cette intéine a été la plus utilisée car ses deux fragments présentent une grande stabilité, et s'assemblent spontanément *in vitro*[188]. Par rapport à la technologie EPL, l'épissage en *trans* offre l'avantage de pouvoir réaliser la ligation avec des concentrations en fragments plus faible car la réaction est conduite par la forte affinité des fragments de l'intéine et non pas par la loi d'action des masses. Par contre, l'inconvénient est que la réaction est limitée par les préférences au niveau du substrat de l'intéine fragmentée.

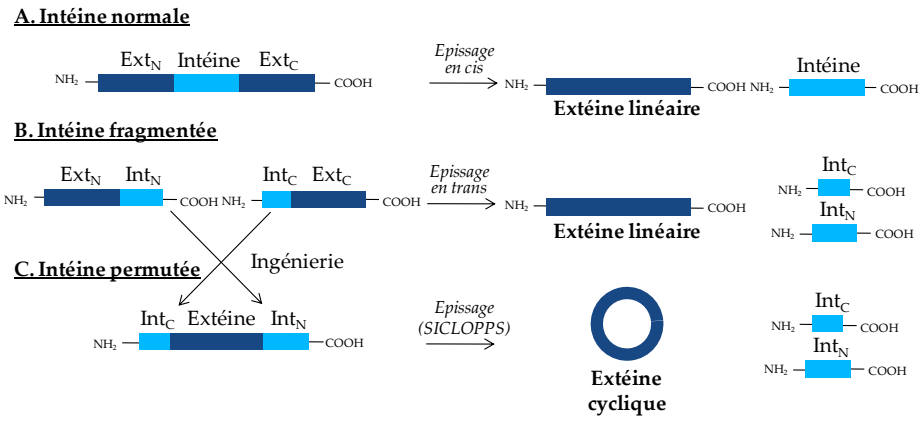


Figure 23 : Schéma de la réaction d'épissage catalysée par une intéine sous la forme A) normale, B) fragmentée et C) permutée.

Tout comme l'EPL, la technologie de l'épissage en *trans* a été utilisée pour la semi-synthèse *in vivo* et *in vitro* d'un grand nombre de protéines. Les exemples principaux sont repris dans la revue de B. Perler[189]. L'épissage en *trans* a également été combiné avec une molécule qui induit la dimérisation pour convertir l'intéine en interrupteur moléculaire *in vivo*[190]. L'intéine Sce VMA a été

fragmentée pour donner deux fragments de faible affinité et chaque fragment a été fusionné soit à la protéine FKBP soit à FRB. FKBP et FRB dimérisent en présence de rapamycine, ce qui enclenche l'épissage *en trans*. Cette application permet d'exercer un contrôle temporel et dépendant de la dose sur la fonction d'une protéine et permet donc d'étudier les effets de l'activation d'une protéine dans une variété de contextes. Une autre application utilisant l'épissage en *trans* est l'immobilisation des protéines sur un support solide[191]. Dans cet exemple, l'extrémité C-terminale du fragment I_C est attachée à la plateforme protéomique, tandis que la protéine à immobiliser est fusionnée à l'extrémité N-terminale du fragment I_N. L'épissage en *trans* permet l'attachement précis de la protéine d'intérêt sur le domaine de liaison à la surface solide.

La technologie SICLOPPS

En 1999, S. J. Benkovic[17] a développé une application additionnelle de l'épissage en *trans* qui a mené à l'utilisation des intéines dans la production de protéines et peptides cycliques. Cette méthode appelée SICLOPPS (Split Intein mediated Circular Ligation Of Proteins and PeptideS) est basée sur la fragmentation de l'intéine et le réarrangement des deux fragments I_N et I_C dans un ordre permuté I_C-I_N avec insertion de la séquence cible au centre I_C-séquence cible-I_N (Figure 23C). L'expression dans ce nouvel arrangement moléculaire résulte en la production d'une protéine de fusion qui cyclise spontanément la séquence cible *in vivo*. Pour réaliser la réaction d'épisage qui mène à la cyclisation, les deux domaines de l'intéine doivent être en phase de lecture avec la séquence cible et correctement repliés. De plus, le premier résidu de la séquence cible doit être une sérine, une cystéine ou une thréonine pour participer à la seconde attaque nucléophile dans la réaction d'épissage.

Dans un premier temps, la technologie SICLOPPS a été appliquée en utilisant une intéine naturellement dimérique, l'intéine DnaE de *Synechocystis*. La disponibilité et la connaissance des deux domaines intéiques fragmentés facilite la stratégie de permutation circulaire. Deux protéines et deux peptides ont pu être cyclisés, la dihydrofolate réductase (DHFR)[17] et la protéine de liaison au maltose (MBP)[153], ainsi que la pseudostellarine F (8aa)[17] et le peptide liant la streptavidine HPQ3-6 (12aa)[192]. Ensuite, la stratégie a été étendue aux intéines monomériques fragmentées artificiellement au niveau du site endonucléase.

Deux tentatives à succès ont été rapportées dans la littérature. L'une a permis de cycliser la protéine DnaB[193] et l'autre, la protéine fluorescente verte[194] (GFP) avec l'intéine DnaB de *Synechocystis* et PfuI de *Pyrococcus furiosis* respectivement. Ces expériences ont révélé la grande flexibilité de la technologie SICLOPPS au niveau de la longueur et de la composition de la séquence cible. Elle permet de cycliser *in vivo* des séquences allant de petits peptides à des protéines et enzymes actives. Une contrainte essentielle au niveau de la séquence cible est d'avoir comme premier résidu un nucléophile qui se retrouve dans le produit cyclique[133].

Ces observations ont permis d'envisager une application prometteuse de la technologie SICLOPPS qui est la création de collections de peptides cycliques *in vivo*. Dans ce cas, une séquence nucléotidique dégénérée est insérée entre les deux fragments de l'intéine permutée. Les collections de peptides cycliques peuvent être potentiellement de toutes les tailles et le peptide lui-même peut contenir un nombre illimité de résidus aléatoires. Cependant, la taille des collections est limitée par l'efficacité de transformation dans *E. coli*. En 2001, S. J. Benkovic a créé les premières collections de peptides cycliques dans le cytoplasme d'*E. coli* avec des séquences allant de quatre à neuf acides aminés et a obtenu des rendements compris entre 10^7 et 10^8 transformants[133,195].

Les peptides cycliques représentent un grand potentiel comme modulateurs de fonction protéique[196]. En effet, comme les sites actifs des enzymes et les interfaces protéine-protéine sont composés d'acides aminés, il est raisonnable de penser qu'un peptide inhibiteur spécifique pourrait être identifié pour n'importe quelle poche de liaison ou surface. De plus, les peptides cycliques ont des avantages par rapport à leurs homologues linéaires[197]. Premièrement, la cyclisation supprime leurs extrémités N- et C-terminales, ce qui permet de les stabiliser contre les exoprotéases[198,199], les endoprotéases qui ont un site actif généralement étendu et le catabolisme cellulaire en général. Deuxièmement, la cyclisation génère une structure plus contrainte conformationnellement, ce qui diminue l'entropie et augmente l'affinité de liaison à la cible[200,201]. Finalement, la cyclisation entraîne la perte des deux charges terminales, ce qui pourrait dans certains cas les rendre aptes à traverser les membranes biologiques.

Le potentiel des collections de peptides cycliques biosynthétiques en tant que source d'un grand nombre de ligands actifs a rapidement été exploité avec le

développement de stratégies de sélection génétique. En utilisant un système "bactérien double-hybride", une stratégie permettant la sélection de peptides cycliques modulateurs des interactions protéine-protéine a été développée[202,203]. Cette stratégie a été validée premièrement sur la disruption de l'ATIC homo-dimère[204] et plus récemment sur l'inhibition de l'interaction Gag-TSG101 qui empêche la propagation du virus HIV[205].

Une stratégie permettant un criblage rapide des collections de peptides cycliques *in vivo* a été développée afin d'identifier des inhibiteurs de la protéase ClpXP [206]. Celle-ci utilise le FACS (Fluorescent Activated Cell Sorter) en combinaison avec une GFP fusionnée à son extrémité 3' avec le peptide tmRNA reconnu par la protéase (GFP-tag). Le peptide qui inhibe la protéase protège la GFP-tag de la dégradation et la cellule fluorescente est ensuite sélectionnée par le FACS.

Plus récemment, des peptides cycliques inhibiteurs de la méthyl transférase EcoDam ont été identifiés grâce à une stratégie qui tire profit du lien existant entre l'activité de la méthyl transférase avec celle de la transposase Tn5 de part sa sensibilité à la méthylation[207]. Les peptides cycliques ont été exprimés dans des conditions qui sont défavorables à la transposition par la transposase Tn5 en raison de l'activité de la méthyl transférase EcoDam. Ceux qui inhibent la méthyl transférase EcoDam ont été sélectionnés car ils ont induit l'activité de transposase sélectionnable *in vivo*.

A coté de ces stratégies conçues pour identifier des peptides agissant sur de cibles cytoplasmiques d'*Escherichia coli*, une collection de peptides cycliques a été produite dans des cellules mammifères grâce à une technologie d'expression rétrovirale et un criblage génétique a été développé pour tester leurs fonctionnalités[208].

Les voies de transport des protéines dans le périplasme d'*Escherichia coli*

Introduction

Le transport des protéines à travers les membranes est essentiel dans tous les systèmes biologiques car il permet aux protéines synthétisées dans le cytoplasme d'atteindre le compartiment dans lequel elles exercent leur fonction métabolique ou structurelle. Chez les bactéries, pas moins de 20% des protéines sont transportées en dehors du cytoplasme et traversent la bicouche phospholipidique de la membrane plasmique[209]. Le transport nécessite des machineries complexes pour ne pas compromettre la structure et la fonction des protéines mais également pour ne pas modifier la perméabilité de la membrane. Dans les bactéries gram-négatives comme *E. coli*, deux machineries de translocation distinctes ont été décrites (Figure 24). La première, qui est la plus connue, est la machinerie d'exportation SecYEG[210-213] (Sécrétion) qui permet le transport de protéines dans une conformation non repliée suivant une modèle post- ou co-traductionnel. La seconde, qui a été découverte plus récemment, est la machinerie de translocation Tat[214] ("Twin-arginine translocation") qui permet l'exportation de protéines repliées et associée à ses cofacteurs.

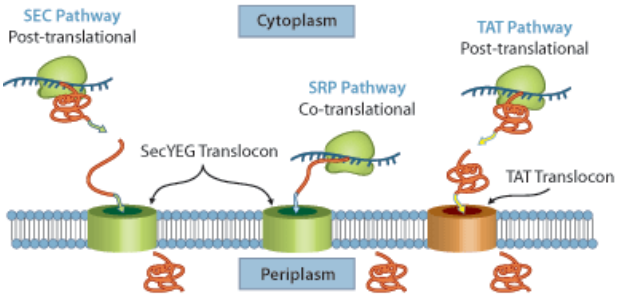


Figure 24 : Représentation schématique des différentes voies de translocation des protéines chez *E. coli* (repris de l'article de S.M Papciak, 2007[215]).

Les différentes voies de translocation coexistent chez *E. coli*. Le choix de la voie à emprunter pour chaque protéine est déterminé par la séquence signal présente en N-terminal de la préprotéine à exporter. Cette séquence est clivée après exportation par une peptidase pour obtenir la protéine mature.

La machinerie de translocation Sec

La machinerie Sec est la voie de translocation la plus utilisée chez les bactéries pour réaliser le transport des protéines à travers la membrane cytoplasmique[216]. Cette machinerie (appelée translocase) est composée principalement d'un complexe hétérotrimérique SecYEG (dénommé translocon) qui forme le canal hydrophilique dans la membrane[217] (Figure 25). Ce complexe SecYEG est très conservé et se retrouve dans les trois domaines du vivant[218-220]. Le cœur du canal de la translocase est formé par deux protéines membranaires SecY et SecE qui ensemble suffisent pour la reconstitution de la translocation *in vitro*[221,222]. Dans la cellule, le complexe SecYE s'associe à une petite protéine membranaire supplémentaire appelée SecG qui n'est pas essentielle mais qui stimule significativement l'activité de translocation[223]. Le complexe SecYEG peut également s'associer à un autre complexe membranaire hétérotrimérique composé des protéines SecD, SecF et YajC[224]. La fonction précise de ces protéines n'est pas encore bien connue, cependant, des modèles ont été proposés dans lesquels SecDF faciliterait l'insertion membranaire de SecA[224-226], qui est la force motrice de la translocation de protéines et interviendrait dans la stabilisation de SecG[227]. En absence de SecDF-YajC, les cellules affichent des sévères défectuosités dans la translocation[228,229]. Une dernière protéine s'associe avec le complexe SecYEG, la protéine YidC dont le rôle est vraisemblablement d'aider la diffusion latérale des fragments transmembranaires dans la bicouche lipidique[230-232].

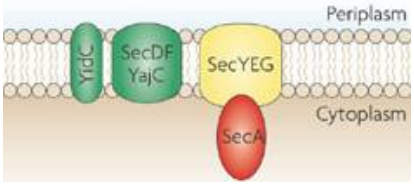


Figure 25 : Représentation schématique de la machinerie d'exportation SecYEG. Elle est composée principalement d'un canal SecYEG et d'un moteur SecA. SecYEG peut s'associer avec des protéines auxiliaires SecDFYajC et YidC (repris de l'article de A. Economou, 2007[233])

Le canal seul est un pore passif : il doit être associé à des partenaires pour lui fournir la force motrice pour la translocation. Chez les bactéries, en fonction du partenaire impliqué, le canal peut fonctionner de deux façon différentes : suivant un modèle co-traductionnel ou post-traductionnel avec comme partenaire le

ribosome ou l'ATPase SecA respectivement[213,234]. Le choix du modèle est déterminé par la séquence signal N-terminale de la préprotéine et intervient lorsque la chaine polypeptidique en croissance commence à émerger du ribosome[235].

Modèle co-traductionnel

La translocation co-traductionnelle est surtout utilisée pour l'intégration de protéines hydrophobes dans la membrane[236] et est généralement peu employée pour l'exportation dans le périplasme ou la membrane externe[237]. Ce modèle fait intervenir une Particule de Reconnaissance du Signal (SRP) dont le rôle est d'amener, durant la traduction, le ribosome et la protéine à exporter au niveau du translocon SecYEG (Figure 26).

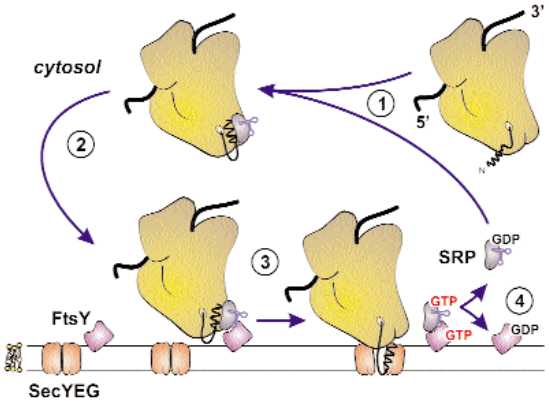


Figure 26 : Modèle co-traductionnel de translocation par la machinerie de translocation SecYEG. La description des étapes est détaillée dans le texte ci-dessous (repris de l'article de C. van de Does, 2001[212]).

Chez *E. coli*, la SRP est composée d'une GTPase appelée Ffh et d'un ARN 4,5 S, qui reconnaissent ensemble et se lient à la séquence signal très hydrophobe de la protéine à exporter alors qu'elle émerge du ribosome[238,239]. Cette interaction entre SRP et le peptide signal retarde la synthèse de la protéine naissante chez les eucaryotes[240] mais, chez les bactéries, l'arrêt de traduction n'a pas encore été prouvé[241]. La SRP amène ensuite le complexe ribosome-chaine naissante au niveau de la membrane cytoplasmique par l'interaction avec son récepteur membranaire FtsY, composé également d'une GTPase[242,243]. Les protéines Ffh et FtsY s'associent en hétérodimère après liaison du GTP et délivrent le complexe

ribosome-chaine naissante au niveau du translocon SecYEG[244]. Après hydrolyse du GTP, la SRP se dissocie de son récepteur FtsY et peut alors reconnaitre et guider une autre protéine vers la membrane[245].

Le complexe ribosome-chaine naissante étant arrivé au niveau de la translocase SecYEG sur la membrane plasmique, l'élongation de la chaine polypeptidique peut se poursuivre en passant directement du tunnel à l'intérieur du ribosome dans le canal membranaire. L'hydrolyse de GTP est nécessaire pour l'élongation de la chaine par le ribosome, mais le mouvement du polypeptide à travers le canal est indépendant de l'hydrolyse nucléotidique[246]. La force motrice est fournie principalement par l'élongation de la chaine bien que la force motrice protonique (PMF) stimule également la vitesse de translocation.

Modèle post-traductionnel

La translocation post-traductionnelle est utilisée essentiellement par les protéines solubles qui possèdent une séquence signal modérément hydrophobe et qui échappe à la reconnaissance par la SRP durant leur synthèse[247,248]. Dans ce cas, c'est la chaperonne cytoplasmique SecB qui intervient[249,250] (Figure 27).

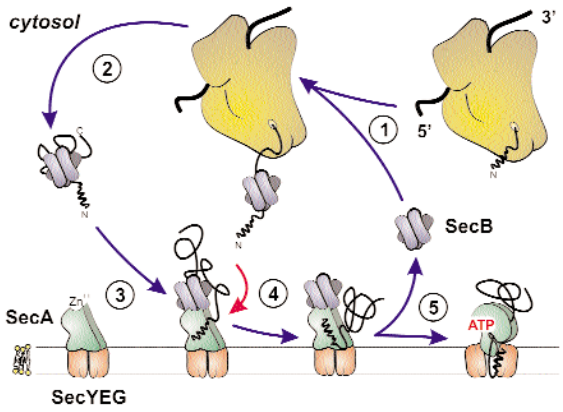


Figure 27 : Modèle post-traductionnel de translocation par la machinerie de translocation SecYEG. La description des étapes est détaillée dans le texte ci-dessous (repris de l'article de C. van de Does, 2001[212]).

SecB n'interagit pas avec la séquence signal mais reconnait des domaines polypeptidiques spécifiques dans la protéine naissante alors qu'elle émerge du ribosome[251]. La liaison à SecB stabilise la préprotéine dans une conformation non

repliée[252] et lui permet d'atteindre la membrane plasmique grâce à la grande affinité de liaison de cette protéine avec l'ATPase SecA liée à la translocase SecYEG[253]. Tandis que la préprotéine est lié à SecB, sa séquence signal crée des interactions avec la protéine SecA[254], ce qui cause une augmentation de l'affinité de liaison SecB-SecA et enclenche la dissociation de la préprotéine de SecB vers SecA Finalement, la liaison de l'ATP à SecA libère SecB pour lui permettre de se lier et guider à une nouvelle protéine naissante[255].

Le transfert du polypeptide à travers le canal SecYEG est initié par l'énergie de liaison de l'ATP et nécessite des cycles de liaison et d'hydrolyse d'ATP par SecA[256-258](Figure 28). En effet, la translocation commence avec la liaison de l'ATP au niveau du site NBF1 de SecA qui cause des changements de conformation et enclenche l'insertion dans le canal de translocation d'une structure en forme d'épingle à cheveux formée par la séquence signal avec le début de la protéine à exporter. Cette étape est stimulée par la force motrice protonique (FMP) qui dirige probablement l'orientation de la séquence signal dans le canal[259,260]. Une fois insérée dans le canal, le substrat est transféré par un mécanisme de "pushing"[225]. Bien que les détails ne sont pas bien connus, un mécanisme plausible a été décrit. Dans ce modèle, SecA lie un segment du substrat et le pousse dans le canal. Le retour en arrière du substrat est réduit par ses interactions avec le canal SecY. Ensuite, après un changement de conformation induit par l'hydrolyse d'ATP, SecA libère le polypeptide et bouge pour se lier au segment suivant du substrat. Le cycle continue jusqu'à ce que le polypeptide entier aie été transloqué.

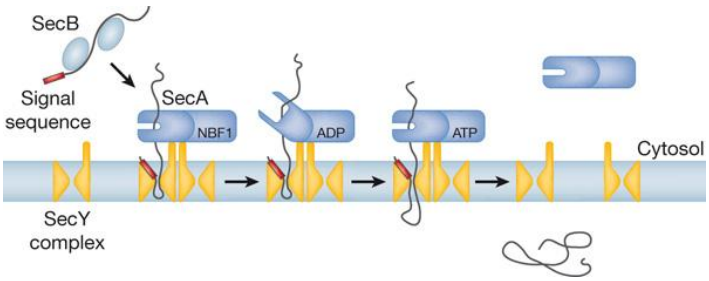


Figure 28 : Représentation schématique du mécanisme proposé pour la translocation catalysée par SecA. La description des étapes est détaillée dans le texte ci-dessus. (repris de l'article de T. Rapoport, 2007[261]).

Sélection du modèle de translocation

La différence principale entre le modèle co- et post-traductionnel est la flexibilité du temps requis pour atteindre le translocon SecYEG. Dans le modèle co-traductionnel, la protéine en train d'être traduite doit être guidée immédiatement vers un translocon vide. Tandis que dans le modèle post-traductionnel, la protéine stabilisée par les chaperonnes SecB peut être transférée vers le translocon plus tardivement. La coexistence des deux modèles de translocation dans *E. coli* est essentielle pour deux raisons[241]. Premièrement, il faudrait un nombre impossible de translocons au niveau de la membrane si la cellule ne disposait pas d'un modèle post-traductionnel permettant de relâcher la pression au niveau du translocon. Deuxièmement, les protéines sujettes à l'agrégation pourraient saturer le système de chaperonnes si la cellule ne possédait pas d'un mécanisme co-traductionnel qui les prend préférentiellement en charge.

La machinerie de translocation Tat

La machinerie de translocation Tat (Twin Arginine Translocation) a été découverte bien après la machinerie Sec. Les premières évidences de l'existence de ce système de transport ont été fournies à la fin des années 80 grâce à des études sur les hydrogénases périplasmiques[262]. Cette machinerie est présente dans la membrane cytoplasmique de la plupart des procaryotes et est utilisée pour le transport des protéines qui lient des cofacteurs et prennent leur "folding" ou même oligomérisent avant d'être transloquées à travers la membrane[263-265]. Les protéines qui empruntent la voie Tat sont reconnues par le motif consensus (SRRXFLK) présent dans leur séquence signal dans lequel se trouvent deux arginines[266].

La translocase Tat d'*E. coli* est composée de trois protéines membranaires TatA, TatB et TatC qui ensemble forment un récepteur et un canal de translocation pour les substrats de Tat[267-270]. La dimension du canal de translocation est variable et dépend de la dimension moléculaire de polypeptide replié à transporter.

Les substrats de Tat utilisent des chaperonnes générales (SlyD ou DnaK) pour se replier dans le cytoplasme, celles-ci interviennent directement à la sortie du tunnel du ribosome[271-273]. En plus d'aider au repliement, ces chaperonnes lient le peptide signal pour l'empêcher de faire des interactions avec la translocase Tat avant l'insertion des cofacteurs et le repliement complet de la protéine à exporter. Une fois que la protéine est prête, la translocation peut démarrer et les chaperonnes libèrent le peptide signal. Celui-ci se lie grâce à son motif "twin-arginine" au complexe TatBC qui est intégré dans la membrane (Figure 29). Le complexe TatBC-substrat s'associe alors avec TatA dans une étape dépendante de la force motrice protonique[274,275]. Cette interaction persiste jusqu'à ce que le substrat ait traversé la membrane et provoque un changement de conformation de TatBC qui pousse le substrat à travers la membrane fragilisée par TatA. Au moment de relâcher le substrat, un canal permettant le flux de protons est formé durant un cours instant. La membrane est ensuite fermée et la séquence signal est clivée par la peptidase[276].

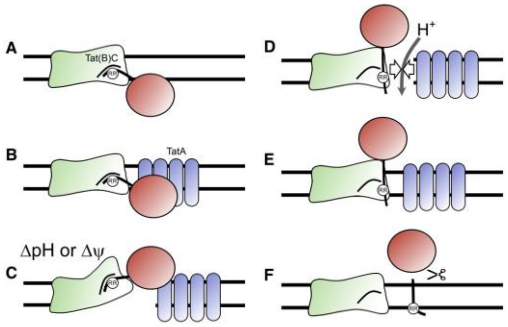


Figure 29 : Modèle de la translocation par la machinerie de translocation Tat. La description des étapes est détaillée dans le texte ci-dessus (repris de l'article de J.M Driessen, 2008[277]).

Comparaison des séquences-signal N-terminales

La translocation des protéines à travers les membranes biologiques requiert l'existence d'une séquence signal N-terminale. Chez *E. coli*, les séquences-signal ont généralement une longueur moyenne de 20 résidus et sont constituées de 3 régions : une région n chargée positivement, suivie d'une région h hydrophobe et finalement une région c polaire contenant le site de clivage (Figure 30). Bien que les séquences-signal soient similaires dans leur structure globale, elles

possèdent des caractéristiques propres qui déterminent la voie de translocation utilisée.

Les séquences-signal de la voie Tat possèdent à l'interface des régions n et h, un motif conservé ZRRxΦΦ où R représente une arginine, Z un résidu polaire et Φ un résidu hydrophobe. Les séquences signal de la voie Sec n'ont pas de motif conservé mais les régions n et h forment des éléments structuraux reconnus par SRP[278,279] (modèle co-traductionnel) et SecA[280] (modèle post-traductionnel). En effet, l'affinité de liaison de SRP augmente avec l'hydrophobicité de la région h, tandis que l'interaction avec SecA augmente avec le nombre de résidus chargés dans la région n. Les régions n des séquences signal de la voie Tat sont généralement plus longues que celles de la voie Sec. Comme ces régions n contiennent les résidus chargés suspectés d'interagir avec la membrane, cela suggère que les séquences signal Tat formeraient plus d'interactions électrostatiques avec la membrane phospholipidique. De plus, les régions h des séquences signal Tat sont en moyenne légèrement plus courtes et moins hydrophobes que celles de Sec. Dans la région c, les résidus en position -3 et -1 par rapport au site de clivage doivent être petits et neutres pour que le clivage se produise[281-283]. On retrouve donc dans cette région le motif "Ala-X-Ala".

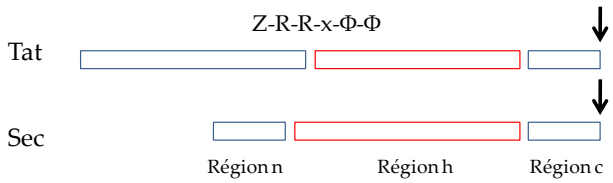


Figure 30 : Comparaison schématique des séquences signal N-terminales des substrats des deux machineries de translocation Sec et Tat. On distingue 3 régions, une région amino-terminale chargée positivement (région n), une région non chargée hydrophobe (région h) et une petite région carboxy-terminale polaire qui contient un site de clivage reconnu par une peptidase (région c).

Le clivage des séquences signal est réalisé par des enzymes appelées peptidases signal de type I (SPases). Elles appartiennent à une classe de protéases à Serine qui réalisent la catalyse en utilisant une dyade sérine/lysine au lieu d'une triade serine/histidine/lysine comme dans la plupart des protéases à Serine[284]. La SPaseI appelée LepI est responsable du clivage des séquences signal d'un grand nombre de protéines dans *E. coli*. Cette enzyme est une endopeptidase liée à la membrane avec une région catalytique qui réside dans le périplasme[285].

II. Résultats et discussion

Régulation de l'expression basale des gènes toxiques

Introduction

Le contrôle de l'expression basale des gènes dans *Escherichia coli* est particulièrement intéressant dans le contexte de la sélection de nouveaux composés antibactériens au sein de collections de peptides synthétisés *in vivo*. En effet, plus la toxicité bactérienne est élevée, plus le système de répression de l'expression doit être maximal car les peptides les plus toxiques, qui sont sans doute les plus intéressants, pourraient rapidement disparaître des collections par sélection *in vivo*. Dans un premier temps, et pour éviter la mort bactérienne, le système doit assurer la répression totale de l'expression. Cette répression totale est cependant très difficile à obtenir avec les vecteurs et les stratégies généralement utilisés, car une expression basale non désirée, dont l'origine peut être multiple est généralement toujours présente. Ensuite, et pour permettre la production d'une certaine quantité de peptide, l'expression du gène doit pouvoir être enclenchée au moment désiré.

Le système que nous souhaitons développer, et qui répond à ces deux exigences, est basé sur la fragmentation du gène toxique. La fragmentation n'est pas aléatoire mais est effectuée de façon à retrouver une séquence commune entre les deux fragments (Figure 31). Cette séquence permettra la reconstitution du gène par recombinaison homologue lorsque les deux fragments seront présents simultanément dans une cellule bactérienne.

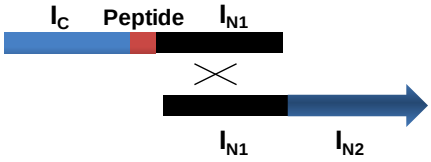


Figure 31 : Représentation schématique de la stratégie de répression totale de l'expression basale par fragmentation du gène. L'exemple est réalisé sur la construction SICLOPPS.

Un fragment sera cloné dans un plasmide, tandis que l'autre fragment sera introduit dans un phage filamenteux ou un phagemide. Le plasmide contenant la première partie du gène toxique sera électroporé dans *E. coli* et produira une protéine fragmentée non toxique. En présence d'une pression de sélection permettant le maintien du plasmide, les cellules transformées pourront donc se

développer normalement. L'induction de l'expression du gène toxique sera enclenchée au moment désiré, c'est-à-dire au moment où la culture aura atteint un nombre de cellules jugé suffisant. Cela se produira grâce à la reconstitution du gène par recombinaison homologue après infection par le phage ou le phagemide contenant le 2ème fragment du gène toxique.

Notre laboratoire s'est intéressé à cette stratégie de reconstitution de gènes par recombinaison homologue impliquant les phages filamenteux à ADN simple brin, suite à l'observation inattendue de recombinaison homologue avec un plasmide[286]. Dans cette expérience, deux vecteurs avaient été conçus de façon à présenter des fragments du gène de la β-lactamase qui se recouvrent partiellement (Figure 32). Une fois dans *E. coli*, cette région de recouvrement de 132 nucléotides a servi de séquence homologue pour la recombinaison et a permis la reconstitution du gène par formation du cointégrat.

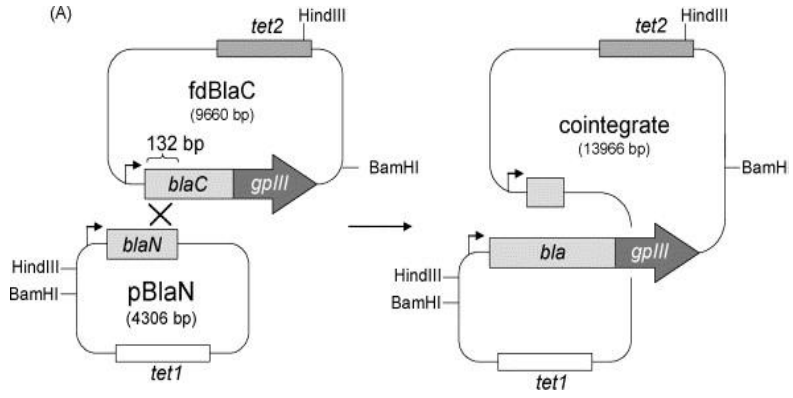


Figure 32 : Design de la recombinaison homologue entre le vecteur fdBlaC et pBlaN. Chaque vecteur contient un allèle non homologue du gène de résistance à la tétracycline (T_C). Ensemble, les deux vecteurs confèrent une résistance faible et thermosensible à l'ampicilline (20 µg/ml à 23°C). La recombinaison homologue dans la région de 132 pb partagée par les deux vecteurs restore un gène *bla* intacte. Les cellules peuvent être alors sélectionnées pour leur haute résistance à l'ampicilline (50 µg/ml) à 37°C (repris de l'article de P. Soumillion, 2006[286])

Dans des conditions où plus de 90% des bactéries sont infectées, des fréquences de recombinaison de 2.10^{-4} et 2.10^{-2} ont été mesurées pour les souches JM109 ($recA^-$) et TG1 ($recA^+$) respectivement. Pour une séquence d'homologie similaire, ces fréquences sont 1000 fois supérieures à celles obtenues lors de la reconstitution du gène de résistance à la tétracycline par recombinaison homologue entre deux plasmides[82]. Ainsi, une augmentation de l'efficacité de la recombinaison a été constatée dans les deux voies dépendantes ou indépendantes de recA, lorsqu'un vecteur phagique est impliqué. Comme recA

initie la recombinaison en liant l'ADN simple brin, la haute efficacité est due probablement à la haute disponibilité de l'ADN simple brin durant la réplication et / ou l'infection des phages.

Pour que cette stratégie de fragmentation et de reconstitution de gènes puisse être appliquée au contrôle de l'expression des collections de peptides cycliques, le système devrait être amélioré pour atteindre des fréquences de recombinaison supérieures. Dans les conditions optimales, il faudrait pouvoir approcher des taux de 100% afin que l'ensemble de la collection soit exprimée et représentée. Cependant, même avec des taux de recombinaison de 2%, cette stratégie pourrait avoir des applications biotechnologiques comme la création de banques combinatoires de grande taille. En effet, l'augmentation de la diversité d'une banque est corrélée à l'augmentation des chances de trouver des clones intéressants. Généralement, la taille des banques est limitée par l'étape de transformation des bactéries et la création de collections de plus de 10^8-10^9 clones est très laborieuse. Grâce à la fragmentation, deux banques partielles pourraient être créées. Si chaque collection contient approximativement 10^6-10^7 clones, la banque combinatoire potentielle résultant de la recombinaison homologue contiendrait 10^{10}-10^{12} clones différents. Dans ce cas, la diversité serait donc limitée par le nombre de bactéries qui peuvent être infectées par un phage.

Objectif

Dans le cadre de la thèse, l'objectif a donc été d'évaluer le taux de recombinaison homologue que l'on peut obtenir avec un système composé de deux nouveaux vecteurs dans lequel des améliorations ont été apportées par rapport au système initialement utilisé et étudié.

Premièrement, la séquence d'homologie a été augmentée pour atteindre 213 nucléotides au lieu de 132 et contrairement au système précédent, cette séquence représente l'unique région homologue entre les deux vecteurs. De plus, la recombinaison homologue a été combinée à la recombinaison site spécifique pour résoudre le cointégrat en deux vecteurs séparés et stables (Figure 33). La suite des deux recombinaisons résulte en l'échange du fragment B sur le plasmide avec le fragment D du phagemide. Finalement, les vecteurs ont été construits de façon à contenir un gène de résistance à un antibiotique différent et des origines de réplication compatibles.

Dans ce nouveau système, la fragmentation a été réalisée au niveau du précurseur DnaBI$_C$-peptide-DnaBI$_N$ en prenant comme séquence d'homologie un fragment du gène DnaBI$_N$ (Figure 34). La réaction de recombinaison homologue restaure donc la construction génétique permettant l'expression et la cyclisation des peptides.

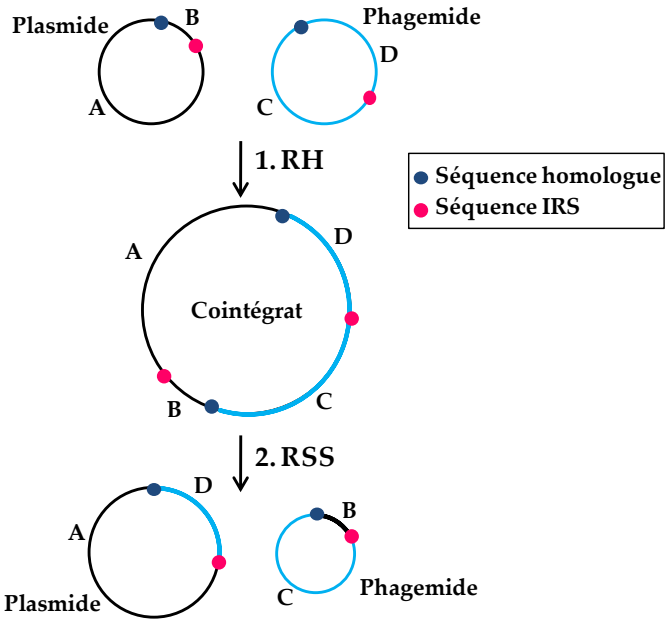


Figure 33 : Représentation schématique du système combinant la recombinaison homologue (RH) avec la recombinaison site spécifique (RSS). La suite des deux recombinaisons résulte en l'échange du fragment B sur le plasmide avec le fragment D du phagemide.

Résultats

Création du système de reconstitution de gène par recombinaison double

Le système est composé d'un plasmide dérivé du vecteur pAP02 nommé p15I$_c$PX et d'un phagemide dérivé du vecteur pHDiex appelé pColYI$_N$ (Figure 34). Ces deux vecteurs possèdent des marqueurs de résistances différents (Tet^R et Kan^R respectivement) et des origines de réplication compatibles (respectivement p15A et ColE1/origine f1). Le plasmide contient la séquence encodant DnaBI$_C$, la collection de peptides et un fragment de DnaBI$_N$ (résidus 1-71). Le phagemide possède la séquence encodant DnaBI$_N$ (résidus 1-106). Les deux vecteurs partagent une séquence identique de 213 nucléotides au niveau de DnaBI$_N$. Afin

de permettre la recombinaison site spécifique, une séquence de recombinaison interne du transposon Tn4430 (IRS) a été insérée sur les deux vecteurs. De plus, le gène encodant la recombinase à tyrosine TnpI a été ajoutée sur le plasmide en aval du promoteur pLac.

Initialement, la recombinaison homologue (RH) va restaurer la construction DnaBI$_C$-peptide-DnaBI$_N$ et permettre la cyclisation des peptides. La recombinaison site spécifique (RSS) va ensuite immédiatement résoudre le cointégrat en deux vecteurs séparés et stables. Sur le phagmide, le terminateur de transcription T7 a été ajouté en amont de DnaBI$_N$ et un gène de résistance au chloramphénicol (Cm^R) en aval. Le terminateur T7 devrait permettre d'empêcher l'expression des gènes avant recombinaison. Tandis que le gène de résistance au chloramphénicol devrait permettre de sélectionner *in vivo* le plasmide résultant de la double recombinaison.

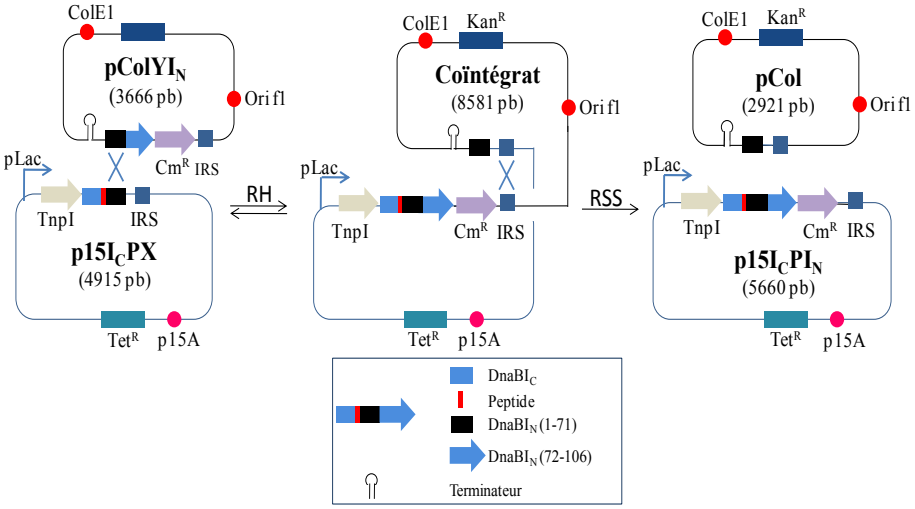


Figure 34 : Présentation du système permettant la reconstitution de gène par la recombinaison double. Le plasmide p15I$_C$PX contient, sous le contrôle de pLac, le gène de la recombinase à tyrosine TnpI, l'intéine DnaBI$_C$, le peptide et le fragment du gène DnaBI$_N$. Ce vecteur confère la résistance à la tétracycline et possède l'origine de réplication p15A. Sur le phagemide pColYI$_N$, en aval d'un terminateur de transcription, ont été clonés le gène de l'intéine DnaBI$_N$ et Cm^R (résistance au chloramphénicol). Ce vecteur confère la résistance à la kanamycine et possède deux origines de réplication, une simple brin (f1 ori) et une double brin (ColE1). Une séquence IRS (séquence de recombinaison interne) a été ajoutée sur les deux vecteurs.

Stratégie pour la construction du plasmide p15I$_C$PX

Le plasmide p15I$_C$PX a été construit en assemblant et clonant des séquences d'ADN provenant de différents vecteurs dans le plasmide pAP02 (Figure 35).

Le promoteur pLac et la recombinase à tyrosine TnpI ont été extraits du vecteur pGIV007, reçu du Prof. B. Hallet, en réalisant une PCR avec les amorces TnpI-fp et TnpI-rp (Tableau 2). De la même façon, l'intéine permutée DnaBI$_C$-DnaBI$_N$ et fragmentée au niveau de la thréonine 70 de DnabI$_N$ a été extraite du vecteur pBI, construit par J.-M. Léonard, en réalisant une PCR avec les amorces Int-fp et Int-rp (Tableau 2). L'amorce TnpI-fp contient à son extrémité le site de restriction ScaI nécessaire pour le clonage ultérieur du fragment. Les amorces TnpI-rp et Int-fp possèdent une séquence complémentaire qui a été utilisée pour assembler les deux fragments par PCR de recouvrement. Cette séquence complémentaire contient un site de fixation pour le ribosome (RBS) afin de permettre la traduction du gène encodant le précurseur intéine-peptide qui suit celui de la TnpI. Le fragment obtenu après la PCR de recouvrement a ensuite été cloné dans le vecteur pGEM-T (Promega). La séquence IRS a été extraite du vecteur pGIV03, reçu du Prof. B. Hallet, par restriction avec les enzymes SalI et PciI et ensuite cloné dans le vecteur pGEM-TA1 entre ces mêmes sites de restriction. Finalement, l'insert présent sur le vecteur pGEM-TA2 entre les sites de restriction ScaI et BamHI a été cloné dans le vecteur pAP02 entre les sites de restriction ScaI et BglII pour donner le plasmide p15I$_C$PX. En effet, les sites BamHI et BglII sont des sites de restriction à bouts collants compatibles.

La construction correcte du plasmide final a été vérifiée par analyse du profil de restriction sur gel d'agarose et par séquençage.

Amorces	**Séquences**
TnpI-fp	5'-TAAGTACTGCTAGCGCACGACAGGTTTCCCGACTG-3'
TnpI-rp	5'-TTCCTCCTGTTAGCCCAATTATAAGAGCTCCATTTT-3'
Int-fp	5'-TTATAATTGGGCTAACAGGAGGAATTAACCATGGGA-3'
Int-rp	5'-TGTTGCCTTGATAGTTCT-3'

Tableau 2 : Séquences des amorces pour la stratégie de construction du plasmide p15I$_C$PX. Le site de restriction ScaI est souligné dans l'amorce TnpI-fp. Sur les amorces TnpI-rp et Int-fp, la séquence complémentaire est en bleu. La séquence RBS (AGGAGG) se situe entre le codon stop et start (en rouge).

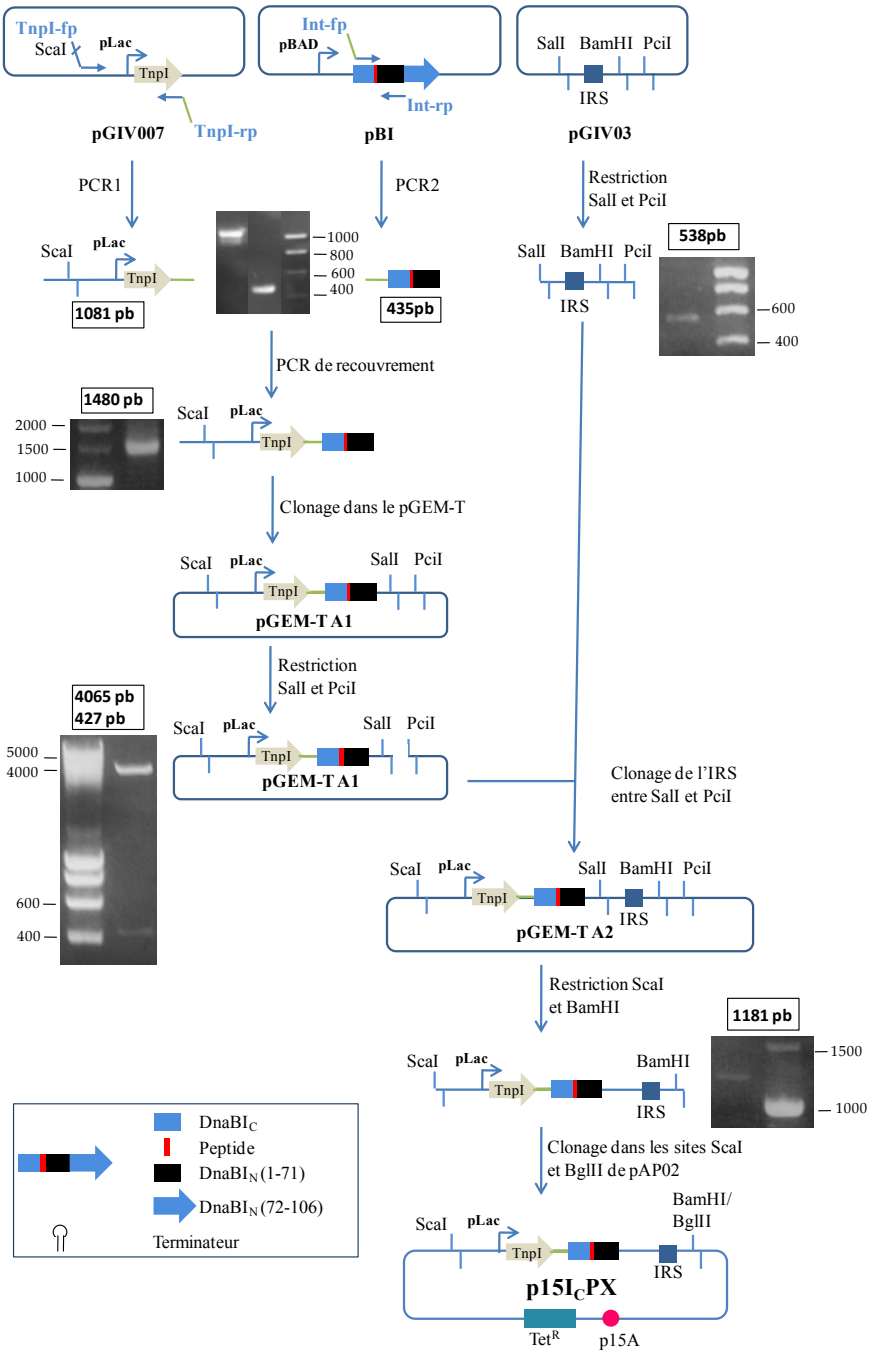


Figure 35 : Stratégie pour la construction du plasmide p15I$_C$PX. . Le promoteur pLac et la TnpI se trouve sur le vecteur pGIV007, l'intéine permutée sur pBADI et l'IRS sur pGIV03. 1. PCR sur le plasmide pGIV007. 2. PCR sur le plasmide pBADI. 3. PCR de recouvrement. 4. Clonage du fragment PCR dans le pGEM-T. 5. Clonage de l'IRS dans le pGEM-T A1. 6. Clonage dans le vecteur pAP02 pour donner p15I$_C$PX .

Stratégie pour la construction du phagemide pColYI$_N$

La stratégie utilisée pour construire le phagemide pColYI$_N$ est fort similaire à celle développée pour le plasmide p15I$_N$PX (Figure 36).

Des séquences provenant de 3 vecteurs différents ont été assemblées et ensuite clonées dans le phagemide pHDiex. Tout d'abord, la séquence encodant l'intéine DnaBI$_N$ a été extraite du plasmide pBI en réalisant une PCR avec les amorces Int2-fp et Int2-rp (Tableau 3). De la même façon, le gène de résistance au chloramphénicol (CmR) a été extrait du plasmide pAP02 en réalisant une PCR avec les amorces Chlo-fp et Chlo-rp (Tableau 3). L'amorce Int2-fp contient à son extrémité le site de restriction MfeI nécessaire pour le clonage ultérieur du fragment. Les amorces Int2-rp et Chlo-fp possèdent une séquence complémentaire qui a été utilisée pour assembler les deux fragments par PCR de recouvrement. Cette séquence complémentaire contient un site de fixation pour le ribosome (RBS) qui permet la traduction du gène de résistance au chloramphénicol. Le fragment obtenu après la PCR de recouvrement a ensuite été cloné dans le vecteur pGEM-T (promega). La séquence IRS a été extraite du vecteur pGIV03, par restriction avec les enzymes XbaI et PciI et ensuite cloné dans le vecteur pGEM-TH1 entre les sites de restriction SpeI et PciI., les sites XbaI et SpeI ayant des bouts collants compatibles. Finalement, l'insert présent sur le vecteur pGEM-TH2 entre les sites de restriction MfeI et EcoRI a été cloné dans le vecteur pHDiex dans le site de restriction unique EcoRI qui se trouve deux nucléotides en aval du terminateur T7 déjà présent sur le vecteur. Cela est possible car les sites MfeI et EcoRI ont des bouts collants compatibles.

La construction correcte du phagemide final pColYI$_N$ a été vérifiée par analyse du profil de restriction sur gel d'agarose et par séquençage.

Amorces	Séquences
Int2-fp	5'-CAATTGGGATGCATCAGTGGAGATAGT-3'
Int2-rp	5'-CTCCATTTTAGCTTCCTTCTATAATTGTAAAGAGGAGCT-3'
Chlo-fp	5'-TTATAGAAGGAAGCTAAAATGGAGAAA-3'
Chlo-rp	5'-TTACGCCCCGCCCTGCCACTC-3'

Tableau 3 : Séquences des amorces pour la stratégie de construction du plasmide pColYI$_N$. Le site de restriction MfeI est souligné dans l'amorce Int2-fp. Sur les amorces Int2-rp et Chlo-fp, la séquence complémentaire est en bleu. La séquence RBS se situe entre le codon stop et start (en rouge).

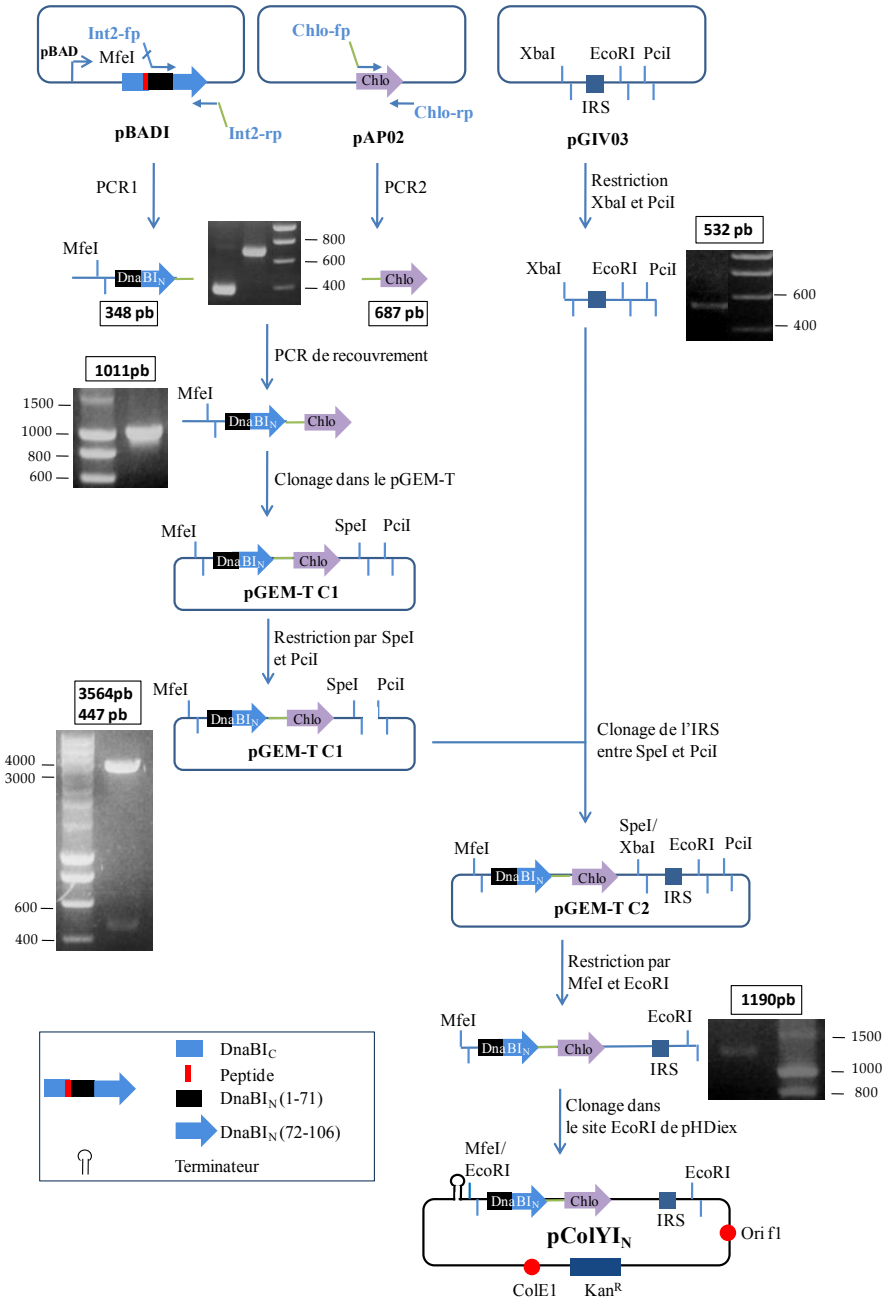


Figure 36 : Stratégie pour la construction du phagemide pColYI$_N$. 1. PCR sur le plasmide pBADI. 2. PCR sur le plasmide pAP02. 3. PCR de recouvrement. 4. Clonage du fragment PCR dans le pGEM-T. 5. Clonage de l'IRS dans le pGEM-T H1. 6. Clonage dans le vecteur pHDiex pour donner le phagemide pColYI$_N$.

Vérification de la sensibilité au chloramphénicol des cellules TG1- pColYI$_N$

La sensibilité au chloramphénicol des cellules TG1 transformées par le vecteur pColYI$_N$ (TG1-(pColYI$_N$)) a été vérifiée pour s'assurer que ce phénotype peut être utilisé pour sélectionner les événements de recombinaison. Les bactéries se sont malheureusement avérées capables de pousser à des concentrations en chloramphénicol allant jusqu'à 100 µg/ml sans perturbation de croissance. Pour les bactéries TG1 non transformées, une concentration minimale inhibitrice (MIC) de 10 µg/ml a été mesurée. Cela indique que, malgré la présence du terminateur T7 en amont du gène, une expression basale permet de produire suffisamment d'acétyltransférase pour conférer une haute résistance à l'antibiotique. Après analyse des séquences, il semble qu'une transcription soit initiée à partir d'un promoteur cryptique présent dans l'intéine DnaBI$_N$ en amont du gène.

Modification des vecteurs pour déplacer la séquence d'homologie au niveau du gène de résistance au chloramphénicol

Afin de pouvoir évaluer la fréquence de recombinaison en utilisant le phénotype de résistance au chloramphénicol, les deux vecteurs ont été modifiés de façon à déplacer la séquence d'homologie à l'intérieur du gène de résistance au chloramphénicol lui-même.

Le plasmide p15CX a été obtenu par une stratégie en deux étapes (Figure 37). Dans la première étape, la séquence présente entre les sites NsiI et NotI sur le vecteur pGEM-T H2 a été clonée dans le vecteur p15I$_N$PX entre ces mêmes sites de restriction. Dans la seconde étape, un fragment de 239 pb entre les sites BsmBI et NotI a été enlevé par restriction. Finalement, les bouts collants obtenus ont été polymérisés et ligués pour donner le plasmide final dans lequel le gène de résistance au chloramphénicol a été tronqué de sa partie C-terminale (résidus 146-220).

La construction du phagemide pColYm s'est effectuée en une seule étape (Figure 38). Le vecteur pColYI$_N$ a été restreint aux niveaux des sites de restriction BglII et BspEI. Les bouts collants obtenus ont été polymérisés et ligués pour donner le vecteur final dans lequel le gène de résistance au chloramphénicol a été tronqué de sa partie N-terminale (résidus 1-70). Les deux vecteurs partagent une

séquence identique de 228 nucléotides (résidus 70-146) à l'intérieur de leur fragment du gène de résistance au chloramphénicol.

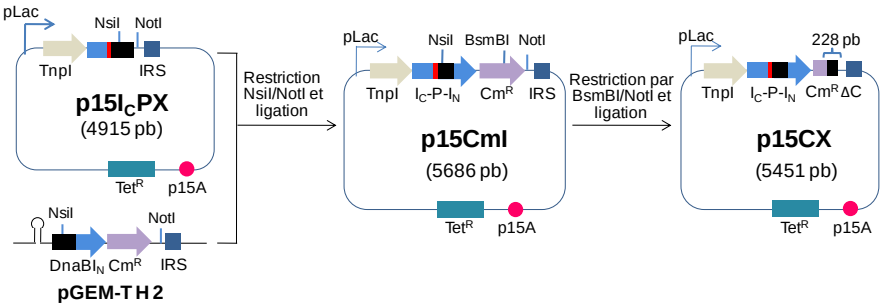


Figure 37 : Stratégie pour la construction du vecteur p15CX dans lequel la séquence d'homologie se trouve dans le gène de résistance au chloramphénicol. Dans la première étape, la séquence présente entre les sites NsiI et NotI sur le pGEM-T H2 a été clonée dans le vecteur p15I$_N$PX. Dans la seconde étape, un fragment du gène de résistance au chloramphénicol a été enlevé par restriction avec les enzymes BsmBI et NotI. Le vecteur a été finalement polymérisé et ligué pour donner le plasmide p15CX

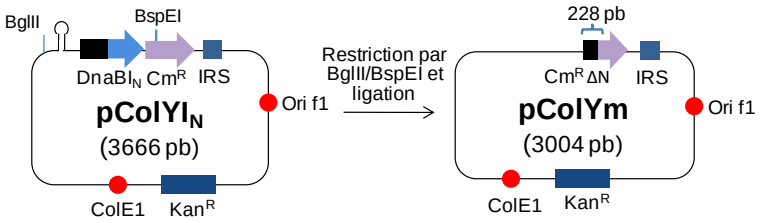


Figure 38 : Stratégie pour la construction du phagemide pColYm dans lequel la séquence d'homologie se trouve dans le gène de résistance au chloramphénicol. Le fragment d'ADN compris entre les sites de restriction BglII et BspEI dans le vecteur pColYI$_N$ est enlevé par restriction. Le vecteur est ensuite polymérisé et ligué pour donner le vecteur pColYm.

Les bactéries TG1-(p15CX) et TG1-(pColYm) ont une sensibilité au chloramphénicol équivalente aux cellules TG1 non transformées avec un MIC de 5 µg/ml. La double recombinaison entre les deux vecteurs donnera le plasmide doublement recombiné p15Cm avec le gène de résistance au chloramphénicol reconstitué qui permettra aux bactéries d'augmenter leur résistance à l'antibiotique (Figure 39). Le plasmide p15CmI, obtenu lors de la construction de p15CX (Figure 37), est identique au plasmide doublement recombiné p15Cm (Figure 39). Ce vecteur a servi de contrôle pour prouver que le gène Cm^R dans cette construction est capable de conférer aux bactéries une résistance au chloramphénicol allant jusqu'à 100 µg/ml.

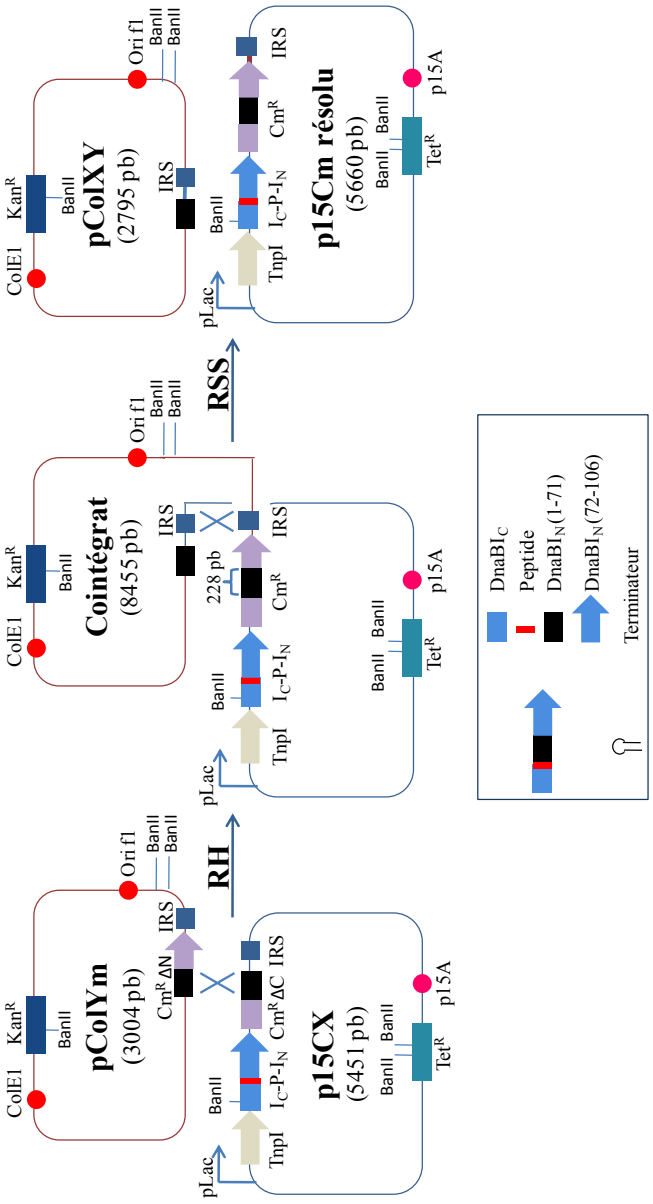


Figure 39 : Présentation du système permettant la reconstitution de gène de résistance au chloramphénicol par la recombinaison double. Le plasmide p15I$_C$X contient le gène de résistance au chloramphénicol tronqué de sa partie C-terminale (résidus 146-220). Le phagemide pColYm contient le gène de résistance au chloramphénicol tronqué de sa partie N-terminale (résidus 1-70). Les deux vecteurs partagent une séquence identique de 228 nucléotides (résidus 70-146).

Evaluation du taux de recombinaison

Des cultures de cellules TG1-(p15CX) ont été infectées par les phagemides pColYm avec une multiplicité d'infection allant de 10 à 10^5. Après 1h d'incubation sans agitation à 37°C et des temps croissants d'incubation allant de 2h à 20h avec une agitation de 180 rpm à 37°C, des étalements ont été réalisés sur des boîtes de Pétri tet7,5/kan50 et tet7,5/kan50/cm15. Les colonies observées sur les boîtes de Pétri avec chloramphénicol ont des tailles fort variables et les plus petites mettent 36h pour pousser. Sans chloramphénicol, les colonies ont des tailles homogènes et mettent 12h pour pousser. La fréquence de recombinaison a été estimée en mesurant le rapport du nombre de colonies poussant sur les boîtes de Pétri avec ou sans chloramphénicol après 36h. Les résultats obtenus sont présentés dans le Tableau 4 ci-dessous.

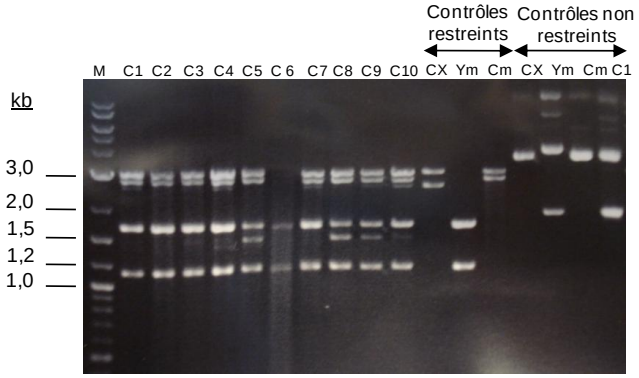
		Temps d'incubation					
Multiplicité d'infection		**2h**	**3h**	**4h**	**5h**	**6h**	**20h**
	10	$2,5\ 10^{-5}$	$1,2\ 10^{-5}$	$1,1\ 10^{-4}$	$1,4\ 10^{-4}$	$5,5\ 10^{-5}$	$1,0\ 10^{-4}$
	10^2	$< 1,0\ 10^{-5}$	$3,2\ 10^{-4}$	$5,2\ 10^{-5}$	$5,0\ 10^{-4}$	$3,0\ 10^{-4}$	$1,0\ 10^{-4}$
	10^3	ND	$2,4\ 10^{-4}$	$3,0\ 10^{-4}$	$3,3\ 10^{-4}$	$9,0\ 10^{-5}$	$1,2\ 10^{-4}$
	10^4	ND	ND	$2,8\ 10^{-4}$	$8,3\ 10^{-5}$	$6,0\ 10^{-4}$	$1,9\ 10^{-4}$
	10^5	ND	ND	ND	$1,7\ 10^{-4}$	$5,0\ 10^{-4}$	$1,9\ 10^{-4}$

Tableau 4 : Présentation des fréquences de recombinaison mesurées après infection des cellules TG1-p15CX par les phagemides pColYm pour des multiplicités d'infection allant de 10 à 10^5 et des temps d'incubation de 2h à 20h. ND : Non Déterminé.

Les taux de recombinaison mesurés sont faibles et compris entre $1,2\ 10^{-5}$ et $6,0\ 10^{-4}$. Ils varient de manière aléatoire et n'augmentent pas en fonction du temps d'incubation ou de la multiplicité d'infection. De plus, lorsque des phages auxiliaires sont ajoutés dans la culture infectée, le taux n'augmente pas, mais reste compris dans les valeurs initialement observées (résultats non montrés). L'hypothèse était que l'ajout de ces phages pourrait augmenter le taux de recombinaison en enclenchant la réplication simple brin du phagemide pColYm. L'expérience pour un temps d'incubation de 3h et une multiplicité de 10 a été répétée 8 fois, un taux de recombinaison moyen de $1,0\ 10^{-4}$ a été mesuré avec un écart type de $7,8\ 10^{-5}$. Par conséquent, cela indique que la manipulation est peu reproductible.

Des analyses sur gel d'agarose des restrictions par BanII réalisées sur l'ADN isolé de 10 colonies poussant sur chloramphénicol à 37°C ont permis de détecter le plasmide et le phagemide résolu (Figure 40). Les bandes à 2711 pb et 1459 pb,

caractérisant respectivement les deux produits de double recombinaison p15Cm et pColXY sont présentes sur le gel pour la majorité des clones. Les différences de concentration des vecteurs dans les cellules peuvent être expliquées par le nombre de copies différents entre les deux vecteurs, le phénomène d'incompatibilité des vecteurs et la pression de sélection liée à la présence de chloramphénicol dans le milieu. Les séquençages sur ces deux vecteurs ont permis de confirmer les résultats.



	BanII
p15CX (CX)	14-2935-2502
p15Cm (Cm)	14-2935-2711
pColYm (Ym)	215-1121-1668
pColXY (XY)	215-1121-1459

Figure 40 : Gel d'agarose 1% présentant le profil de restriction par BanII de 10 clones poussant sur chloramphénicol à 37°c, ainsi que les contrôles restreints et non restreints. Puit 1 : 2-Log DNA Ladder (0.1–10.0 kb), New England Biolabs

Comme la sélection des événements de recombinaison par le phénotype de résistance au chloramphénicol pourrait sous estimer le taux de recombinaison du à la mort cellulaire, des extractions d'ADN plasmidiques ont été réalisées sur des cellules infectées par les phagemides sans réaliser de sélection sur chloramphénicol. Le profil de restriction de l'ADN par BanII a ensuite été analysé sur gel d'agarose 1%. Les bandes correspondant aux produits de double recombinaison p15Cm et pColXY n'ont pas pu être observées indiquant que le taux de recombinaison est trop faible pour être détecté sur gel (résultats non montrés).

L'ADN du phagemide pColYm a été électroporé dans des cellules TG1-(p15CX) de façon à ce que les deux vecteurs se comportent en plasmide. Des étalements

sur boîtes de Pétri avec et sans chloramphénicol à une concentration de 15µg/ml ont été réalisés après des temps d'incubation à 37°C et 180 rpm allant de 5min à 20h. Pour cette expérience et contrairement à celle précédente avec infection, les colonies observées sur les boîtes de Pétri avec chloramphénicol ont des tailles homogènes et poussent en une nuit. Les résultats de l'expérience répétée 4 fois sont présentés dans le Tableau 5 ci-dessous.

	5min	30min	1h	3h	5h	8h	20h
Expérience 1	$< 1{,}0\ 10^{-7}$	$8{,}6\ 10^{-7}$	$4{,}4\ 10^{-6}$	$5{,}9\ 10^{-5}$	$6{,}3\ 10^{-5}$	$6{,}6\ 10^{-4}$	$4{,}8\ 10^{-4}$
Expérience 2	$< 1{,}0\ 10^{-7}$	$< 1{,}0\ 10^{-7}$	$4{,}4\ 10^{-7}$	$7{,}4\ 10^{-6}$	$4{,}7\ 10^{-6}$	$1{,}1\ 10^{-4}$	$1{,}4\ 10^{-4}$
Expérience 3	$< 1{,}0\ 10^{-7}$	$< 1{,}0\ 10^{-7}$	$4{,}3\ 10^{-6}$	$7{,}0\ 10^{-5}$	$7{,}9\ 10^{-5}$	$2{,}7\ 10^{-5}$	$1{,}3\ 10^{-4}$
Expérience 4	$< 1{,}0\ 10^{-7}$	$2{,}5\ 10^{-6}$	$1{,}2\ 10^{-5}$	$1{,}1\ 10^{-4}$	$1{,}5\ 10^{-4}$	$1{,}0\ 10^{-4}$	$3{,}4\ 10^{-5}$
Moyenne	$< 1{,}0\ 10^{-7}$	$8{,}5\ 10^{-7}$	$5{,}3\ 10^{-6}$	$6{,}2\ 10^{-5}$	$7{,}5\ 10^{-5}$	$2{,}3\ 10^{-4}$	$2{,}0\ 10^{-4}$
Ecart Type	$< 1{,}0\ 10^{-7}$	$1{,}2\ 10^{-6}$	$4{,}8\ 10^{-6}$	$4{,}3\ 10^{-5}$	$6{,}0\ 10^{-5}$	$2{,}9\ 10^{-4}$	$2{,}0\ 10^{-4}$

Tableau 5 : Présentation des fréquences de recombinaison mesurées au cours du temps après l'électroporation de p15CX dans des cellules TG1 contenant le vecteur pColYm.

Les taux de recombinaison sont compris entre $8{,}5\ 10^{-7}$ et $2{,}9\ 10^{-4}$ et augmentent au cours du temps pour se stabiliser après 8h d'incubation. L'analyse des écarts types révèle que les expériences sont à nouveau peu reproductibles. Pour la souche JC8679 mutée dans des gènes de recombinaison connus *RecB RecC* et *sbcA* (reçue du Prof. S. Colloms), le taux de recombinaison observé après 20h d'incubation est $5{,}8\ 10^{-4}$ et est donc similaire aux taux de la souche TG1 (résultats non montrés).

Discussion

Dans ce travail, nous avons tenté d'apporter des améliorations à une stratégie de reconstitution de gène par recombinaison homologue intermoléculaire, décrite par A. Brédo, dans le but d'augmenter son efficacité et de permettre son utilisation pour le contrôle de l'expression des collections de peptides cycliques. Une des améliorations principales a été de combiner la recombinaison homologue intermoléculaire avec une recombinaison site spécifique intramoléculaire afin de résoudre le cointégrat en deux vecteurs séparés et stables.

Les deux vecteurs résolus ont été identifiés dans des cellules ayant poussé sur un milieu sélectif, ce qui prouve que le système est fonctionnel. De plus, le cointégrat n'a jamais été observé, ce qui indique que la recombinaison site

spécifique est rapide et unidirectionnelle et que l'étape limitante de vitesse dans le système est la recombinaison homologue. Cependant, ce nouveau système est moins performant que le système précédent car les taux de recombinaison mesurés ne sont pas reproductibles et sont en moyenne 100x inférieurs à ceux initialement mesurés par A. Brédo même en variant les temps d'incubation et les multiplicités d'infection. De plus, l'ajout de phage auxiliaire dans la culture infectée pour enclencher la réplication simple brin des phagemides dans la cellule, n'influence pas le taux mesuré. Ce qui indiquerait que l'augmentation de l'efficacité de recombinaison dans le cas de l'implication d'un phage a plutôt lieu lors de l'infection que lors de la réplication. Comme dans nos expériences, la préparation de phagemides est contaminée à ~ 10 % par des phages auxiliaires, la surinfection des cellules par les phagemides pourrait être empêchée et l'ADN simple brin recombinogénique pourrait être moins disponible. Pour vérifier l'hypothèse, une expérience dans laquelle la préparation de phagemides serait débarrassée des phages helper pourrait être envisagée.

Après électroporation de l'ADN du phagemide dans la cellule, le système plasmide/plasmide a été étudié et des taux de recombinaison similaires à ceux obtenus pour le système phagemide/plasmide précédemment décrit ont été mesurés. Par conséquent, cela apporte une preuve supplémentaire que dans ce cas précis, l'intervention d'un vecteur phagique ne permet pas d'augmenter les taux de recombinaison homologue.

Dans cette première approche, la fonctionnalité de la stratégie de reconstitution de gène par recombinaison double a été prouvée. Cependant, sa faible efficacité empêche son utilisation pour la régulation de l'expression des banques de peptides cycliques. Pour la suite du projet, nous avons décidé de travailler avec le vecteur pBAD couramment utilisé et dont l'efficacité de répression du promoteur a été prouvée. En présence de glucose, ce vecteur a notamment permis de maintenir le niveau d'expression de la protéine CcdB en dessous de son seuil de toxicité en condition de répression[42] et de produire la RnaseP archéale hautement toxique en quantité suffisante pour les études structurales[43,44].

La recombinaison site spécifique étant plus efficace que la recombinaison homologue, il serait intéressant d'étudier un système qui impliquerait deux recombinases site spécifique : une intégrase pour la reconstitution du gène

d'intérêt (GOI) par cointégration et la recombinase TnpI pour la résolution du cointégrat (Figure 41). Une intégrase particulièrement intéressante serait celle du bactériophage mv4. En effet, une étude réalisée par P. Ritzenthaler[287] a démontré que les éléments de recombinaison *attP* et *attB* de cette intégrase sont fonctionnels dans *E. coli*. De plus, lorsque ces deux éléments sont clonés sur des vecteurs différents, la recombinaison site spécifique catalysée par l'intégrase mv4 dans *E. coli* mène à la formation d'un cointégrat visible sur gel d'agarose.

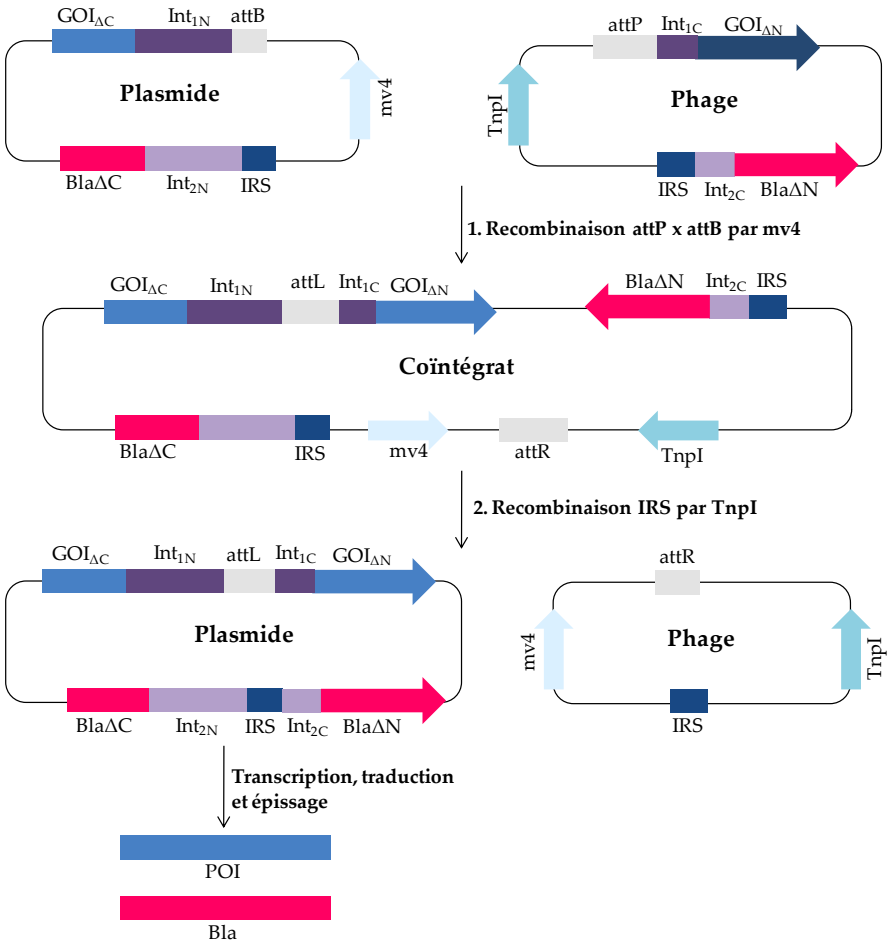


Figure 41 : Représentation schématique de la stratégie de reconstitution d'un gène d'intérêt (GOI) par un système qui implique deux recombinases site spécifique. La première recombinaison catalysée par l'intégrase mv4 au niveau des sites attB et attP forme un cointégrat qui est ensuite résolu par la résolvase TnpI au niveau des sites IRS. Les sites de recombinaison sont entourés par les fragments de l'intéine qui s'épisse après transcription et traduction pour donner la protéine d'intérêt mature (POI).

Pour cette nouvelle stratégie, deux vecteurs devront être créés. Un plasmide qui contiendra la partie N-terminale du gène d'intérêt avec le site de recombinaison *attB* et un phage qui contiendra le site de recombinaison *attP* avec la partie C-terminale du gène. Des sites de recombinaison IRS seront également ajoutés sur les deux vecteurs afin de permettre la résolution du cointégrat après la recombinaison *attB* x *attP*. Les sites de recombinaison (*attP*, *attB* et IRS) seront entourés par les fragments d'une intéine (Int_N et Int_C) qui s'épisseront après transcription et traduction pour donner la protéine mature. Les vecteurs résolus pourront être sélectionnés sur ampicilline grâce à la reconstitution du gène de la β-lactamase après résolution du cointégrat. Cette stratégie devrait permettre d'obtenir des rendements de reconstitution de gènes élevés car la recombinaison site spécifique est très efficace et unidirectionnelle.

Malgré sa faible efficacité, une application intéressante pourrait être envisagée pour la stratégie de reconstitution de gène par recombinaison double, il s'agit de l'assemblage d'opérons de grande taille *in vivo*. Le clonage d'opérons ou de gènes de grandes tailles dans un vecteur d'expression est souvent difficile chez E. coli car cela nécessite la présence de sites de restrictions spécifiques en bordure de l'opéron qui sont rarement présents. Une PCR doit alors être réalisée pour introduire ces sites, ce qui génère fréquemment, pour les fragments de grande taille, des mutations non désirées dans la séquence.

La stratégie que nous envisageons permettrait d'assembler l'opéron directement dans le vecteur d'expression en écartant le problème des mutations par ajout successif de fragments de petite taille (Figure 42). Les fragments d'opérons seront créés par PCR de façon à ce que chacun contienne une séquence homologue avec le fragment qui suit dans l'opéron. Le premier fragment de l'opéron sera cloné dans un vecteur d'expression en aval d'un promoteur et en amont d'un gène qui confère la résistance à la tétracycline et une séquence IRS. Les fragments suivants de l'opéron seront clonés par TA-cloning dans un vecteur avec une origine de réplication thermosensible en amont d'un gène qui confère la résistance à l'ampicilline pour les fragments pairs ou la tétracycline pour les fragments impairs et une séquence IRS. Tout d'abord, les deux premiers vecteurs contenant le fragment d'opéron 1 et 2 seront mis en présence dans *E. coli* pour enclencher la réaction de recombinaison double. Le vecteur d'intérêt 1b sera sélectionné en cultivant les cellules à une température non permissive et avec l'antibiotique approprié (ampicilline). Ensuite, le vecteur 1b sera mis en

présence du vecteur 3 pour donner le vecteur 1C. La démarche sera répétée jusqu'à ce que tout l'opéron soit assemblé.

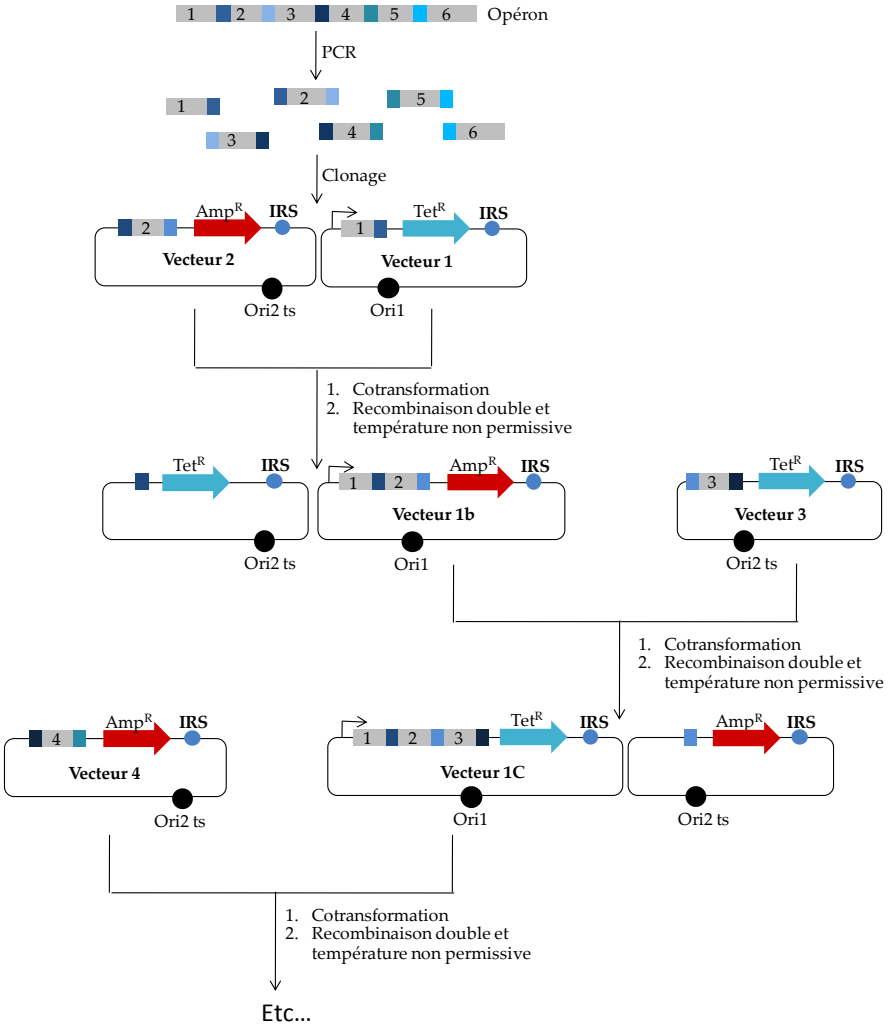


Figure 42 : Représentation schématique de la stratégie d'assemblage d'un opéron par recombinaison double (homologue et ensuite site spécifique).

Biosynthèse de collections de peptides cycliques dans le périplasme d'*Escherichia coli*

Introduction

En produisant des collections diverses de molécules, les méthodes chimiques et biologiques fournissent des outils pour contrôler ou comprendre les processus biologiques, moduler les interactions moléculaires et potentiellement traiter des maladies dans des applications thérapeutiques[288-290]. Les collections de petites molécules synthétiques peuvent être riches en diversité mais nécessitent souvent des méthodes laborieuses de criblage. En revanche, des composés biosynthétiques très complexes peuvent être produits par des organismes vivants et le lien génotype-phénotype permet généralement d'envisager la conception d'approches de sélection ou de criblage à haut débit.

Dans ce contexte, la biosynthèse de produits naturels "non naturels" est un domaine très actif de recherche. Un des succès les plus impressionnant est l'ingénierie génétique des synthases de polycétides (PKS) et de polypeptides non ribosomiaux (NRPS), qui a mené à la biosynthèse de plusieurs centaines de nouveaux composés[291-293]. Cependant, comme ces composés sont biosynthétisés par des systèmes multienzymatiques spécifiques et complexes, des approches combinatoires basées sur une dégénérescence du code génétique ne peuvent pas être développées facilement. A ce propos, la technologie SICLOPPS est une alternative intéressante car elle permet la biosynthèse de peptides cycliques stables à partir de précurseurs ribosomaux encodés[17,133]. Jusqu'à présent, la technologie SICLOPPS a été utilisée pour produire des banques de peptides cycliques dans le cytoplasme *d'E. coli* et a permis l'identification de peptides agissant sur des cibles co-localisées.

Objectif

Nous souhaitons développer une nouvelle application de la technologie qui permettrait la biosynthèse de collections directement dans le périplasme d'*E. coli*. Cette possibilité d'utiliser la chimie des intéines dans le périplasme ouvre un large potentiel d'applications comme la sélection de peptides cycliques qui devraient traverser la bicouche lipidique avant d'agir sur une cible

cytoplasmique, ou la recherche de composés qui inhiberaient des protéines exportées, ou interagiraient avec des domaines périplasmiques des protéines membranaires. C'est aussi la première étape vers la sécrétion de collections de composés biosynthétiques vers le milieu extracellulaire.

Le challenge est d'exporter les précurseurs intéine-peptide avant la cyclisation et de réaliser l'épissage avant l'inactivation de l'intéine par oxydation de sa cystéine essentielle dans le périplasme. En effet, le milieu périplasmique est oxydant et contient notamment des enzymes qui catalysent la formation de ponts disulfures. Dans ce travail, la mini-intéine Ssp DnaB de la cyanobactérie *Synechocystis sp.* PCC6803 a été utilisée pour cycliser la β-lactamase TEM-1 et pour créer des collections d'hexa-, hepta-, et octa-peptides cycliques dans le périplasme d'*E. coli*. Les constructions ont été réalisées dans le vecteur pBAD qui permet un contrôle étroit de l'expression des gènes[39]. L'exportation a été assurée par ajout d'un peptide signal en amont du précurseur intéine-peptide.

Résultats

Production d'une β-lactamase cyclique dans le périplasme

Nous avons décidé de travailler avec la mini-intéine Ssp DnaB artificiellement permutée de la cyanobactérie *Synechocystis sp.* PCC6803 car il a été démontré antérieurement que cette intéine sous forme permutée permet d'obtenir un haut rendement de protéine cyclisée[193]. La β-lactamase TEM-1 (Bla) a été choisie comme modèle de protéine périplasmique car son activité et sa stabilité dans la forme cyclique a déjà été démontrée dans le cytoplasme[294] et car sa localisation périplasmique peut être facilement démontrée par le phénotype de résistance à l'ampicilline.

En se basant sur le travail de X.Q. Liu[154], le domaine central de l'intéine Ssp DnaB a été supprimé pour créer une mini-intéine de 154 acides aminés divisée en deux parties, un fragment N-terminal de 106 résidus ($DnaBI_N$) et un fragment C-terminal de 48 résidus ($DnaBI_C$). Les deux fragments ont été arrangés dans l'ordre $DnaBI_C$-$DnaBI_N$ et cloné dans un vecteur d'expression en aval d'un promoteur pBAD contrôlé à l'arabinose (Figure 43A). Pour éviter tout problème potentiel d'oxydation dans le périplasme, la cystéine 50 non essentielle du fragment $DnaBI_N$ a été mutée en alanine. Le gène encodant Bla a été introduit

entre les deux fragments encodant l'intéine en fixant un codon sérine et glycine respectivement comme le premier résidu de l'extéine au niveau de la jonction N-terminale et le dernier résidu de l'extéine à la jonction C-terminale (position +1 et -1 dans la Figure 43B). En effet, la sérine est un résidu essentiel impliqué dans la réaction d'épissage[143] et la glycine a été rapportée comme importante pour l'efficacité de l'épissage de l'intéine Ssp DnaB[295,296]. Des séquences encodant cinq acides aminés additionnels ont été ajoutées sur les deux cotés de la séquence Bla pour fournir une séquence de douze résidus qui connectera les extrémités N et C originales de la protéine comme dans la construction cytoplasmique précédemment décrite[294]. Finalement, une séquence encodant le peptide signal de la β-lactamase TEM1 a été fusionné au coté 5′ de l'ORF (vecteur pPBla). Dans cette construction, la réaction d'épissage doit se produire après l'exportation pour que l'enzyme soit localisée dans le périplasme. Un vecteur exprimant un précurseur sans séquence signal a été également construit et utilisé comme contrôle négatif (vecteur pCBla).

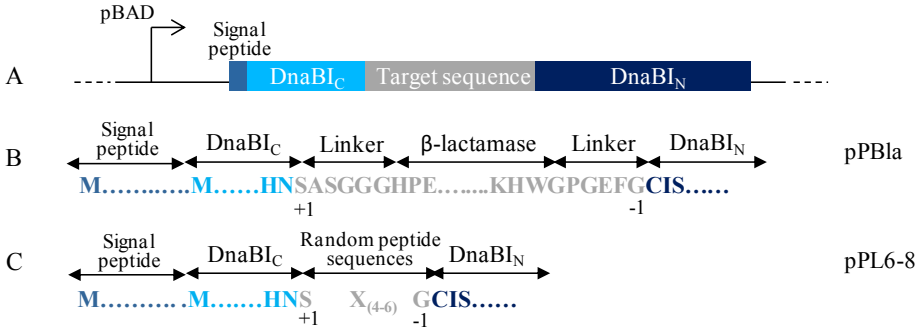


Figure 43 : Constructions pour la cyclisation de protéines et peptides dans le périplasme d'*E. coli*. (A) La séquence cible pour la cyclisation est clonée en respectant la phase de lecture entre les gènes encodant les deux fragments de la mini-intéine permutée Ssp DnaB (DnaBI$_C$ et DnaBI$_N$) dans un ordre permuté et en aval du promoteur pBAD contrôlé à l'arabinose. L'épissage est rendu possible par la présence d'une sérine (+1) et une glycine (-1) aux jonctions d'épissage (B) Détails pour le précurseur intéine-Bla. Les linkers permettent la connection des extérmités N- et C- terminales. (C) Les constructions pPL6-8 pour la cyclisation des collections d'hexa-, hepta-, et octa-peptides dans le périplasme.

Stratégie de construction

Le travail initial de construction du vecteur pBAD avec l'intéine permutée (pBI) a été réalisé par J.M. Léonard (Figure 44). Le gène Ssp $DnaBI_N$ a été amplifié à partir de la souche PCC6803 de la cyanobactérie Synechocystis sp. (Laboratoire O. Vallon) avec les amorces IN-fp et IN-rp (Tableau 6) qui introduisent les sites de restrictions 5' BbsI et 3' XbaI. Le gène Ssp $DnaBI_C$ a été amplifié avec les amorces IC-fp et IC-rp (Tableau 6) qui introduisent les sites de restriction 5' NcoI et 3' BbsI. Grâce à leur séquence complémentaire, les deux fragments PCR ont été assemblés dans l'ordre $DnaBI_C$-$DnaBI_N$ par PCR de recouvrement. Le produit PCR final a été cloné dans le vecteur pBAD entre les sites NcoI et XbaI, pour donner le plasmide pBI. Le vecteur d'expression pBAD est un dérivé de pBAD/myc-HisB (Invitrogen) dans lequel le marqueur de résistance à l'ampicilline a été remplacé par un gène de résistance à la tétracycline.

La mutation cystéine en alanine dans le gene $DnaBI_N$ (C50A) a été réalisée par mutagenèse "Quick-Change" avec les amorces QC-fp et QC-rp (Tableau 6) en suivant les instructions fournies par le fabricant (Stratagène) pour donner le plasmide pBICA. Le gène Bla a été amplifié à partir du plasmide pBADTEM1 avec les amorces Bla-fp et Bla-rp (Tableau 6) qui introduisent des sites de restriction 5' et 3' BbsI, un lieur N-terminal (SASGGG) et C-terminal (GPGEFG). Le produit PCR a été cloné dans le vecteur pBICA entre les sites BbsI pour donner le plasmide pCBla. La séquence du peptide signal de Bla a été extrait du plasmide pBADTEM1 par PCR avec les amorces PS-fp et PS-rp (Tableau 6) qui permettent d'introduire le site de restriction 5' NcoI. Le fragment PCR a ensuite été restreint par NcoI et introduit dans le vecteur pCBla dans le site unique NcoI pour donner le vecteur pPBla. Finalement, les vecteurs pCBla et pPBla ont été électroporés dans des cellules TOP10 d'*E. coli*.

La construction correcte des deux plasmides finaux pCBla et pPBla a été vérifiée par analyse du profil de restriction sur gel d'agarose et par séquençage.

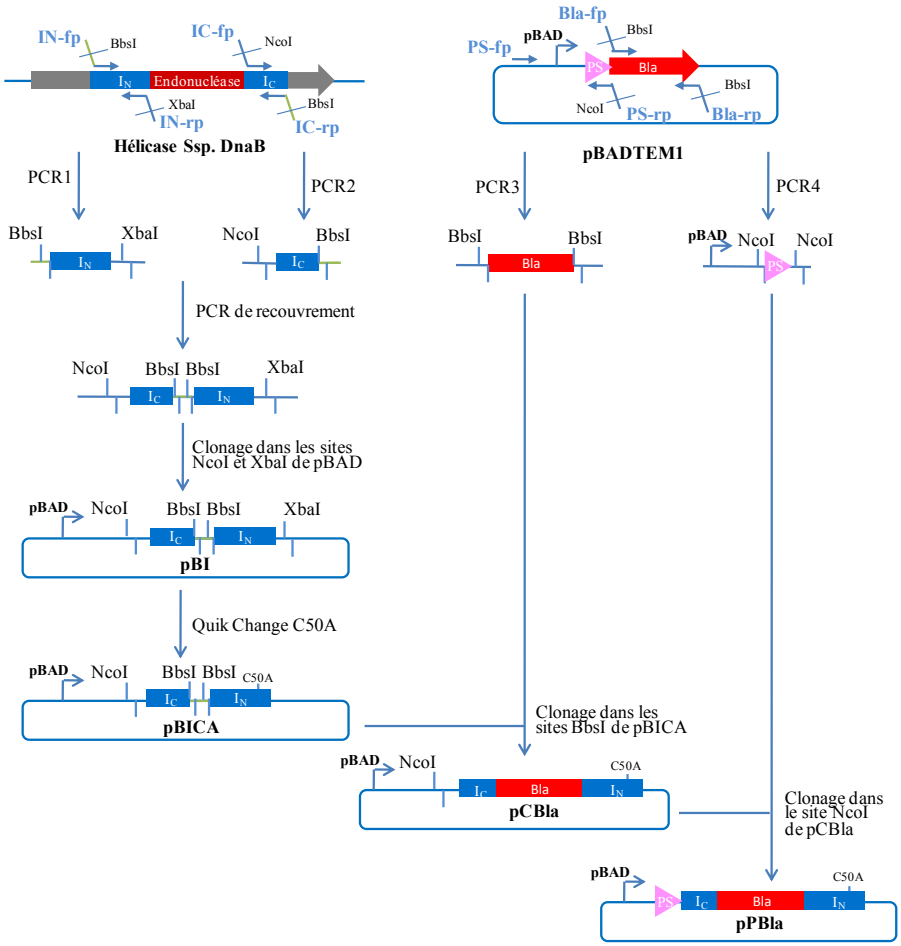


Figure 44 : Stratégie pour la construction des vecteurs pCBla et pPBla. Les étapes de la construction sont expliquées dans le texte ci-dessus.

Amorce	Séquence
IN-fp	5′-AACAGTGGGTCTTCGTGAAGACGCGGATGCATCAGTGGAGATAG-3′
IN-rp	5′-ATCTAGACTATAATTGTAAAGAGGAGCT-3′
IC-fp	5′-GCCCATGGGATCACCAGAAATAGAAAA-G-3′
IC-rp	5′-GCATCCGCGTCTTCACGAAGACCCACTGTTATGGACAATGATGTC-3′
QC-fp	5′-GTTAGTCGTGTATTTGCTACTGGCAAAAAGCTAG-3′
QC-rp	5′-CTAGCTTTTTGCCAGTAGCAAATACACGACTAAC-3′
Bla-fp	5′-CTGAAGACAACAGTGCTAGCGGTGGTGGTCATCCAGAAACGCTGGTG-3′
Bla-rp	5′-TACGAAGACGCATCCGAATTCCCCGGGTCCCCAATGCTTAATCAG-3′
PS-fp	5′-AAAACGCGTAACAAAAGTGTC-3′
PS-rp	5′-GCCTTCCTGTTTTTGCCATGGGAT-3′

Tableau 6 : Séquences des amorces utilisées pour la stratégie de construction des plasmides pCBla et pPBla. Les séquences reconnues par les enzymes de restriction utilisées sont soulignées. La séquence complémentaire est en bleu.

Au contraire des transformants pCBla, les bactéries transformées par le vecteur pPBla sont résistantes à des hautes concentrations en ampicilline (100 µg/ml, Figure 45). Par conséquent, cela indique une localisation périplasmique de la β-lactamase.

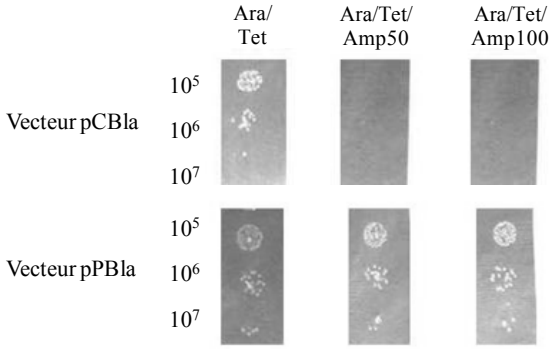


Figure 45 : Phénotype de résistance à l'ampicilline des cellules d'*E. coli* transformées par le vecteur d'expression de la β-lactamase cyclique cytoplasmique (pCBla) ou périplasmique (pPBla). Des dilutions 10x des cellules ont été étalées sur des boîtes de Pétri avec de la tétracycline (7,5 µg/ml) et du L-(+)-arabinose 0,2%, soit avec ou sans ampicilline (50 ou 100 µg/ml).

Des analyses SDS-PAGE des extraits cytoplasmiques et périplasmiques de cellules induites à l'arabinose pour les deux constructions ont confirmés que la localisation de Bla est dépendante du peptide signal (Figure 46, puits 3 et 9). Dans la construction périplasmique pPBla, la β-lactamase n'est pas détectable dans le cytoplasme après induction (Figure 46, comparaison puits 6 par rapport à 7). Ces conclusions ont été confirmées par des tests qualitatifs à la nitrocéfine et par des expériences de cinétique enzymatique d'hydrolyse de la pénicilline G. Pour la construction périplasmique pPBla, une concentration en β-lactamase 5 fois plus faible a été calculée dans le cytoplasme par rapport au périplasme (résultats non montrés). Cette faible concentration en enzyme est probablement due à la contamination de l'extrait cytoplasmique par des protéines périplasmiques lors de la préparation des extraits. Comme le peptide signal est fusionné à l'intéine, l'ensemble des résultats indiquent que l'épissage se déroule principalement après l'exportation.

La bande de la β-lactamase cyclique attendue après épissage de l'intéine apparait autour de 30 kDa (PM Bla cyclique = 29849,9 Da). Deux autres bandes

correspondant aux deux fragments de l'intéine $DnaBI_N$ et $DnaBI_C$ (11985 Da et 5544 Da) sont également observées à environs 12 kDa et 5,5 kDa. Des analyses par spectrométrie de masse (Louvain-la-Neuve, P. Morsomme et H. Degand) ont indiqué que ces bandes correspondent bien aux fragments de l'intéine (Voir travail de thèse de J.M. Léonard). Il est important de remarquer qu'aucun produit alternatif provenant d'une réaction d'épissage inefficace ou partielle n'a été détecté sur le gel.

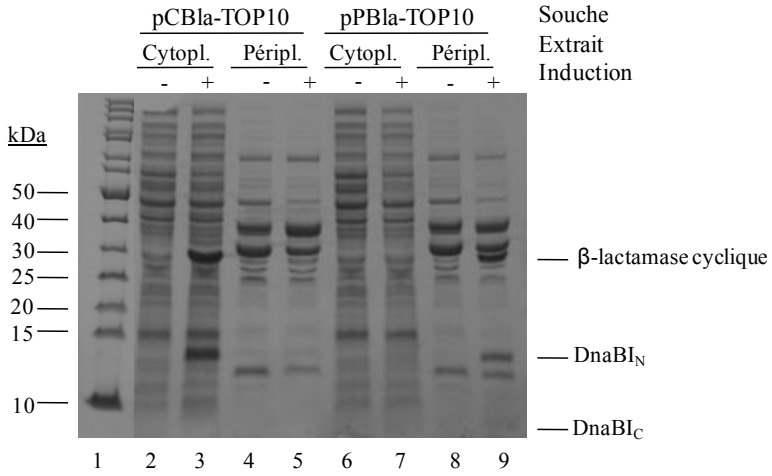


Figure 46 : Les extraits cytoplasmiques et périplasmiques des cellules TOP10 d'*E. coli* transformées par les vecteurs conçus pour l'expression de la β-lactamase cyclique dans le cytoplasme (pCBla) ou le périplasme (pPBla) en condition d'induction (arabinose) ou de non induction ont été analysés par SDS-PAGE. En condition d'induction à l'arabinose, des bandes correspondant à la β-lactamase cyclique et les fragments de l'intéine sont observés dans les extraits attendus. Puit1 : PageRuler™ Prestained Protein Ladder, Fermentas.

Pour prouver la cyclisation de la β-lactamase, l'enzyme a été purifiée de l'extrait périplasmique sur colonne de nickel et analysée au spectromètre de masse après une digestion à la trypsine. La purification sur colonne de nickel en absence de His-Tag est rendue possible par la présence de deux paires d'histidines à la surface de l'enzyme. Les masses des fragments observées sur le spectre de masse permettent de déterminer si la β-lactamase est cyclique et provient d'une réaction d'épissage ou si elle est linéaire et est le produit de réactions de clivage (Figure 47). Un peptide de 2107,02 Da correspondant à la séquence prédite pour la réaction d'épissage qui entraine la ligation des extrémités N- et C-terminales de la protéine a été clairement observé. Les peptides qui pourraient provenir de produits non épissés ou clivés (886 Da et 1239 Da) n'ont pas été détectés par spectrométrie de masse (Figure 48).

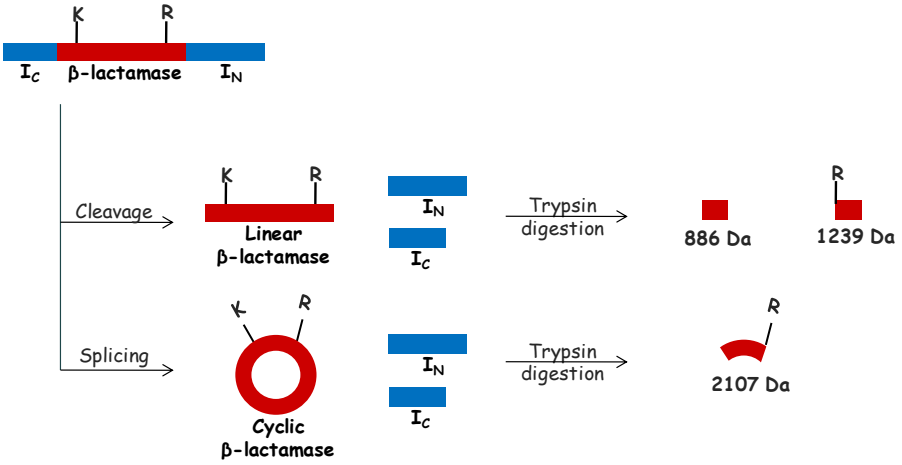


Figure 47 : Représentation schématique des deux réactions (clivage et épissage) qui pourraient se produire *in vivo* à partir des constructions pPBla et pCBla. Les réactions de clivage génèrent deux fragments de 886 Da et 1239 Da. La réaction d'épissage donne un seul fragment de 2107 Da dont la masse est égale à la somme des masses des deux framents résultant du clivage.

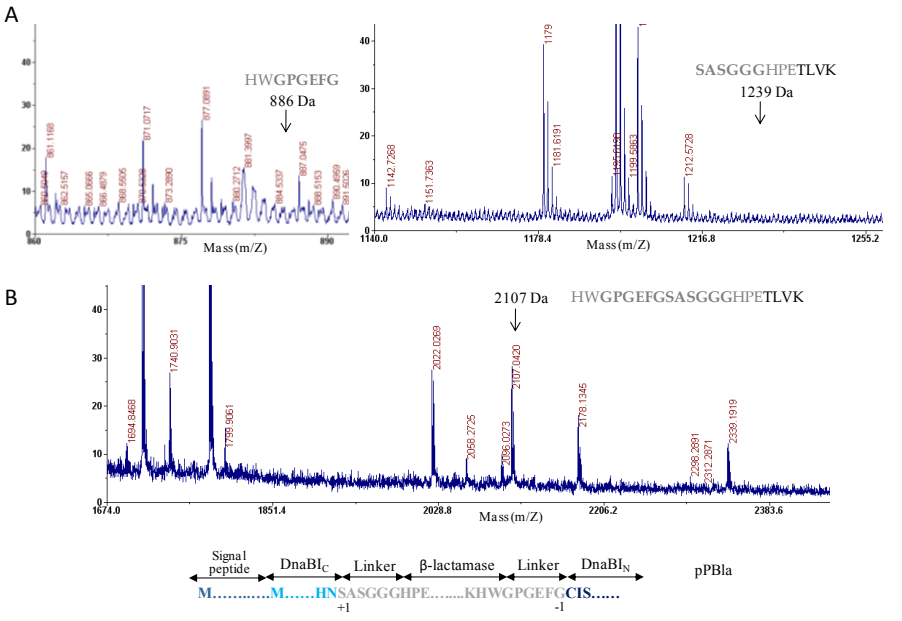


Figure 48 : Spectres de masse de la β-lactamase contenue dans l'extrait périplasmique après digestion à la trypsine. A. Les peptides qui pourraient provenir de produits non épissés ou clivés (886 Da et 1239 Da) n'ont pas été détectés. B. Le peptide de 2107,02 Da correspondant à la séquence prédite qui est attendue après la ligation des extrémités N- et C-terminales de la protéine a été clairement observé.

Production de collections d'hexa-, hepta- et octa-peptides cycliques dans le périplasme

En se basant sur la démonstration de la fonctionnalité de l'intéine dans le périplasme, des collections de peptides cycliques ont été construites en introduisant, à la place du gène de β-lactamase, des codons pour quatre, cinq ou six acides aminés entre les codons des résidus sérine et glycine (Figure 43C). Une dégénérescence de codon NNB a été choisie (B = C, T, G) pour fournir 48 codons qui encodent les 20 acides aminés et seulement 1 stop ambre TAG.

Stratégie de construction des collections

Pour construire les collections de peptides cycliques, le plasmide pBICA préalablement construit pour créer les vecteurs pCBla et pPBla a été utilisé. Tout d'abord, la séquence du peptide signal de Bla a été insérée dans le site unique NcoI de ce vecteur pour donner le plasmide pPBICA (Figure 49). Ensuite, à partir de ce vecteur, deux PCR ont été réalisées. La première utilise les amorces L1-fp et L2-rp (Tableau 7) qui amplifient un fragment contenant le promoteur P_{BAD}, la séquence du peptide signal de Bla et le gène $DnaBI_C$ avec en bordure les sites de restriction 5' MluI et 3' BbsI. La seconde utilise les amorces $L3_{(4\text{-}6)}$-fp et L4-rp (Tableau 7) qui amplifient un fragment contenant la séquence dégénérée et le gène $DnaBI_N$ avec en bordure les sites de restriction 5' BbsI et 3' XbaI. Une troisième PCR avec les amorces L5-fp et L4-rp (Tableau 7) a été réalisée sur le fragment obtenu après la PCR2 pour allonger la séquence en aval du site BbsI. Les fragments PCR1 et PCR3 ont ensuite été restreints par BbsI, ligués et finalement clonés dans le vecteur pBAD entre les sites de restriction MluI et XbaI pour donner les vecteurs pPL4-6. Finalement, les vecteurs ont été électroporés dans des cellules TOP10 d'*E. coli*.

Caractérisation des collections avec le peptide signal de Bla

Pour les trois banques créées, 10 clones piqués au hasard ont été analysés par restriction et séquençage. La construction correcte des vecteurs au niveau de l'insert $DnaBI_C$-peptide-$DnaBI_N$ a été confirmée pour les 30 clones. Les diversités des banques ont été estimées en encodant le nombre de transformants obtenus pour chaque banque dans l'algorithme GLUE-IT[297-299] (Tableau 8).

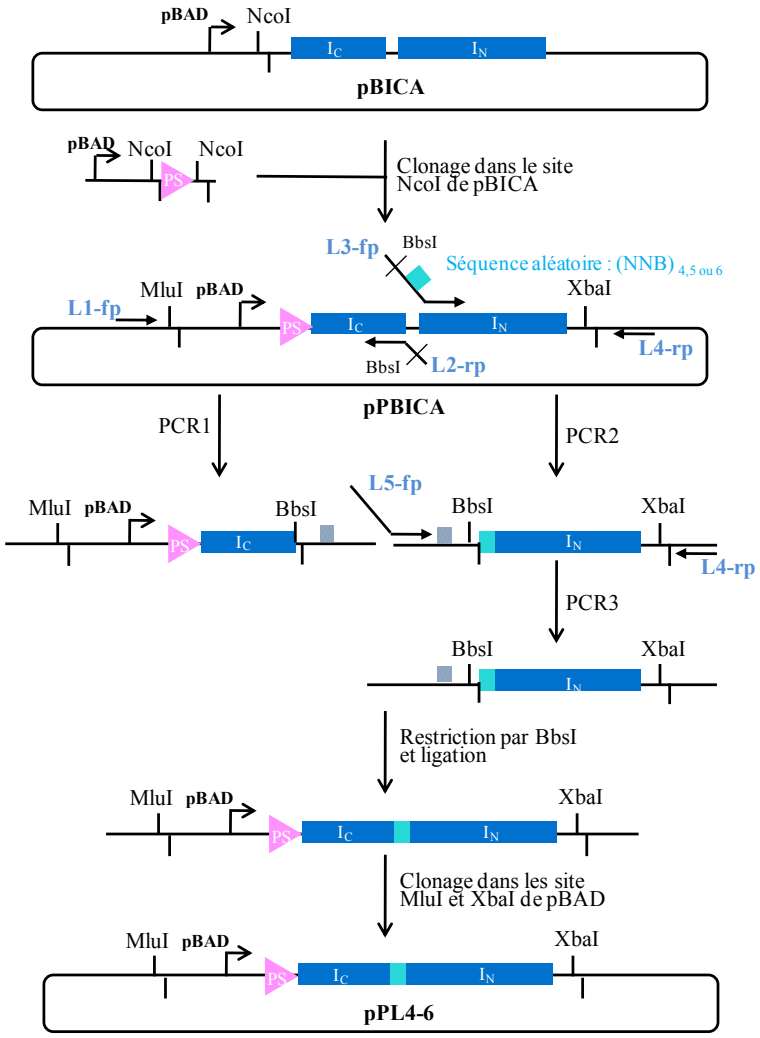


Figure 49 : Stratégie de construction des banques d'hexa-, hepta- et octa-peptides avec expression dans le périplasme. Les étapes de la construction sont expliquées dans le texte ci-dessus.

Amorce	Séquence
L1-fp	5'- CCCCGCTTATTAAAAGCATTC-3'
L2-rp	5'-CAGCATTACCCAGTCACCATGAAGACCCACTGTTATGG -3
L3$_{(6\text{-}8)}$-fp	5'-GGGGGAAGACTCCAGT(NNB)$_{4\text{-}6}$GGATGCATCAGTGGAG-3'
L4-rp	5'-GCCAAGCTGGAGACCGTTTAACT -3'
L5-fp	5'-GCATCAGTAGTCACGGGGGAAGA-3'

Tableau 7 : Séquences des amorces utilisées pour la stratégie de construction des collections de peptides cycliques. Les séquences soulignées dans l'amorce L2-fp et L3-rp sont les sites reconnus par l'enzyme BbsI.

Bien que les banques d'hexa- et hepta- peptide soient de 90 à 100% complète, la banque d'octa-peptide contient approximativement 40% de la diversité maximale. Dans l'ensemble, les trois collections contiennent environs 30 millions de clones différents.

Collection	Nombre de transformants	Diversité théorique	Diversité de la banque
pPL6 (SX_4G)	$6{,}7\ 10^7$	$0{,}16\ 10^6$	$0{,}16\ 10^6$
pPL7 (SX_5G)	$2{,}9\ 10^7$	$3{,}2\ 10^6$	$2{,}8\ 10^6$
pPL8 (SX_6G)	$6{,}8\ 10^7$	$6{,}4\ 10^7$	$2{,}7\ 10^7$

Tableau 8 : Diversités des collections d'hexa-, hepta- et octa-peptide avec expression périplasmique promu par le peptide signal de Bla. La diversité de chaque banque a été obtenue en rentrant le nombre de transformants dans l'algorithme GLUE-IT.

Dans les analyses SDS-PAGE des extraits périplasmiques de la collection d'hexa-peptides induite, les deux fragments de l'intéine libérés après l'épissage n'ont pas été observés (Figure 50, puits 2-3), contrairement aux résultats de la β-lactamase (Figure 50, puits 8). De plus, ces bandes ne sont pas présentes dans l'extrait cytoplasmique ou dans la fraction insoluble (Figure 50, puits 4-7). Les mêmes résultats ont été obtenus pour les collections d'hepta- et octa-peptides (Figure 51A) et pour des clones individuels piqués au hasard. Un problème d'exportation a été suspecté car les fragments de l'intéine ont été observés pour les mêmes collections mais dans les versions cytoplasmiques (Figure 51B et travail de thèse de J.M. Léonard).

En analysant la littérature plus en détails, nous avons réalisé qu'il existe deux voies de sécrétion, une post- et une co-traductionnelle. Par ailleurs, le peptide signal de la β-lactamase est connu pour promouvoir un transport post-traductionnel à travers la membrane interne[300]. Nous avons donc émis l'hypothèse que les précurseurs intéine-peptide libérés dans le cytoplasme pourraient prendre leur "folding" trop rapidement, ce qui empêcherait la translocation. Cette hypothèse est appuyée par un autre exemple dans la littérature dans lequel la fusion de la protéine cytoplasmique thiorédoxine avec le peptide signal post-traductionnel PhoA n'a donné qu'un très faible rendement de protéines dans le périplasme[249,301]. Dans ce cas, le remplacement de la séquence signal post-traductionnelle par une séquence co-traductionnelle a permis d'augmenter fortement le rendement d'exportation dans le périplasme. Donc, le peptide signal de la protéine DsbA qui est connu pour promouvoir une exportation co-traductionnelle[301-303] a été testé.

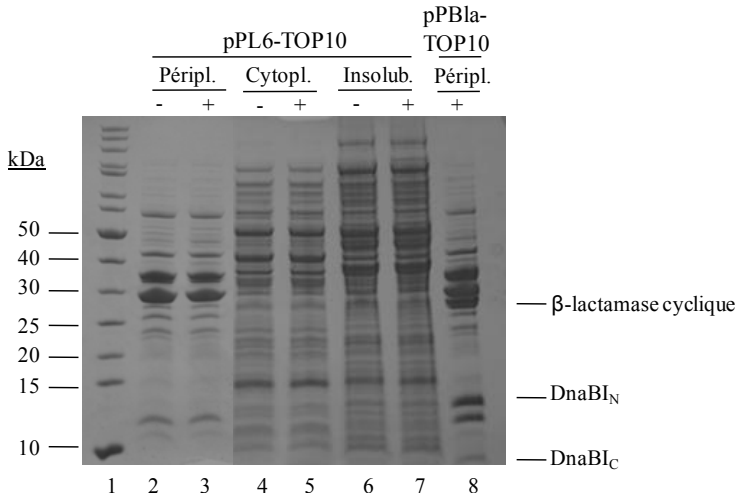


Figure 50 : Les fractions insolubles, les extraits cytoplasmiques et périplasmiques de cellules TOP10 d'*E.coli* transformées par les vecteurs d'expression périplasmique de la collection d'hexapeptides cycliques aléatoire (pPL6) en condition d'induction (arabinose) ou de non induction ont été analysés par SDS-PAGE. Les fragments de l'intéine et des précurseurs ne sont pas observés dans aucun des échantillons. L'extrait périplasmique obtenu pour le clone exprimant la β-lactamase cyclique périplasmique est montré comme control (puit 8). Puit1 : PageRuler™ Prestained Protein Ladder, Fermentas.

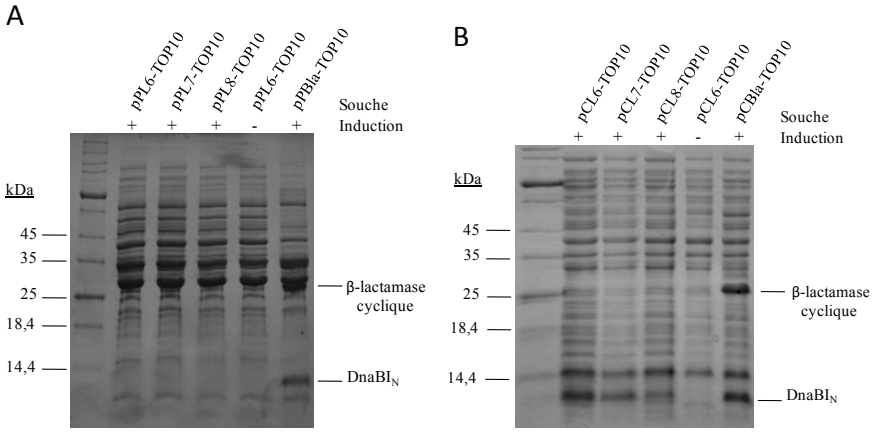


Figure 51 : (A) Les extraits périplasmiques de cellules TOP10 d'*E.coli* transformées par les vecteurs d'expression des collections pPL6, pPL7 et pPL8 ont été analysés par SDS-PAGE. (B) Les extraits cytoplasmiques de cellules TOP10 d'*E.coli* transformées par les vecteurs d'expression des collections pCL6, pCL7 et pCL8 ont été analysés par SDS-PAGE. L'extrait cytoplasmique ou périplasmique du clone pCBla ou pPBla est montré comme control.

Stratégie de construction des collections avec le peptide signal de DsbA

Des nouvelles collections d'hexa-, hepta- et octa-peptides cycliques ont été construites en utilisant la même stratégie mais en réalisant les PCR avec les amorces L1-fp, L2-rp et $L3_{(4-6)}$-fp, L4-rp et L5-fp (Tableau 7) sur le plasmide pBICA. Les deux fragments PCR1 et PCR3 ont été retreints par BbsI, ligués et finalement clonés dans le vecteur pPDsbA entre les sites de restriction NcoI et XbaI pour donner les vecteurs pPL4-6D. Le vecteur pPDsbA a été construit par G. Woronoff en clonant la séquence du peptide signal de la protéine DsbA contenant les bouts collants BspHI/NcoI dans le vecteur pBAD juste après le promoteur au niveau du site unique NcoI (BspHI et NcoI ayant des bouts collants compatibles). Une construction contrôle pPBlaD a été réalisée à partir du vecteur pPBla en remplaçant le peptide signal de Bla par celui de DsbA. Finalement, les vecteurs pPL4-6D et pPBlaD ont été électroporés dans des cellules TOP10 d'*E. coli*.

Caractérisation des collections avec le peptide signal de DsbA

Les diversités des banques ont été estimées en encodant le nombre de transformants obtenus pour chaque banque dans l'algorithme GLUE-IT (Tableau 8). La banque d'hexa-peptide est à 100% complète, tandis que les banques d'hepta- et d'octa-peptides contiennent respectivement 60 et 25% de la diversité maximale. Dans l'ensemble, les trois collections contiennent environ 17 millions de clones différents.

Collection	Nombre de transformants	Diversité théorique	Diversité de la banque
pPL6D (SX_4G)	$4{,}2\ 10^6$	$0{,}16\ 10^6$	$0{,}16\ 10^6$
pPL7D (SX_5G)	$5{,}8\ 10^6$	$3{,}2\ 10^6$	$1{,}9\ 10^6$
pPL8D (SX_6G)	$24\ 10^6$	$64\ 10^6$	$15\ 10^6$

Tableau 9 : Diversités des collections d'hexa-, hepta- et octa-peptide avec expression périplasmique promu par le peptide signal DsbA. La diversité de chaque banque a été obtenue en rentrant le nombre de transformants dans l'algorithme GLUE-IT.

Pour les trois banques créées, 20 clones piqués au hasard ont été analysés par restriction et séquençage. La construction correcte des vecteurs au niveau de l'insert $DnaBI_C$-peptide-$DnaBI_N$ a été confirmée pour 59 clones. Le clone mal construit provient de la collection d'octapeptides et correspond au vecteur sauvage pBADDsbA. Parmi les 59 clones correctement construits, huit clones (13,5%) contiennent un codon stop ambre dans la séquence dégénérée. Les

peptides cycliques de ces clones ne seront donc pas exprimés dans la souche TOP10 d'*E. coli*. Cependant, l'utilisation de souches suppressives qui permettent de continuer la traduction au travers du codon stop ambre pourrait être envisagée pour exprimer ces peptides et augmenter la diversité des banques.

Le pourcentage réel de chaque acide nucléique (A, T, C et G) pour chaque position dans le codon a été estimé à partir des séquences des 59 clones analysés (Figure 52). Théoriquement, les acides nucléiques pour les positions 1 et 2 devraient être en pourcentage de 25% chacun. Tandis que pour la position 3, l'adénosine ayant été supprimée, les trois acides nucléiques restant devraient être en proportion de 33,3% chacun. On constate que la cytosine est toujours sous-représentée, tandis que la thymine et la guanine sont sur-représentées. Au niveau des codons, cela se répercute principalement par une sur-représentation des glycines, des valines et des cystéines et une sous-représentation des prolines et des thréonines par rapport au pourcentage théorique (Figure 55).

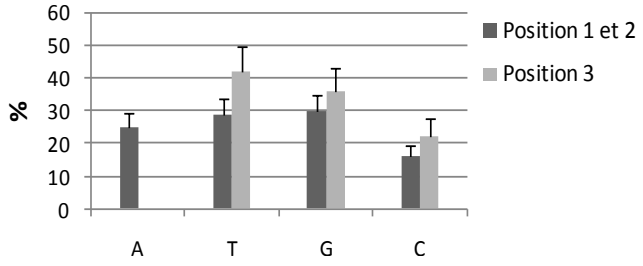


Figure 52 : Représentation du pourcentage expérimental de chaque acide nucléique pour chaque position dans le codon estimé à partir des séquences des 59 clones analysés.

L'ADN plasmidique d'un échantillon de cellules représentatif de chaque banque a été analysé par séquençage. Les électrophérogrammes résultants ont validé les résultats des pourcentages d'acides nucléiques estimés à partir des 59 clones. En effet, le pourcentage de cytosine est toujours plus faible à chaque position. De plus, comme attendu, l'adénosine n'est jamais détectée en 3ème position.

Comme montré dans la Figure 53 (puits 3-5), le changement du peptide signal mène à l'observation du fragment $DnaBI_N$ libéré après l'épissage dans le périplasme. Une autre bande est également observée à environ 18 kDa. L'analyse de cette bande par spectrométrie de masse, après découpage sur le gel et digestion à la trypsine, a permis de démontrer qu'il s'agit du précurseur non épissé $DnaBI_C$-Peptide-$DnaBI_N$ (résultat non montré). A part le précurseur non épissé, aucun produit dérivé de la réaction d'épissage n'a été détecté sur le gel

SDS-PAGE. Pour les constructions contrôles avec le gène *bla*, les deux peptides signaux Bla et DsbA produisent la β-lactamase cyclique dans le périplasme et dans les deux cas, le précurseur non épissé n'est jamais observé sur le gel SDS-PAGE (Figure 50, puit 8 et Figure 53, puit 6).

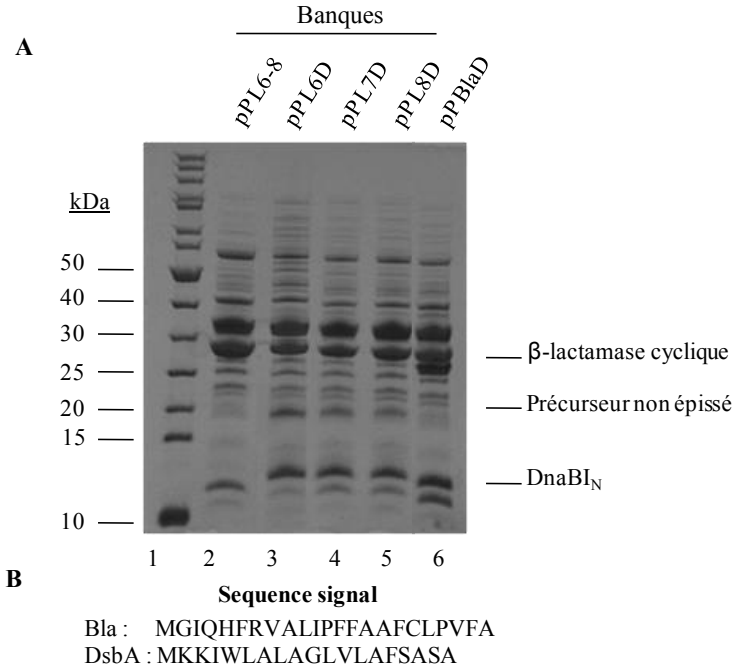


Figure 53 : (A) Comparaison des extraits périplasmiques obtenus à partir des collections induites à l'arabinose exprimant les précurseurs fusionnés au peptide signal de la β-lactamase (puits 2) ou de DsbA (puits 3-5). Dans le puits 2, les collections encodant les hexa-, hepta- et octa-peptides aléatoires ont été mélangées. Pour les constructions avec le peptide signal DsbA, chaque collection (pPL6D, pPL7D et pPL8D) a été analysée séparément. Les bandes correspondant au fragment DnaBI$_N$ et au précurseur non épissé sont observées quand le peptide signal DsbA est utilisé. L'extrait périplasmique obtenu à partir du clone exprimant le précurseur intéine-Bla avec le peptide signal DsbA est montré comme control (Puits 6). (B) Séquences du peptide signal de Bla et DsbA.

Les extraits périplasmiques préparés à partir de cellules induites de 34 clones sans codon stop dans l'ORF du peptide ont été analysés par gel SDS-PAGE (huit d'entre eux sont montrés dans la Figure 54). La présence et l'intensité des bandes correspondant au fragment DnaBI$_N$ et/ou au précurseur non épissé sont fortement variables d'un clone à l'autre. Pour tous les clones, aucun produit dérivé de la réaction d'épissage, à part le précurseur non épissé, n'a été observé sur le gel SDS-PAGE. Pour 8 clones, le fragment DnaBI$_N$ n'a pas été détecté.

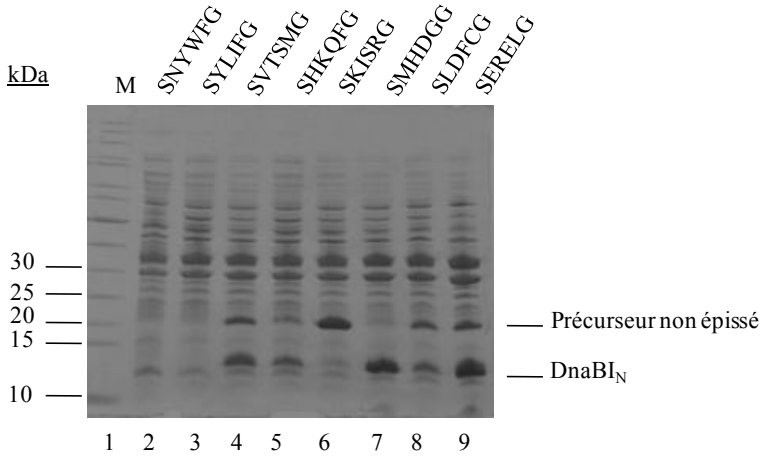


Figure 54 : Analyses SDS-PAGE des extraits périplasmiques préparés à partir des cultures induites à l'arabinose de huit clones piqués au hasard dans la collection d'hexa-peptide pPL6D. Les séquences des peptides sont montrées au dessus du gel. En fonction du clone, le fragment DnaBI$_N$ et le précurseur non épissé sont ou ne sont pas observés.

Peptide	Séquence
P6-8	SNYWFG
P6-9	SYLIFG
P6-10	SFCVNG
P7-9	SYCYTIG
P7-10	SHSYYVG
P8-12	SHTHCALG
P8-13	STYLNKSG
P8-14	SWETNTSG

Tableau 10 : Séquences des peptides provenant de clones qui n'ont pas montré une activité détectable de l'intéine.

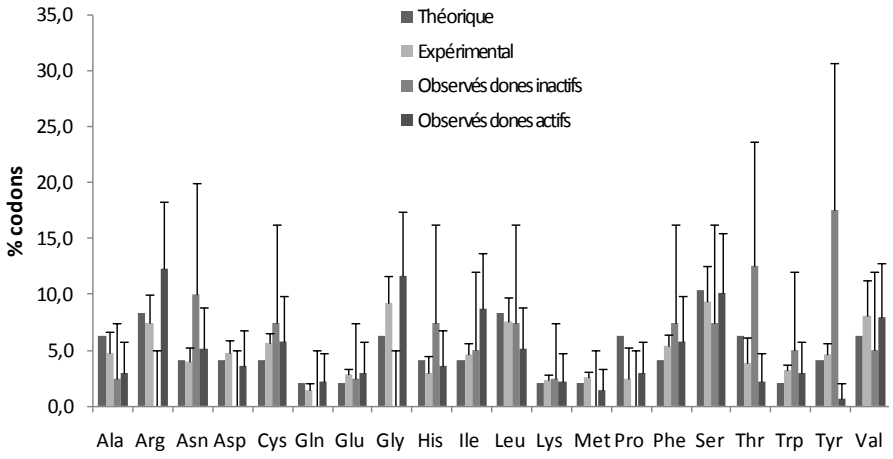
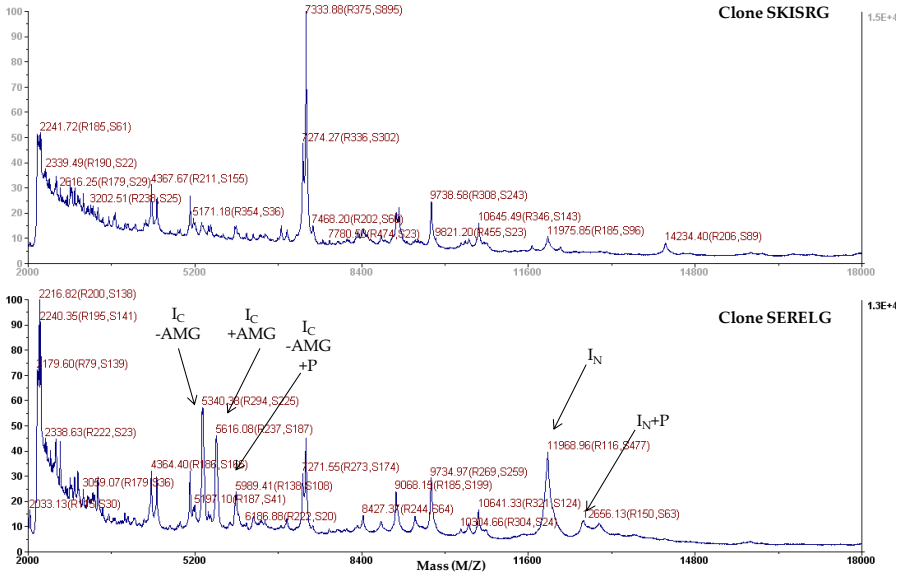


Figure 55 : Représentation en histogramme du pourcentage des codons dans les clones des banques périplasmiques. Nombre de codons théorique, expérimental et présent dans les clones analysés par SDS-PAGE comme actifs et inactifs au niveau de l'épissage.

Les séquences de ces clones ont été analysées et ont révélé une haute fréquence de tyrosines (17,5% : 7 sur 40 résidus aléatoires, Tableau 10) en comparaison à la dégénérescence expérimentale attendue du codon (4,6% de codon Tyr, Figure 55). Bien que l'échantillon soit légèrement trop petit pour que cette fréquence soit significative, elle est cependant corrélée à la très faible fréquence de tyrosine dans les séquences des clones montrant une activité de l'intéine (0,7% : 1 Tyr sur 138 résidus aléatoires, Figure 55). Une tendance inverse mais statistiquement encore un peu moins significative est observée pour les arginines et les glycines (Figure 55).

Les extraits périplasmiques des clones correspondant aux puits 6 et 9 sur le gel de la Figure 54, qui ont montré une activité de l'intéine, ont été concentrés 50 fois et analysés par spectrométrie MALDI-TOF directement comme échantillon brut pour tenter de détecter les différents fragments résultant de l'épissage ou du clivage (Figure 56 et figure en annexe). Le rapport signal/bruit des pics correspondant aux fragments de l'intéine pour le clone SERELG est meilleur que pour le clone SKISRG.



IC = **AMG**SPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHN = 5597,22 Da
IN = CISGDSLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFATGKKLVYILKTRLGRTIKATANHRF
LTIDGWKRLDELSLKEHIALPRKLESSSLQL = 11985,01 Da

Figure 56 : Spectres de masse de l'extrait périplasmique des clones SERELG et SKISRG concentré 50 fois et analysé directement comme échantillon brut. Le spectre permet d'observer les fragments de l'intéine I_C et I_N libérés après épissage ou clivage.

Le spectre du clone SERELG a permis d'observer que le fragment I_C, que l'on obtient après l'exportation et l'épissage, est réparti sur quatre pics différents dans une gamme de masse allant de 5340 à 5616 Da (Figure 57). Cela peut s'expliquer par le fait que le peptide signal de DsbA, nécessaire à l'exportation et qui se trouve à l'extrémité N-terminale, est clivé à deux endroits différents dans le polypeptide par la peptidase, avant l'analine ou après la glycine. De plus, l'asparagine au niveau de l'extrémité C-terminale est soit sous la forme cyclique ou soit a subit une hydrolyse. Le pic moins intense qui possède une masse de 6989 Da pourrait correspondre à un produit de clivage au niveau de l'extrémité C-terminale du peptide pour donner I_C+P. Le pic du fragment I_N réduit qui a une masse théorique de 11985 Da est clairement détecté sur le spectre. Un pic beaucoup moins intense qui a une masse augmentée de 688 Da, pourrait correspondre à un produit de clivage au niveau de l'extrémité N-terminale du peptide pour donner I_N+P. Les pics correspondant aux fragments du précurseur réduits ou oxydés n'ont pas été observés sur le spectre.

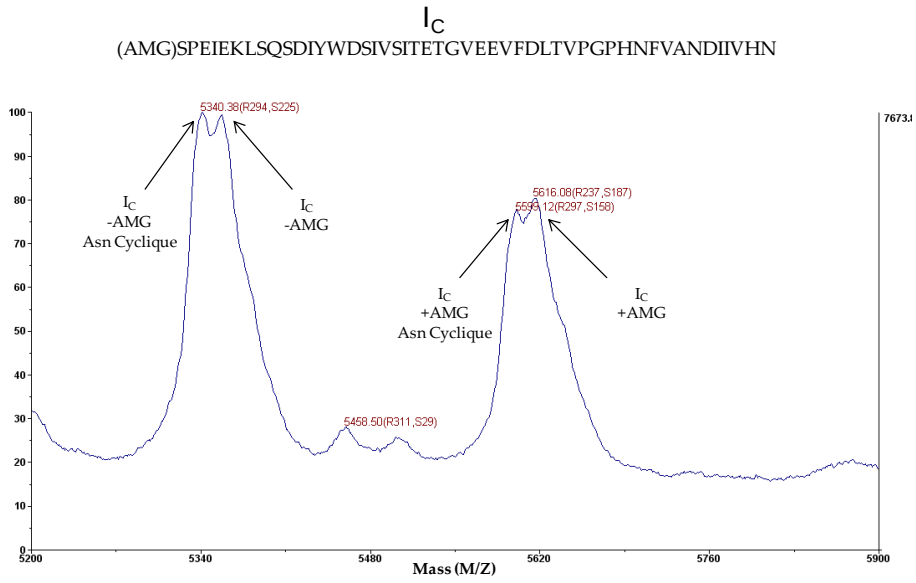


Figure 57 : Spetre de masse du clone SERELG zoomé sur la région permettant d'observer les quatres pics correspondant aux quatre formes du fragment I_C après exportation et épissage.

Des expériences ont été réalisées pour tenter d'expliquer la présence de précurseurs non épissés dans les extraits périplasmiques de cellules induites de certains clones. L'hypothèse suggérée est que l'intéine est probablement inactivée par oxydation de sa cystéine essentielle. En effet, le milieu

périplasmique est oxydant et contient notamment des enzymes qui catalysent la formation de ponts disulfures. Des extraits périplasmiques de cellules induites de quatre clones individuels pour lesquels un précurseur avait été observé (Figure 54) ont été analysés une seconde fois par SDS-PAGE mais cette fois en comparant la migration des bandes des échantillons en présence ou en absence d'agent réducteur, le DTT (Figure 58). Il est intéressant de remarquer que les expériences sont reproductibles car les quatre clones qui ont été analysés deux fois par SDS-PAGE (Figure 54 et Figure 58) montrent à chaque fois un même profil d'intensité pour les bandes du fragment $DnaBI_N$ et du précurseur. La bande du précurseur migre de la même façon dans les deux conditions, et aucune bande de poids moléculaire plus élevé n'est observée. Cela apporte une première évidence que s'il y a bien inactivation de l'intéine par oxydation de sa cystéine, cela est réalisé par une molécule de faible poids moléculaire et il n'y a pas de dimérisation du précurseur.

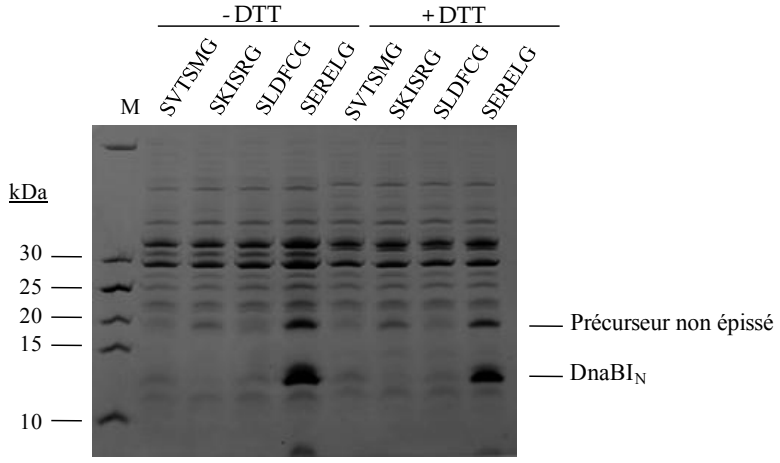


Figure 58 : Analyse SDS-PAGE des extraits périplasmiques traités au non au DTT préparés à partir des cultures induites à l'arabinose de quatre clones du gel précédent. Les séquences des peptides sont montrées au dessus du gel.

La bande à 18 kDa présente dans les extraits périplasmiques traités ou non au DTT des clones SERELG et SKISRG a été découpée, digérée à la trypsine et analysée par spectrométrie de masse. Les spectres à la Figure 59 présentent les résultats obtenus pour le clone SERELG avec en plus les spectres du clone SKISRG qui servent de contrôle négatif. Dans les échantillons non traité au DTT et pour les deux clones, un pic déplacé de 305,06 Da par rapport à la taille attendue du fragment contenant la cystéine a été observé (résultat pour le clone

SERELG, Figure 59A et pour le clone SKISRG, voir annexe). Dans les échantillons traités au DTT, ce même pic est absent (Figure 59B) mais le pic correspondant à la taille du fragment réduit a été observé (résultat non montré). De plus, dans les spectres du peptide SKISRG avec et sans DTT, le pic à 2056 Da est absent (Figure 59C et D). Le glutathion ayant une masse moléculaire de 307,323 Da pourrait être la molécule responsable de l'oxydation de la cystéine via la formation d'un pont disulfure. Une analyse par spectrométrie de masse en tandem (MS/MS) a confirmé que l'ion identifié correspond bien au fragment du précurseur modifié au niveau de la cystéine par glutathionylation (résultat non montré).

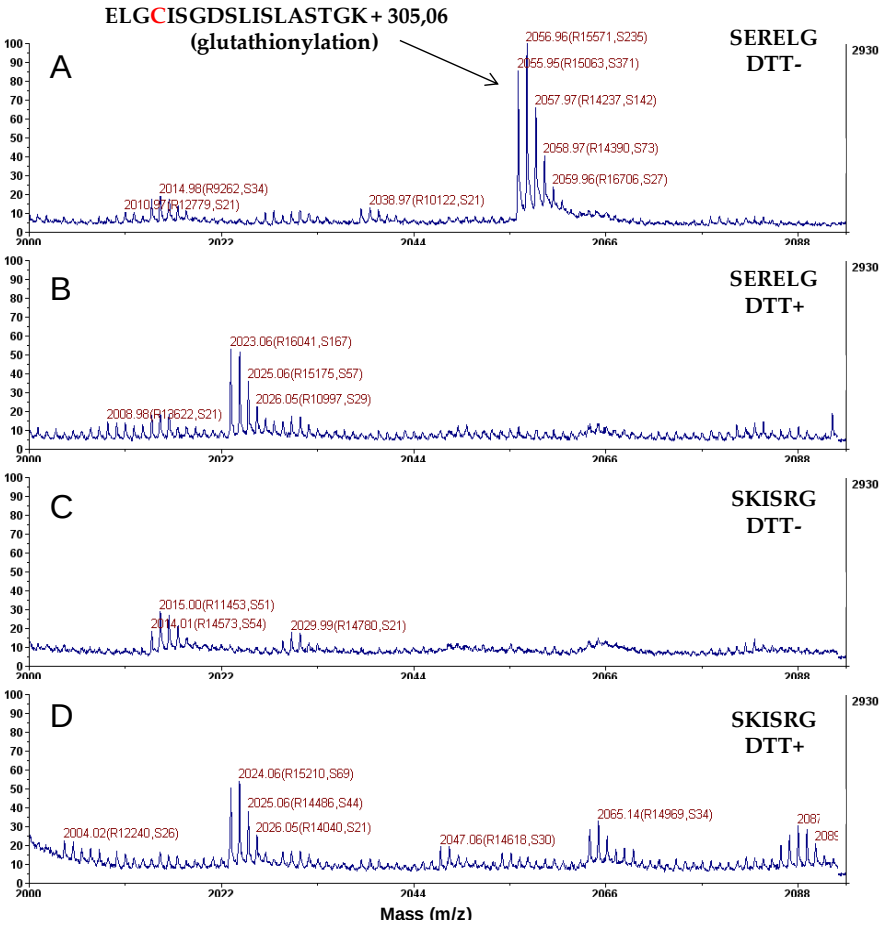


Figure 59 : Spectre de masse du précurseur non épissé du clone SERELG et SKISRG traité au DTT (B et D) ou non traité au DTT (A et C) après digestion à la trypsine. Les spectres du clone SKISRG servent de controles négatifs. Le pic à 2056 Da présent dans l'échantillon du clone SERELG a permis d'identifier la modifiation par le gluatathion.

Les extraits périplasmiques, préparés à partir de cellules induites des 26 clones qui ont montré une activité de l'intéine, ont été concentrés 50 fois et analysés par spectrométrie MALDI-TOF directement comme échantillon brut et/ou après purification sur une mini colonne C_{18}. Les résultats sont présentés dans le Tableau 11.

Peptide	Séquence	Masse calculée	Masse observée	Δ masse obs-calc	Rapport signal/bruit
P6-1	SH**K**QFG	685,33	685,35	0,02	871
P6-2	S**K**IS**R**G	629,36	629,38	0,02	182
P6-3	SM**H**DGG	585,2	585,22	0,02	222
P6-4	SE**R**ELG	672,32	672,31	-0,01	2130
P7-1	SL**RR**GGG	684,38	684,39	0,01	590
P7-2	SGL**R**GVG	627,35	627,37	0,02	204
P7-3	SGDEP**R**G	699,3	699,31	0,01	898
P7-4	S**R**HWSAG	782,36	782,38	0,02	257
P8-1	SW**R**TQIGG	886,42	886,46	0,04	643
P8-2	SSSSIP**R**G	772,38	772,40	0,02	165
P8-3	SNF**R**WVIG	960,49	960,49	<0,01	247
P8-4	SI**R**VQGVG	797,45	797,46	0,01	276
P8-5	SFG**H**ESGG	759,29	759,32	0,03	172
P8-6	S**R**PSLIVG	810,47	810,49	0,02	83
P8-7	S**H**NDTSYG	862,32	862,33	0,01	53
Not detected					
P6-5	SNFSTG	594,24			
P6-6	SVTSMG	563,24			
P6-7	SLDFCG	623,24			
P7-5	SIVFLGG	674,38			
P7-6	SGGNGCG	533,17			
P7-7	SN**R**SCIG	718,32			
P7-8	S**R**C**K**F**R**G	835,42			
P8-8	SLVLNSDG	786,39			
P8-9	SVPCFA**R**G	818,39			
P8-10	SCG**R**SAVG	718,32			
P8-11	SANCGICG	706,25			
Unspliced peptides with disulfide bridge					
P7-8a	S**R**C**K**F**R**GCI	1067,53	1067,54	0,01	172
P7-8b	S**R**C**K**F**R**GCISGDSL	1526,72	1526,77	0,05	56
P7-8c	S**R**C**K**F**R**GCISGDSLI	1639,81	1639,80	-0,01	35
P8-9a	SVPCFA**R**GCI	1050,49	1050,49	<0,01	115
P8-9b	SVPCFA**R**GCISGDSL	1509,69	1509,69	<0,01	578
P8-9c	SVPCFA**R**GCISGDSLI	1622,77	1622,76	-0,01	582
P8-9d	NSVPCFA**R**GCISGDSLI	1736,81	1736,82	0,01	321

Tableau 11 : Analyses MALDI-TOF des peptides cycliques exprimés dans des clones individuels d'une banque criblée par SDS-PAGE pour l'activité d'épissage de l'intéine.

Les produits cycliques $(M+H)^+$ ont été détectés pour les quinze peptides qui comportent au moins un résidu ionisable (Lys, Arg ou His) et aucune cystéine dans leur séquence. Pour trois d'entre eux, le signal du peptide était le plus intense de tout le spectre indiquant une concentration relativement élevée. En

plus des extraits périplasmiques des cellules non induites, chaque échantillon a servi de contrôle négatif pour les autres. L'absence systématique de signal pour les peptides cycliques ne contenant pas de résidus Lys, Arg et His suggère que l'énergie du laser du spectromètre n'est pas suffisante pour l'ionisation positive. D'un autre coté, quand un ou plusieurs résidus cystéine sont encodés dans l'ORF dégénérée, les peptides cycliques ne sont également pas détectés même si des résidus basiques sont présents ou si l'échantillon a été réduit avec du DTT avant l'analyse. Cela est plus que probablement lié à des modifications dans l'environnement oxydant du périplasme.

Pour les clones P7-8 et P8-9, respectivement quatre et trois fragments du précurseur non épissé ont été détectés (Tableau 11). Comparés aux peptides attendus, tous ces fragments contiennent quelques résidus additionnels au niveau de l'extrémité C-terminale et un pont disulfure qui résulte de l'oxydation de la cystéine essentielle de l'intéine avec la cystéine dans le peptide. Pour six fragments, DnaBI$_C$ a été clivé et les masses observées peuvent correspondre soit à des peptides linéaires oxydés intra-moléculairement soit à des peptides cycliques oxydés inter-moléculairement avec des petits fragments N-terminaux de DnaBI$_N$ (Figure 60). Après réduction et traitement à l'iodoacétamide, une masse correspondant au peptide linéaire avec une augmentation de 116 daltons due à la modification au niveau des deux cystéines par un groupement CH_2CONH_2 a été observée, indiquant une oxydation intramoléculaire.

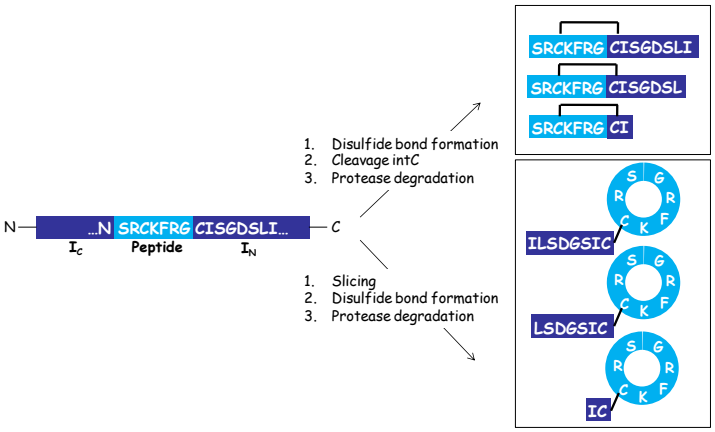


Figure 60 : Représentation schématique des deux pistes qui pourrait expliquer la masse observée par spectrométrie de masse pour les six peptides contenant une cystéine dans la région aléatoire en plus de celle qui intervient dans la réaction d'épissage. On obtient soit un peptide linéaire oxydé intra-moléculairement soit à un peptide cycliques oxydé inter-moléculairement avec des petits fragments N-terminaux de DnaBI$_N$

Des étalements sur boîtes de Pétri de culture de cellules des clones pP6-1, pP7-1 et pP8-1 en condition d'induction et répression ont révélé que les colonies présentent un phénotype particulier en condition d'induction. Celles-ci sont en effet plus petites et plus transparentes qu'en condition de répression. De plus, les mesures de la densité optique à 600 nm de ces cultures au cours du temps, ont montré qu'après 2h d'incubation la cultrure induite entre en phase stationnaire alors que le contrôle non induit continue de pousser. Après 4h d'incubation, une différence d'environ 1 unité d'absorbance est observée entre la culture induite et le contrôle non induit (Figure 61). Cependant, des étalements sur boîtes de Pétri de ces cultures induites et réprimées ont démontré qu'à densité optique égale, le nombre de colonies formées est similaire, indiquant une faible mortalité. Pour un clone avec expression cytoplasmique, pC6-1, la différence de densité optique entre la culture en condition d'induction ou de répression est moins marquée, environ 0,2 unités d'absorbance (Figure 61). L'ensemble de ces résultats indiqueraient donc que l'utilisation de la séquence signal de DsbA pour l'exportation causerait un ralentissement de la croissance et/ou de la division des cellules induites.

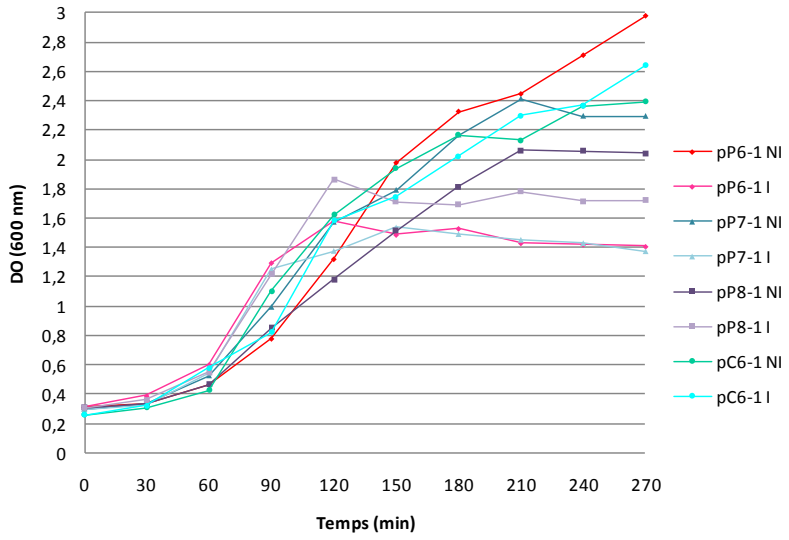


Figure 61 : Mesures toutes les 30 minutes de la densité optique à 600 nm de cultures de cellules en condition d'induction (I) et de répression (NI) à 37°C. Les bactéries ont été d'abord cultivées en présence de glucose 1 % jusqu'à une densité optique de 0,6. Après centrifugation et incubation à 37°C pendant 30 mintues dans du nouveau milieu de culture, le glucose 1 % ou l'arabinose 0,2 % ont été ajoutés dans les échantillons. L'expérience a été réalisée sur trois clones avec expression périplasmique (pP6-1, pP7-1 et pP8-1) et sur un clone avec expression cytoplasmique (pC6-1).

Discussion

L'activité d'épissage de la mini-intéine permutée Ssp DnaB a été combinée à une voie de transport co-traductionnel pour créer une méthode générale pour la production de protéines et de peptides cycliques dans le périplasme d'*Escherichia coli*. Avec un minium d'exigence dans la séquence du polypeptide à cycliser, une glycine et une sérine respectivement comme premier et dernier résidu, une cyclisation périplasmique efficace a été démontrée pour une protéine modèle et pour une grande proportion des peptides piqués au hasard dans des collections d'hexa-, hepta- et octa-peptides.

La cyclisation se produit spontanément sans aucune preuve, sur gel SDS-PAGE ou par analyse au spectromètre de masse, d'accumulation de produits dérivés qui pourraient provenir d'une réaction de clivage. Cependant, pour un grand nombre de clones des collections de peptides mais pas pour la protéine modèle, des précurseurs non épissés (DnaBI$_C$-peptide-DnaBI$_N$) ont été observés sur gel SDS-PAGE. Les analyses par spectrométrie de masse ont permis de démontrer que cette accumulation du précurseur non épissé est due à l'inactivation de l'intéine par glutathionylation de sa cystéine essentielle. Des rapports d'intensités forts variables des bandes du précurseur non épissé sur le fragment DnaBI$_N$ ont été observés d'un clone à l'autre. Une hypothèse serait que la séquence peptidique à cycliser de chaque clone pourrait créer un contexte différent et spécifique qui influencerait le rapport de vitesse de l'épissage sur la glutathionylation et conduirait à l'observation de différentes concentrations en précurseur non épissé. Pour un même clone, un même rapport d'intensité des bandes du précurseur non épissé sur le fragment d'intéine DnaBI$_N$ a toujours été observé d'une expérience à l'autre. Ce qui suggère que les expériences sont reproductibles. Nous disposons donc d'un système qui pourrait servir de sonde pour étudier la réaction de glutathionylation dans le périplasme.

Les résultats de spectrométrie de masse permettant de mettre en évidence l'inactivation de l'intéine par glutathionylation ont été obtenus très récemment au laboratoire et nécessitent d'être complétés par d'autres expériences. Cependant, cette découverte inattendue est enthousiasmante, car il s'agit du premier exemple de polypeptide glutathionylée identifiée dans le périplasme. Il sera important dans la suite du travail de tenter de déterminer le rôle physiologique de cette modification protéique dans le périplasme. Deux pistes

de réflexion inspirées par deux articles récents de J. Beckwith[304] et de J.F. Collet[305] sont proposées.

Dans la première piste de réflexion, la glutathionylation aurait pour fonction de réguler les réactions d'oxydo-réductions. En effet, dans l'article de J. Beckwith, ils ont démontré que du glutathion oxydé serait présent dans le périplasme et collaborerait avec la glutaredoxine pour oxyder les protéines. De plus, ils ont observé que l'activité *in vivo* de la glutaredoxine exportée n'est pas dépendante de DsbB mais nécessite une seule cystéine dans le site actif et est dépendante de la voie de biosynthèse cytoplasmique du glutathion. La Figure 62 ci-dessous présente le modèle proposé pour expliquer la catalyse de formation de pont disulfure par la glutaredoxine utilisant un méchanisme impliquant un monothiol. La glutaredoxine glutathionylée est attaquée par la cystéine d'un substrat, produisant une glutaredoxine réduite et le substrat glutathionylé. Dans une seconde étape, une seconde cystéine dans le substrat attaque le pont disulfure glutathion-substrat pour donner le substrat oxydé. Dans ce modèle, si le substrat ne contient qu'une seule cystéine, il restera glutathionylé.

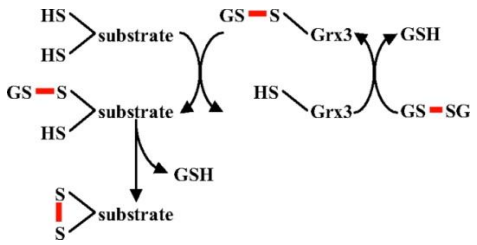


Figure 62 : Modèle pour la formation de pont disulfure catalysé par la glutaredoxine et utilisant un méchanisme impliquant un monothiol. La glutaredoxine glutathionylée est attaquée par la cystéine d'un substrat, produisant une glutaredoxine réduite et le substrat glutathionylé. Dans une seconde étape, une seconde cystéine dans le substrat attaque le pont disulfure glutathion-substrat pour donner le substrat oxydé (repris de l'article de J. Beckwith, 2009[304]).

Dans la seconde piste de réflexion, la glutathionylation jouerait un rôle au sein d'un système de réduction permettant de protéger de l'oxydation irréversible les protéines qui possèdent un seul résidu cystéine ou un nombre impair de cystéines. En effet, comme ces cystéines ne sont pas impliquées dans la formation de ponts disulfures, elles pourraient être vulnérable à l'oxydation en acides sulféniques et ensuite être modifiées irréversiblement en acides sulfiniques et sulfoniques[305-308] (Figure 63). Il a été démontré que deux protéines apparentées aux thiorédoxines, DsbG et DsbC, contrôlent le contenu en acide suflénique du périplasme et protègent ces résidus cystéine unique en les

réduisant[305]. Le glutathion sous sa forme réduite pourrait également jouer un rôle dans la protection de ces protéines et intervenir en formant un substrat glutathionylé qui devrait ensuite être réduit par une enzyme présente dans le périplasme.

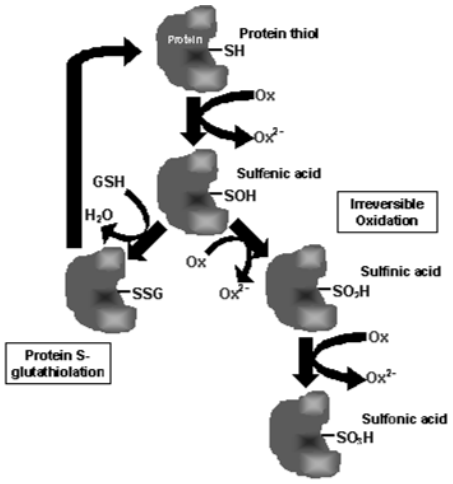


Figure 63 : Représentation schématique de l'oxydation irréversible des protéines en acide sulfinique et sulfonique. Intervention du glutathion et d'autres ensymes pour protéger ces protéines et pour les maintenir sous la forme réduite (repris de l'article de J. Momand, 2003)[309]

Dans le cadre des banques, une perspective intéressante serait d'utiliser une autre intéine capable de réaliser l'épissage à l'aide de deux sérines. Cela permettrait d'éviter l'inactivation d'une partie des précurseurs par l'oxydation de l'intéine dans le périplasme. Des expériences réalisées par A. Brédo ont démontré que l'intéine Ssp DnaB perd son activité d'épissage lorsque l'on remplace son résidu cystéine par une sérine. Les deux autres intéines qui ont été largement étudiées, la Ssp DnaE et la PI-PfuI possèdent également une cystéine dans leurs résidus catalytiques. Il faudrait donc choisir une intéine avec deux sérines dans celles qui n'ont encore été que peu étudiées et réaliser toute l'étude de l'épissage dans le contexte d'une extéine non naturelle. De plus, il faudrait s'assurer que l'intéine conserve son activité après la fragmentation et la permutation.

Bien que l'efficacité de l'épissage est dépendante de la séquence du peptide et que des précurseurs non épissés sont souvent observés, les rendements d'épissage apparaissent assez élevé à en juger par la détection facile des produits cycliques dans les extraits périplasmiques bruts. Pour les deux autres intéines

qui ont été utilisées pour la cyclisation dans le cytoplasme, l'intéine naturellement dimérique DnaE de *Synechocystis* et l'intéine fragmentée artificiellement PI-PfuI de *Pyrococcus furiosus*, le précurseur non épissé était également observé[17,153,194]. De plus, avec l'intéine Ssp DnaE, le rendement de produit cyclique était réduit par des produits de clivage détectables sur gel SDS-PAGE[17,153].

Un problème de ralentissement de croissance en condition d'induction a été mis en évidence pour les trois collections ainsi que pour des clones isolés. Ce ralentissement de croissance pourrait être causé par l'expression trop importante de précurseurs fusionnés à la séquence signal de DsbA qui entrainerait la saturation de la voie d'exportation SRP avec pour conséquence une monopolisation importante des ribosomes de la bactérie et des problèmes d'exportation pour les autres protéines périplasmiques. Dans ce contexte, il est important d'envisager des méthodes pour diminuer l'expression des précurseurs à exporter. Le promoteur pBAD fonctionne comme un interrupteur tout ou rien et, lorsqu'il est induit, mène à une expression très forte. La solution serait de diminuer le niveau d'expression en découplant le transport de l'inducteur de sa présence dans la cellule et ce par induction constitutive du transporteur[310,311]. Une autre solution serait de remplacer le promoteur pBAD par le promoteur inductible au propionate P_{prpB} qui est moins efficace au niveau de la répression mais qui permet d'induire à des niveaux intermédiaires[312-316]. On pourrait également envisager de travailler dans des souches qui permettent de réduire le nombre de copie du plasmide[317-319].

Une banque d'expression de peptides cycliques contenant environ 17 millions de clones différents a été construite. Le nombre de transformants obtenu pour la banque d'hexapeptides a permis d'avoir une banque à 100% complète. Cependant, pour les banques d'hepta- et octa-peptides, il serait intéressant d'envisager de construire des collections avec un nombre de transformants plus élevés pour atteindre également la diversité théorique totale. Les 17 millions de clones obtenus peuvent être divisés en trois groupes. Approximativement 50% des clones biosynthétisent effectivement les peptides cycliques dans le périplasme, 30% ont un codon stop ambre dans l'ORF du peptide ou sont exportés comme des précurseurs inactifs qui s'accumulent dans le périplasme, et environ 20% encodent des peptides avec une ou plusieurs cystéines qui sont

sujettes à des modifications. En tout, entre 8 et 10 millions de peptides cycliques différents sont biosynthétisés dans la banque.

Comme les peptides cycliques sont génétiquement encodés, des stratégies de sélection peuvent être conçues pour rechercher des molécules qui se lient à des cibles spécifiques. La localisation périplasmique est intéressante à plusieurs niveaux. Premièrement, la banque peut être utilisée pour rechercher des peptides agissant sur des cibles co-localisées, par exemple des protéines exportées naturelles ou recombinantes. Deuxièmement, la sélection de peptides périplasmiques contre une cible cytoplasmique pourrait être envisagée et nécessiterait automatiquement le passage à travers la membrane interne. Cela est particulièrement intéressant si l'on recherche des molécules avec un potentiel thérapeutique. Cependant, comme la sélection ne concernera que les peptides capables de diffuser, la diffusion à l'extérieur de la bactérie pourrait être un problème et une sélection sur un milieu solide devra être préférée pour augmenter la concentration locale des peptides co-localisés sur les colonies qui les produisent. Les protéines membranaires pourraient être également très attractives comme cibles, en particulier les protéines des membranes plasmiques des eucaryotes. En effet, s'il est possible de les exprimer fonctionnellement dans la membrane interne d'*E. coli,* un peptide agissant du périplasme pourrait être équivalent à un peptide venant de l'extérieur de la cellule eucaryote qui se lie directement sur un site externe de la protéine ou pénètre dans la membrane ou le cytosol pour rechercher une cible interne.

La biosynthèse de peptides cycliques dans le périplasme est également une première étape vers la sécrétion dans le milieu externe. Alors qu'il semble difficile d'amener les précurseurs intéine-peptide dans le milieu extracellulaire avant l'épissage, l'utilisation de mutants d'*E. coli* qui présentent une perméabilité accrue au niveau de la membrane externe pourrait libérer les peptides du périplasme. Cela ouvrirait des nouvelles possibilités de concevoir des expériences de sélection Darwinienne comme une compétition entre la banque et des pathogènes spécifiques.

La détection, dans les extraits périplasmiques, de peptides individuels allant de 585 à 960 daltons avec des rapports signal/bruit relativement hauts et similaires (Tableau 11) suggère qu'une majorité des clones dans les collections exprime des peptides qui s'accumulent dans ce compartiment et ne peuvent pas diffuser dans

le milieu. En effet, il est généralement accepté que la membrane externe d'*E. coli* n'est pas perméable aux molécules hydrophiles dont la masse est supérieure à 600 daltons, une limite qui a été évaluée en utilisant des oligosaccharides linéaires[320,321]. Comme les collections contiennent des peptides contraints structurellement avec des hydrophilicités très variables, il est néanmoins possible que certains d'entre eux sont capables de diffuser à travers les membranes internes et/ou externes. La bicouche lipidique pourrait être perméable à au moins quelques peptides hydrophobes et des porines non spécifiques de la membrane externe pourraient également faciliter le passage de certains petits peptides hydrophiles.

Mise au point d'une approche de sélection anti-darwinienne pour identifier des peptides toxiques potentiellement antibiotiques

Introduction

L'émergence de bactéries pathogènes résistantes aux antibiotiques a augmenté l'intérêt dans la découverte de nouveaux composés antibiotiques[322]. Antérieurement, les produits naturels ont joué un rôle important dans la découverte d'agents thérapeutiques contre les maladies infectieuses. Cependant, comme il est devenu de plus en plus difficile de trouver de nouveaux agents antimicrobiens dans la nature, il est primordial de développer des stratégies pour augmenter les chances d'identifier de nouveaux composés avec un potentiel thérapeutique. Une stratégie intéressante est de mettre au point des méthodes de sélection innovantes pour rechercher des molécules d'intérêt dans des collections de composés de grande diversité. Les collections de peptides cycliques cytoplasmiques et périplasmiques générées par la technologie SICLOPPS sont d'un intérêt tout particulier car étant encodées génétiquement, le lien génotype-phénotype permet d'envisager la conception d'approches de sélection ou de criblage à haut débit. De plus, ayant des poids moléculaires inférieurs à 1500 Da, les peptides cycliques ont la capacité de moduler les processus biologiques à travers des interactions avec des macromolécules comme les protéines et l'ADN[289].

Objectifs

Nous souhaitons développer des stratégies innovantes de sélection au sein des collections de peptides cycliques cytoplasmiques et périplasmiques pour tenter d'identifier de nouveaux composés antibiotiques biosynthétiques. Alors que la sélection "Darwinienne" est basée sur la survie ou sur un avantage biologique, ces stratégies viseront à sélectionner pour la mort cellulaire. Dans le cadre des projets du laboratoire, deux approches de sélection différentes ont été étudiées. Une stratégie est applicable pour les peptides bactériolytiques (voir travail de thèse de J.M. Léonard) et l'autre, développée plus en détail ci-après, pour les peptides bactéricides non bactériolytiques. Dans le cadre de l'identification de peptides bactériolytiques, l'objectif sera de sélectionner pour la lyse bactérienne, ou plus précisémment de récupérer les séquences génétiques encodant les

peptides responsable de cette lyse. S'il y a lyse bactérienne, le matériel génétique sera libéré dans le milieu de culture et pourra être simplement récupéré dans le surnageant après centrifugation et amplifié par réplication *in vitro* (PCR) ou *in vivo* (transformation). Dans le cadre de l'identification de peptides bactéricides (non bactériolyques), l'objectif sera de récupérer les bactéries qui "se suicideront" à cause de l'effet antibiotique d'un peptide et d'extraire leur ADN plasmidique afin d'isoler les séquences génétiques encodant les peptides d'intérêt. La stratégie consiste en la répétition de cycles de quatre étapes (Figure 64) :

- Dans la première étape, les collections de bactéries sont cultivées en condition de forte répression de l'expression des peptides.
- Ensuite, après l'induction de l'expression et marquage fluorescent à l'iodure de propidium (IP), les bactéries mortes sont récupérées par un tri au cytomètre de flux. En effet, l'IP ne peut pénétrer les membranes des bactéries vivantes mais traverse les membranes fragilisées et dépolarisées des bactéries mortes[323].
- L'étape suivante est d'amplifier, *in vitro*, l'information génétique contenue dans les bactéries récupérées.
- Finalement, la dernière étape consiste à transformer de nouvelles bactéries avec le matériel sélectionné.

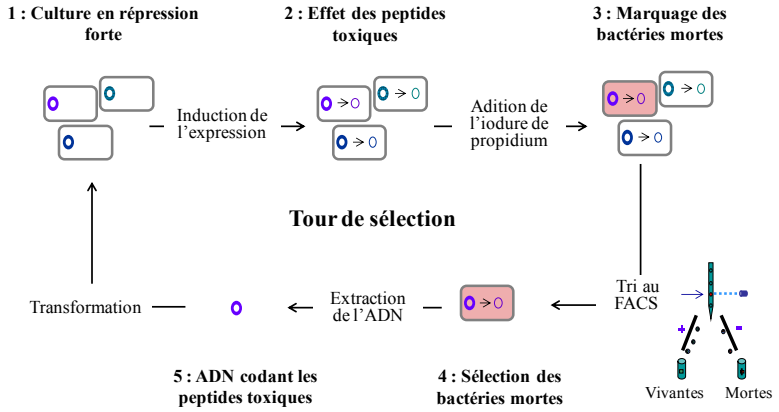


Figure 64 : Représentation schématique des étapes d'un tour de sélection de la stratégie de sélection "Anti-Darwinienne".

Avant de travailler avec les collections de peptides, les différentes étapes de cette méthode de sélection ont été testées sur un système modèle. L'efficacité de la stratégie a été déterminée en mesurant le facteur d'enrichissement après un tour de sélection. Une fois mise au point, la stratégie a été appliquée à la sélection de peptides d'intérêt au sein des collections.

Résultats

Mise au point de la stratégie sur un système modèle

Le système modèle est composé d'un mélange de deux souches d'E.coli TOP10 transformées chacune par un plasmide (pGIV012 ou pUC18Ery) portant un marqueur de résistance différent. Ces plasmides pGIV012 et pUC18Ery, reçus du professeur B. Hallet, dérivent tous les deux du plasmide pUC18 qui porte l'origine de réplication pMBI et le gène de résistance à l'ampicilline. Dans le cas de pGIV012, le gène de résistance à l'ampicilline a été remplacé par celui de la kanamycine et pour pUC18Ery par celui de l'érytromycine. La première souche contient le plasmide avec le gène de résistance à l'érythromycine (TOP10-Ery^R) et la seconde possède celui avec la kanamycine (TOP10-Kan^R).

Les bactéries sont cultivées sans antibiotique et l'ajout de kanamycine tue les bactéries sensibles, à savoir celles contenant le marqueur de résistance à l'érythromycine (TOP10-Ery^R) Le but étant de récupérer spécifiquement le plasmide pUC18Ery, les bactéries mortes et vivantes sont triées par cytométrie de flux (FACS) après marquage à l'IP. L'ADN plasmidique des bactéries récupérées est extrait et puis réintroduit dans des nouvelles bactéries par transformation. L'augmentation du rapport du nombre de bactéries résistantes respectivement à l'érythromycine (TOP10-Ery^R) et à la kanamycine (TOP10-Kan^R) après et avant le cycle de sélection reflète l'efficacité de la méthode. Les expériences ont été réalisées de manière à évaluer la puissance de la méthode de sélection, notamment en mesurant les facteurs d'enrichissement lorsque les bactéries cibles sont présentes en très faible proportion dans le mélange modèle.

Test de l'effet de l'antibiotique

L'effet de la kanamycine sur les deux types de bactéries a été testé en réalisant des étalements sur des boîtes de Pétri et des mesures de densité optique au cours du temps (Figure 65). Pour les mesures de densité optique, les bactéries des deux souches ont été cultivées séparément et la kanamycine à une concentration de 50 µg/ml a été ajoutée dans chacune des cultures lorsque la densité optique à 600 nm était d'environ 0,6. Pour les étalements sur boîtes de Pétri, un mélange des bactéries des deux souches a été réalisé avec une proportion de 10% de TOP10-Ery^R. Des échantillons de la culture dilués d'un facteur 10 à 10^5 ont été

étalés sur boites de Pétri avant et 1h après l'ajout de l'antibiotique dans les cultures.

La kanamycine, à une concentration de 50 µg/ml, cause la mort des bactéries sensibles (TOP10-Ery^R) 15 à 30 minutes après son addition dans le milieu de culture sans perturber la croissance des bactéries résistantes (TOP10-Kan^R). De plus, son effet est bactéricide mais non bactériolytique car la densité optique de la culture de TOP10-Ery^R n'augmente plus après l'ajout de l'antibiotique mais ne chute pas. L'addition d'un antibiotique est donc un bon modèle pour simuler l'effet de l'expression de peptides toxiques bactéricides.

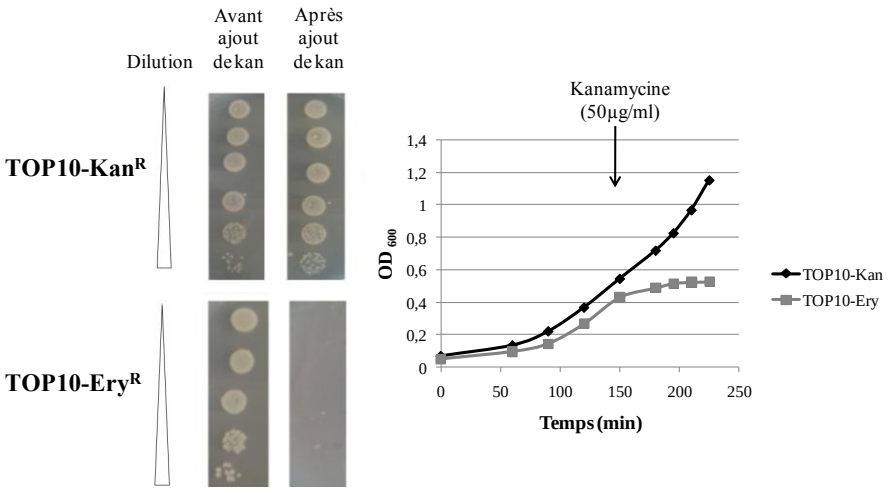


Figure 65 : Effet de l'ajout de kanamycine (50 µg/ml) dans une culture de bactéries sensibles (TOP10-Ery^R) et résistantes (TOP10-Kan^R).

Test du marquage à l'IP

Le marquage à l'IP (fluorescent dans le rouge) a été testé sur des cultures qui ont été traitées à la kanamycine 50 µg/ml et qui contiennent soit la souche TOP10-Kan^R ou TOP10-Ery^R, soit un mélange 50/50 afin de vérifier si ce marquage permet la discrimination des bactéries TOP10-Ery^R tuées à la kanamycine. Les bactéries ont été cultivées jusqu'à une une densité optique d'environ 0,6, la kanamycine a ensuite été ajoutée et après 1h d'incubation, les mesures au microscope ont été réalisées.

Le Syto9 (fluorescent dans le vert) a été utilisé comme contrôle positif car il est capable de pénétrer dans toutes les bactéries. Les images obtenues au

microscope à fluorescence (Figure 66) indiquent que l'IP (C_f = 30µM) pénètre bien les bactéries mortes et qu'après excitation à 480 nm, les bactéries émettent dans le rouge. Un bruit de fond de bactéries mortes (± 5%) est retrouvé dans l'échantillon des bactéries vivantes. De plus, les images indiquent que lorsque l'IP est présent dans la bactérie, celle-ci fluoresce moins dans le vert. Les deux marqueurs étant des agents intercalant de l'ADN, il pourrait y avoir compétition entre eux pour la liaison à l'ADN.

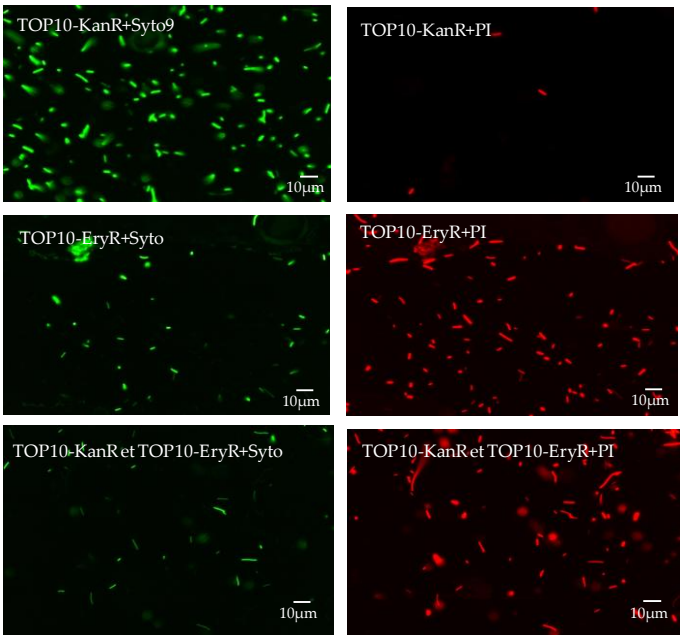


Figure 66 : Images au microscope à fluorescence des échantillons des cultures traités à la kanamycine 50 µg/ml de bactéries TOP10-Kan^R et TOP10-Ery^R ainsi que d'un mélange 50/50.

Test d'analyses au FACS

Avant de réaliser des expériences de tri au FACS, des analyses de mélanges de bactéries TOP10-kan^R et TOP10-Ery^R, traités en ajoutant de la kanamycine, ont été effectuées sur un FACS Calibur (spécialisé dans l'analyse plutôt que le tri). Des mélanges de bactéries contenant des proportions en bactéries TOP10-Ery^R de 0%, 1%, 10%, 50%, 80% et 100% ont été préparés et les cellules ont été marquées à l'IP (C_f = 30 µM). Un contrôle négatif a été également réalisé en passant au FACS des bactéries non marquées à l'IP. Les résultats sont présentés dans le Tableau 12 et la Figure 67 ci-dessous.

Une certaine proportion de bactéries fluorescentes, correspondant à environ 15%, sont retrouvées dans l'échantillon des bactéries TOP10-Kan^R (Tableau 12). De la même manière, une proportion de bactéries non-fluorescente, correspondant à environs 12%, sont retrouvées dans l'échantillon des bactéries TOP10-Ery^R (Tableau 12). Cependant, on peut voir clairement un accroissement de la hauteur du pic correspondant aux bactéries fluorescentes lorsque le pourcentage en bactéries TOP10-Ery^R augmente dans les mélanges (Figure 67). Ces résultats sont encourageants car ils indiquent que le FACS est capable de discriminer les bactéries marquées à l'IP de celles non marquées.

	% bactéries marquées	% bactéries non marquées
Bactéries non marquées	0	100
TOP10-Ery^R	88	12
TOP10-Kan^R	15	85
Mélange 1%	17	83
Mélange 10%	21	79
Mélange 50%	42	58
Mélange 80%	70	30

Tableau 12 : Présentation des résultats du FACS calibur obtenus sur les bactéries TOP10-Kan^R et TOP10-Ery^R, ainsi que les mélanges contenant un pourcentage croissant en bactéries TOP10-Ery^R (en présence de kanamycine 50 µg/ml).

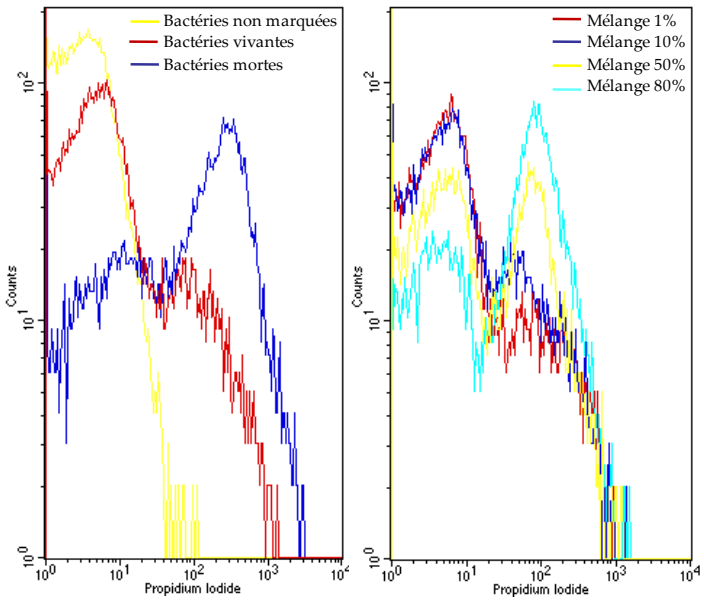


Figure 67 : Présentation des résultats du FACS calibur obtenus sur les bactéries TOP10-Kan^R et TOP10-Ery^R, ainsi que les mélanges contenant un pourcentage croissant en bactéries TOP10-Ery^R (en présence de kanamycine 50 µg/ml).

Test de tri au FACS

Des expériences de tri ont ensuite été réalisées sur un FACS Vantage (Louvain en Woluwé, Institut Ludwig, A. Tonon). Ce FACS est capable de distinguer deux populations différentes dans un mélange, de les trier et finalement les récolter dans deux tubes collecteurs différents. Il est donc théoriquement possible de récupérer sélectivement les bactéries fluorescentes. Les expériences ont été réalisées avec des bactéries en proportion de 99 TOP10-Kan^R /1 TOP10-Ery^R. En effet, l'objectif est de tester la puissance de sélection du FACS en effectuant des expériences dans lesquelles les bactéries cibles, c'est-à-dire les bactéries mortes sont en faible proportion. Les expériences ont été répétées trois fois pour tester la reproductibilité.

Les mélanges de bactéries ont été préparés et cultivés jusqu'à une densité optique d'environ 0,6. Une fraction de chaque culture a été étalée sur des boîtes de Pétri Kan50 et Ery250 avant l'ajout de l'antibiotique pour déterminer la proportion de départ des bactéries TOP10-Ery^R et TOP10-Kan^R dans le mélange (Tableau 13, point 1). La kanamycine 50 µg/ml a ensuite été ajoutée et après 1h d'incubation, les cellules ont été lavées, resuspendues dans le tampon de coloration et marquées avec l'IP (Cf = 30 µM). Finalement, les échantillons de bactéries, à une densité optique de 1, ont été analysés et triés au FACS pendant 30 minutes. Comme le FACS Vantage est capable d'analyser et trier 20000 événements/seconde, environ 2.10^7 cellules ont donc été analysées et réparties durant le tri dans les deux tubes collecteurs. Pour les deux mélanges, l'affichage du FACS a confirmé que le nombre total de 2.10^7 bactéries ont bien été analysées et triées. Dans chaque tube collecteur obtenu, une fraction des cellules a été ré-analysée au FACS Vantage et étalée sur des boîtes de Pétri kan50 et Ery250 pour contrôler le tri (Tableau 13, point 2).

Pour déterminer le facteur d'enrichissement, la nouvelle proportion en bactéries TOP10-Ery^R et TOP10-Kan^R présente dans le tube collecteur des cellules triées comme marquées a été calculée. Pour ce faire, une extraction plasmidique a été effectuée. Pour atteindre un nombre de cellules total de 10^9 nécessaire pour l'extraction plasmidique, des cellules contenant un plasmide "carrier" portant un gène de sélection différent (tet^R) ont été ajoutées. Ces plasmides "carrier" ont ensuite été éliminés de l'échantillon par traitement de l'ADN avec une enzyme de restriction qui possède des sites de reconnaissance dans ce plasmide mais pas

dans les vecteurs d'intérêts. L'ADN a été électroporé dans des cellules TOP10 et finalement, les cellules ont été étalées sur des milieux kan50 et Ery250 (Tableau 13, point 4).

	Exp 1	**Exp 2**	**Exp 3**
1. Etalements avant le tri au FACS			
Nb de bactéries TOP10-Kan^R	$8{,}4.10^7$	$8{,}3.10^7$	$3{,}4.10^7$
Nb de bactéries TOP10-Ery^R	$6{,}1.10^5$	$2{,}8.10^6$	$3{,}4.10^5$
% TOP10-Ery^R/ TOP10-Kan^R	0,8%	3,4%	1%
2. Analyses au FACS avant et après le tri			
% des marquées dans mélange de départ	1,1	1,3	4,2
% des marquées dans tube des marquées après tri	18	ND	ND
% des marquées dans tube des non marquées après tri	0,04	ND	ND
Nb total cellules récoltées dans tube des marquées	$1{,}5.10^5$	$2{,}5.10^5$	6.10^5
Nb total cellules récoltées dans tube des non marquées	$1{,}5.10^7$	$1{,}5.10^7$	$1{,}5.10^7$
% cellules récultées dans tube marquées/ non marquées	1%	1,7%	4%
3. Etalements après le tri			
Nb TOP10-Ery^R vivantes dans tube des marquées	0	0	0
Nb TOP10-Ery^R vivantes dans tube des non marquées	0	0	0
Nb TOP10-Kan^R vivantes dans tube des marquées	$6{,}8.10^4$	$2{,}9.10^4$	$7{,}2.10^3$
Nb TOP10-Kan^R vivantes dans tube des non marquées	$3{,}7.10^6$	$1{,}2.10^7$	$2{,}1.10^6$
4. Etalements après électroporation			
Nb TOP10-Kan^R dans tube des bactéries marquées	ND	$1{,}5.10^4$	$1{,}1.10^5$ $2{,}9.10^4$
Nb TOP10-Ery^R dans tube des bactéries marquées	ND	$4{,}2.10^3$	$1{,}9.10^4$ $8{,}8.10^3$
% TOP10-Ery^R/TOP10-Kan^R	ND	28%	17% 30%
Nb TOP10-Kan^R dans tube des bactéries non marquées	ND	$1{,}6.10^5$	$5{,}7.10^5$ $5{,}7.10^5$
Nb TOP10-Ery^R dans tube des bactéries non marquées	ND	$4{,}6.10^3$	$1{,}1.10^4$ 0
% TOP10-Ery^R/TOP10-Kan^R	ND	2,8%	1,9% 0%
5. Facteur d'enrichissement dans tube des bactéries marquées	**ND**	**8,3**	**17** **30**

Tableau 13 : Résultats obtenus pour les trois expériences de tri au FACS. Le facteur d'enrichissement est calculé est divisant la proportion de bactéries TOP10-Ery^R/TOP10-Kan^R obtenu dans le tube des bactéries marquées par la proportion de bactéries TOP10-Ery^R/TOP10-Kan^R dans le mélange de départ.

Les analyses aux FACS sont reproductibles, le pourcentage de bactéries marquées dans l'échantillon de départ indiqué par le FACS est compris entre 1% et 4% (Tableau 13, point 2à). Après le tri, une augmentation d'un facteur 16 de la proportion en bactéries fluorescentes est observée dans le tube des cellules triées comme marquées (Tableau 13, point 2). De plus, la proportion en bactéries fluorescentes présente dans le tube des cellules triées comme non marquées est faible. Ces résultats indiqueraient que le tri des bactéries non fluorescentes est plus efficace que celui des fluorescentes.

Les résultats des étalements sur boîtes de Pétri kan50 effectués directement après le tri au FACS ne sont quant à eux que faiblement reproductibles (Tableau 13, point 3). Bien que le FACS ait indiqué qu'un nombre équivalent de cellules a été récolté dans les tubes collecteur de bactéries non marquées chaque expérience (Tableau 13, point 2), le nombre de bactéries TOP10-Kan^R obtenu par étalement est fort variable d'une expérience à l'autre. Ce nombre est compris entre 20% et 80% du nombre indiqué par le FACS (Tableau 13, comparaison point 2 et 3). Deux façons d'interpréter ces résultats seraient qu'en fonction des conditions expérimentales, un nombre plus ou moins important de bactéries meurent dans le laps de temps entre le tri et l'étalement ou que le FACS prend en compte dans son analyse des événements qui ne sont pas des bactéries. Un pourcentage de bactéries vivantes TOP10-Kan^R est retrouvé dans le tube des cellules triées comme marquées. Ce qui indiquerait que les bactéries fluorescentes entrainent avec elles des bactéries non marquées lors du tri.

Après l'extraction de l'ADN des cellules présentes dans chaque tube et l'électroporation, des problèmes ont été rencontrés avec les bactéries TOP10-Ery^R. Elles ont du mal à pousser et la taille des colonies n'est pas homogène sur les boîtes de Pétri. Malgré cela, la nouvelle proportion de bactéries TOP10-Ery^R et TOP10-Kan^R présente dans chaque tube collecteur a pu être déterminée pour chaque expérience (Tableau 13, point 5). Des facteurs d'enrichissement compris entre 8 et 30 ont été obtenus.

Le Tableau 14 présente les facteurs d'enrichissement théoriques que l'on obtiendrait dans les conditions où le tri dans les deux tubes collecteur serait réalisé avec une efficacité de 100%. Le facteur diminue fortement avec l'augmentation du pourcentage en bactéries TOP10-kan^R non désirées présentes dans le tube collecteur des cellules triées comme marquées. En présence d'un

bruit de fond de 5% de bactéries mortes TOP10-Kan^R non désirées, comme nous l'avions observé lors des expériences au microscope, un facteur théorique de 20 est obtenu.

Bactérie	**% dans le mélange**	**% de bactéries mortes**					
TOP10-Kan^R	99	0	1	2	3	4	5
TOP10-Ery^R	1	1	1	1	1	1	1
Facteurs d'enrichissement		**∞**	**99**	**49,5**	**33**	**24,7**	**19,8**

Tableau 14 : Facteurs d'enrichissement théoriques obtenus dans des conditions où la sélection est optimale pour des mélanges contenant 1% de cellules mortes d'intérêt mais avec un pourcentage croissant de bactéries mortes sans intérêt qui forment un bruit de fond.

Les facteurs d'enrichissement expériementaux que nous avons obtenus, bien qu'ils soient faibles, sont encourageants pour la stratégie car ils montrent qu'il y a bien sélection. Avec nos facteurs d'enrichissement compris entre 10 et 30, on devrait théoriquement pouvoir sélectionner un peptide toxique dans une collection de 1 million après 4 à 6 tours de sélection pour peu que la récupération des plasmides soit efficace.

Test de mise au point de la récupération des plasmides après le tri

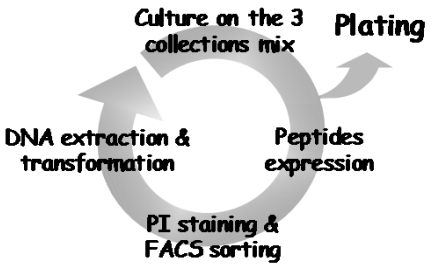
La stratégie de sélection nécessite une méthode de récupération efficace des plasmides présents dans les bactéries triées comme fluorescentes et donc mortes. Or, ces cellules sont fortement diluées et en très faible concentration dans l'échantillon après le tri.

Dans l'expérience modèle, les bactéries fluorescentes étaient récupérées à une concentration approximative de 2.10^5/ml. Or, comme on travaille avec un vecteur possédant une origine de réplication ColE1 et qui possède donc environ 100 copies de plasmide par cellule[63], on se trouve avec un échantillon contenant 2.10^7 molécules de vecteur/ml. C'est une concentration beaucoup trop faible pour pouvoir réaliser une extraction plasmidique à l'aide d'un kit de purification commercialisé. La méthode de l'ajout de cellules portant un plasmide "carrier" a été utilisée Ces cellules ont été ajoutées juste avant l'étape de centrifugation des bactéries. Après l'extraction, 100µl d'ADN ont été récupérés et 10 µl ont été électroporés dans des cellules TOP10. Un nombre de bactéries d'environ 1.10^4 pour les 100µl d'ADN a été récupéré par cette méthode, et correspond à 10% des cellules triées comme fluorescentes. Au niveau des plasmides, cela donne un rendement de 0,1%. Ce rendement est faible mais est fortement dépendant de

l'efficacité de la purification et de la transformation. Une expérience contrôle a été réalisée dans laquelle une extraction plasmidique a été effectuée sur un échantillon contenant 1 ml de bactéries du mélange de départ avant le tri (1.10^8 bactéries/ml) et 1 ml de cellules contenant le plasmide "carrier". Un nombre de bactéries d'environ 1.10^7 pour les 100µl d'ADN a été récupéré par cette méthode, et correspond également à 10% des cellules de départ. Ce qui indiquerait que dans nos conditions expérimentales, un rendement supérieur ne peut pas être obtenu.

La méthode d'amplification de l'ADN par PCR a été également testée pour l'expérience modèle. La PCR directement sur les cellules n'a pas été évaluée. Des PCR sur l'ADN récolté des préparations de plasmides ont été effectuées. A partir d'1 µl d'ADN des 100 µl obtenus par l'extraction et en travaillant avec différentes polymérases et conditions de PCR, nous n'avons pas réussi à amplifier la séquence du gène de résistance à l'érythromycine.

Application de la stratégie de sélection sur 3 collections de peptides cycliques avec expression cytoplasmique.



Deux tours de sélection ont été réalisés sur les collections d'hexa-, hepta- et octa- peptides avec expression cytoplasmique (construites par J.M. Léonard) dans le but d'identifier des peptides bactéricides. Ces tours ont été réalisés en induisant l'expression des peptides en milieu solide. En effet, le confinement des bactéries en colonie devrait faciliter l'action des peptides capables d'exercer une toxicité en *trans*. De plus, cela permet de conserver le lien génotype-phénotype. Les banques ont été diluées et mélangées de façon à ce que, pour chaque banque, le nombre de bactéries soit 10x supérieur à la diversité de la banque (Tableau 15). Les ~ 5.10^8 bactéries TOP10-pCL6-8 ainsi obtenues ont été étalées au dessus d'une membrane de nitrocellulose sur une boîte de Pétri contenant de la tétracycline 7,5 µg/ml et du glucose 1%. La boîte a alors été incubée une nuit à 37°C pour permettre la formation d'un tapis homogène de cellules. L'induction a été enclenchée en transférant la membrane de nitrocellulose sur une boîte de Pétri contenant de l'arabinose 0,2% et en incubant

2h30 à 30°C. Les cellules ont été raclées, lavées et finalement mises dans le tampon de coloration pour atteindre une densité optique de 1.

	Concentration des bactéries	Diversité des banques	Dilution
pCL_4	10^{11}/ml	$1.5\ 10^5$	10^{-5}
pCL_5	10^{11}/ml	$3.2\ 10^6$	10^{-4}
pCL_6	10^{11}/ml	$5.6\ 10^7$	10^{-3}

Tableau 15 : Dilutions réalisées pour chaque collection cytoplasmique (pCL4, pCL5, pCL6) pour que le nombre de bactéries soit 10x supérieur à la diversité des banques.

Un échantillon de cellules TOP10-pCL4-6 a été prélevé pour réaliser une analyse SDS-PAGE des protéines totales après induction à l'arabinose. Des cellules non induites ont été également analysées et utilisées comme contrôle négatif (Figure 68). La bande correspondant au fragment d'intéine $DnaBI_N$ (~12 kDa) est présente dans l'échantillon après induction à l'arabinose et absente dans le contrôle négatif. Cela indique que l'induction en milieu solide par transfert de bactéries via une membrane de nitrocellulose permet donc l'expression des précurseurs et un épissage probable des peptides.

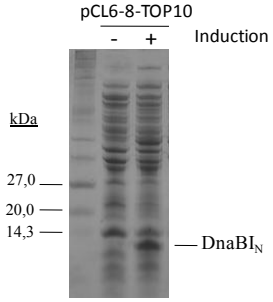


Figure 68 : Analyse SDS-PAGE de l'induction à l'arabinose 0,2% ou de la répression au glucose 1% en milieu solide de cellules TOP10-pCL6-8.

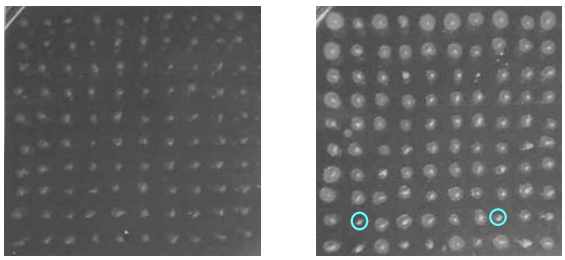
Après ajout de l'iodure de propidium (C_f = 30 µM), les cellules ont été triées au FACS Vantage. Ce FACS est capable d'analyser et trier 2.10^8 cellules en 2 heures. L'expérience a donc été réalisée en 4 fois pour trier un nombre total de 8.10^8 cellules. Le pourcentage de bactéries fluorescentes dans les échantillons de départ est de 4% et il augmente pour atteindre 6,5% après un tour de sélection. L'extraction d'ADN sur les fractions de cellules triées comme fluorescentes (mortes) après le FACS a été réalisée en présence de cellules contenant un plasmide "carrier". Après restriction de l'ADN "carrier" par une enzyme de restriction ne possédant pas de site de reconnaissance dans le plasmide d'intérêt, l'ADN a été finalement électroporé dans de nouvelles cellules TOP10 d'*E. coli*. Le

nombre total de bactéries triées et récoltées indiqué sur l'affichage du FACS, et le nombre de transformants obtenus après électroporation sont présentés dans le Tableau 16. Comme pour l'expérience modèle, le nombre de transformants est 100x inférieur au nombre de cellules indiqué par le FACS pour le tour 1. Pour le tour 2, il y a un facteur 1000. Une expérience a été réalisée pour vérifier si le nombre de transformants ne pourrait pas être augmenté en électroporant l'ADN dans des cellules compétentes traitées au glucose. En effet, l'ADN ayant été sélectionné car il cause la mort de la bactérie hôte, l'électroporation doit être réalisée dans des conditions de répression maximale de l'expression du peptide. Cependant, un nombre de transformants du même ordre de grandeur a été obtenu.

	Tour 1	**Tour 2**
Nbre total de bactéries triées (FACS)	$8\ 10^8$	$5\ 10^7$
Nbre de bactéries récoltées comme fluorescentes (FACS)	$2\ 10^7$	$2{,}5\ 10^6$
Nbre de transformants	$2\ 10^5$	$2{,}4\ 10^3$

Tableau 16 : Comparaison entre le tour 1 et le tour 2 du nombre de bactéries triées et récoltées comme fluorescentes indiqué par l'affichage du FACS, et du nombre de transformants.

Des repiquages sur boîte de Pétri contenant du glucose 1% (répression) ou de l'arabinose 0,2% (induction) ont été effectués sur 100 clones isolés après le second tour de sélection (Figure 69) et sur 100 clones des collections de départ comme contrôle. Sur arabinose, les 100 clones contrôles poussent normalement. Par contre, 2 clones forment des colonies plus petites et semblent avoir plus de mal à pousser parmi les 100 clones isolés après le second tour de sélection. Ces clones ont été nommés pCT22 pour celui de gauche et pCT24 pour celui de droite. Ces deux clones proviennent de la collection d'octapeptides et ont comme séquence peptidique SCMLPCRG pour le clone pCT22 et SIDLGYDG pour le clone pCT24.



Milieu de répression
Glucose 1%

Milieu d'induction
Arabinose 0,2%

Figure 69 : Réplicas sur boîtes de Pétri en milieu d'induction (arabinose 0,2%) ou de répression (glucose 1%) de 100 clones isolés après le second tour de sélection.

Des analyses par SDS-PAGE (Figure 70) des protéines présentes dans l'extrait cytoplasmique de cultures de ces deux clones en condition de répression et d'induction ont permis de mettre en évidence la présence de l'intéine $DnaBI_N$ et donc un épissage probable des peptides. La confirmation de l'épissage et de la cyclisation par spectrométrie de masse n'a pas été réalisée. En effet, des expériences effectuées par J.M. Léonard ont montré que la détection des peptides est moins évidente dans le milieu complexe cytoplasmique que dans le milieu périplasmique. Un peptide cyclique dont le signal est significativement plus élevé que le bruit de fond n'a jamais été détecté dans le cytoplasme (voir travail de thèse J.M. Léonard).

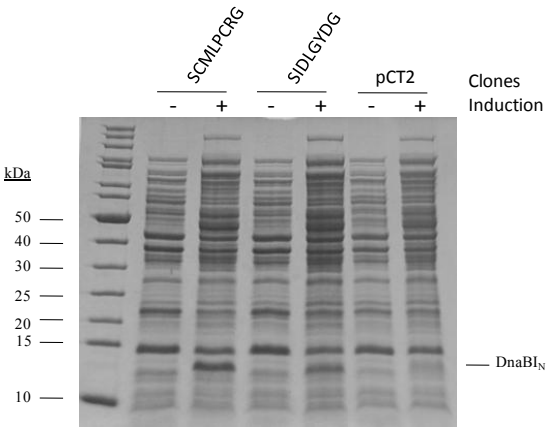


Figure 70 : Comparaison des extraits cytoplasmiques obtenus à partir des cultures induites et réprimées de clones identifiés après 2 tours de sélection. Une culture de l'ensemble des clones du tour T2 a été utilisée comme contrôle.

Des étalements sur boîtes de Pétri et des mesures de la densité optique ont été réalisées sur des cultures de ces deux clones en conditions de répression ou d'induction afin de mesurer l'effet de l'expression du peptide sur la croissance bactérienne. Les bactéries ont été d'abord cultivées en présence de glucose 1 % jusqu'à une densité optique de 0,6. Après centrifugation et incubation à 37°C pendant 30 mintues dans du nouveau milieu de culture, le glucose 1 % ou l'arabinose 0,2 % ont été ajoutés dans les échantillons. Pour les étalements, même si le nombre de colonies observé est identique, les colonies sont plus petites et transparentes sur le milieu arabinose par rapport au glucose. Pour les densités optiques (Figure 71), la courbe de croissance des cultures en condition d'induction ne sont pas significativement différentes de celles contrôles en condition de répression ni de celles de la culture contrôle exprimant un peptide

non toxique pC6-1. On observe cependant que la culture atteint plus rapidement la phase stationnaire en condition d'induction pour les deux clones.

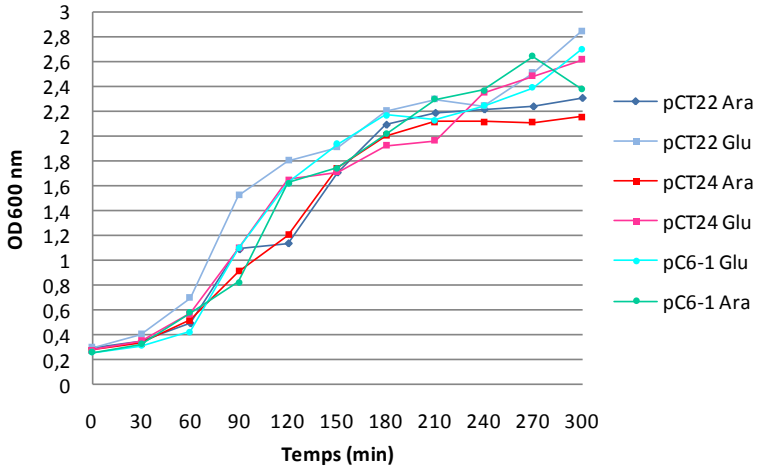


Figure 71 : Mesure de la densité optique au cours du temps de cultures des clones pCT22 et pCT24 en condition d'induction (Ara) et de répression (Glu). Le glucose 1 % ou l'arabinose 0,2 % ont été ajoutés après 30 minutes dans les échantillons. Une culture de cellules exprimant un peptide non toxique a été utilisée comme contrôle pC6-1.

Des cultures des clones pCT22 et pCT24 en condition de répression et d'induction ont été marquées à l'IP (C_f = 30 µM) et analysées au FACS Calibur (Figure 72). Les bactéries obtenues après le second tour de sélection (pCT2) ainsi que celles de la collection d'octapeptides pCL8 ont été également analysées et utilisées comme contrôles. En conditions de répression, le pourcentage de bactéries marquées se trouve autour des 5%. En conditions d'induction, le pourcentage augmente dans les quatre échantillons pour atteindre les 20% et est légèrement plus élevé pour les deux clones sélectionnés.

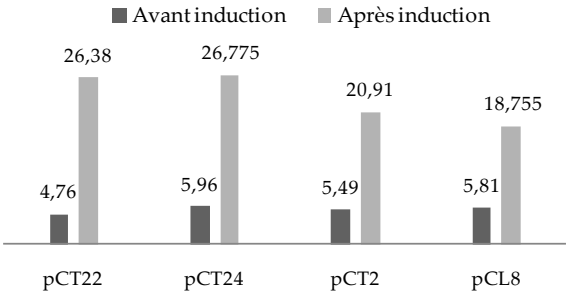


Figure 72 : Analyses au FACS Calibur du pourcentage de bactéries marquées à l'iodure de propidium avant et après induction à l'arabinose pour le clone pCT22 et pCT24, ainsi que les bactéries obtenues après le second tour de sélection et de la collection pCL8.

Discussion

Dans ce travail, nous avons réalisé les premières mises au point d'une stratégie de sélection novatrice qui vise à identifier, dans des collections de peptides cycliques, ceux qui causent la mort de leur bactérie hôte. Ces mises au point ont été effectuées sur un système modèle et ont eu pour objectif de tester les différentes étapes du protocole. Deux tours de sélections ont été réalisés sur les collections de peptides cycliques cytoplasmiques et ont permis d'identifier deux clones dont la croissance est perturbée en condition d'induction.

Par des mesures au microscope à fluorescence et des analyses au FACS, nous avons démontré que l'IP est un bon choix de molécule fluorescente pour marquer les bactéries mortes. Lorsque des mélanges de populations de bactéries mortes et vivantes sont analysées au FACS, le pic des bactéries fluorescentes se distingue clairement des cellules non marquées.

Lors des expériences au FACS, nous avons remarqué que le tri des bactéries non fluorescentes est plus efficace que celui des fluorescentes. L'hypothèse avancée serait que les bactéries s'associent en groupe en passant devant le laser. Le marquage ne permettant pas de discriminer les bactéries vivantes, les groupes contenant un mélange de bactéries mortes et vivantes seraient dirigés vers le récipient collecteur des bactéries fluorescentes. Une solution serait de travailler avec un deuxième marquage fluorescent spécifique des bactéries vivantes. L'utilisation de la fluorescéine pourrait être envisagée car c'est un marqueur fluorescent qui s'échappe dans le milieu extracellulaire quand la viabilité cellulaire est compromise[324]. On pourrait également utiliser des cellules d'E. coli transformées pas un plasmide exprimant la GFP[325].

La stratégie de récupération du matériel génétique présent dans les bactéries mortes après le tri doit encore être optimisée. En effet, un nombre de bactéries correspondant à 1% du nombre de cellules considérées comme fluorescente sur l'affichage du FACS ont été récupérées. Il faudrait pouvoir arriver à augmenter ce pourcentage d'un facteur 10, voire 100 pour que la diversité des cellules triées comme fluorescentes soit représentée. Une première amélioration pourrait être apportée à la stratégie que nous avons utilisée qui consisterait à ajouter les cellules portant le plasmide "carrier" directement dans le tube récolteur avant le tri. De cette façon, on éviterait que les cellules en très faible concentration ne

soient pas récupérables car elles seraient collées aux parois du récipient. Une autre solution serait d'ajouter de la BSA.

Des expériences plus approfondies pour tenter d'amplifier l'ADN par PCR devraient également être réalisées avec les polymérases plus performantes commercialisées actuellement. La manipulation est très délicate car l'ADN d'intérêt se trouve en très faible concentration et est, de plus, mélangé avec du plasmide carrier qui pourrait perturber l'hybridation des amorces. Si l'on récupère 10^6 cellules fluorescentes contenant chacune 100 copies du plasmide d'intérêt après le tri au FACS et que l'on réalise l'extraction plasmidique de manière optimale, on aurait dans notre échantillon $6{,}5.10^{-4}$ ng d'ADN/µl. Or, les réactions de PCR sont généralement réalisées avec des concentrations en ADN allant de 0,1 à 10 ng. De plus, l'inconvénient de cette méthode est que les fragments d'ADN doivent être reclonés dans un vecteur avant de pouvoir être électroporés.

La difficulté principale de la stratégie est donc le très faible nombre de bactéries fluorescentes récoltées après le tri au FACS. Etant donné que la vitesse du tri est limitée à 20000 événements/seconde et que les bactéries d'intérêts ne se trouvent qu'à une proportion de 4% dans le mélange, un tri de 8h ne permet de récupérer que 2.10^6 bactéries dans 1 ml. Ce nombre de bactéries ne peut pas être augmenté facilement et est un des obstacles majeurs de la stratégie. Une perspective serait d'utiliser un FACS plus performant au niveau de l'analyse et du tri. Le MoFlo, par exemple, permet d'atteindre des vitesses de 200000 événements/seconde.

Les expériences de sélection réalisées sur les trois banques cytoplasmiques ont confirmé les difficultés à récupérer un nombre suffisant de bactéries après le tri. Cependant, après deux tours de sélection et des repiquages, deux clones dont la croissance cellulaire est ralentie en conditions d'induction ont été identifiés. L'effet négatif de l'induction du peptide sur la croissance de la bactérie a été observé par différentes expériences.

III. Conclusion et perspectives

Durant ces quatre années de thèse, le travail réalisé a été consacré au développement de trois stratégies novatrices pour tenter d'identifier des peptides cycliques biosynthétiques ayant une activité antibiotique causant la mort de leur bactérie hôte.

La première stratégie visait à créer un système permettant une répression totale de l'expression des peptides cycliques. En effet, dans le contexte de la recherche de peptides causant la mort de leur hôte, un tel système est indispensable si on veut éviter de perdre rapidement les peptides les plus intéressants par sélection *in vivo*. Or, comme nous l'avons vu dans l'introduction, les systèmes d'expression commercialisés actuellement présentent toujours une expression basale même avec les promoteurs les plus fortement réprimés. La fonctionnalité de notre système de répression totale, basé sur une reconstitution de gène par double recombinaison homologue et site spécifique, a été prouvée. Cependant, son efficacité est 100x inférieure à celle du système antérieurement décrit par A. Brédo et il ne permet de reconstituer les gènes dans les meilleures conditions que dans 6.10^{-4} des bactéries. A cause de sa faible efficacité, cette stratégie n'a pas pu être utilisée pour la régulation de l'expression des collections de peptides cycliques. Elle pourrait tout de même mener à des applications d'intérêt au niveau biotechnologique de part sa particularité de permettre l'échange de fragments d'ADN entre deux vecteurs. Un exemple intéressant et prometteur d'application serait l'assemblage d'opérons de grande taille *in vivo*.

La seconde stratégie a été de créer une méthode générale pour la production de collections de peptides cycliques dans le périplasme d'*E. coli*. Cette méthode permettrait de donner l'accès à plusieurs millions de molécules différentes et stables à ce milieu périplasmique très spécifique et unique aux bactéries. Après un échec en combinant la technologie SICLOPPS à une voie de transport post-traductionnelle, l'objectif a été atteint en combinant la même technologie à une voie de transport co-traductionnelle. Cette approche constitue une grande avancée technologique car elle permet de réaliser deux modifications post-traductionnelle d'un peptide, l'exportation et la cyclisation, à partir d'un précurseur.

Une collection d'hexa-, hepta- et octa-peptides cycliques de 8 à 10 millions de composés différents a été créée et pour un grande nombre de clones, les analyses au spectromètre de masse des extraits périplasmiques ont permis de détecter le

peptide cyclique avec un rapport signal/bruit relativement haut, ce qui suggère que l'exportation et l'épissage s'effectuent avec une efficacité élevée. Des précurseurs non épissés ont cependant été observés pour la plupart des clones et leur analyse par spectrométrie de masse a permis de faire une découverte inattendue et enthousiasmante. Ces précurseurs sont inactivés par glutathionylation de leur cystéine, une modification encore très peu étudiée dans le périplasme. Il serait interessant dans la suite du travail de confirmer ce résultat et de réaliser des expériences pour tenter de déterminer si la glutathionylation joue un rôle dans la régulation des réactions d'oxydo-réductions ou dans le système de réduction qui protège de l'oxydation les protéines qui possèdent un nombre impair de résidu cystéine. Une première approche envisageable serait de déterminer si l'inactivation du précurseur par oxydation est toujours observée dans différentes souches d'*E. coli* dépourvues, par exemple, de la glutaredoxine ou de DsbA ou de DsbG ou de la γ-glutamylcystéine synthase qui est la première protéine impliqué dans la voie du biosynthèse du glutathion. On pourrait également réaliser des analyses pulse-chase avec du glutathion marqué pour déterminer si le précurseur modifié par le glutathion reste oxydé ou s'il est réduit par une enzyme pour pouvoir réaliser la réaction d'épissage.

La localisation dans le périplasme des peptides cycliques permet d'ouvrir des portes à de nouvelles applications pour la recherche de composés biologiquement actifs. En effet, les peptides cycliques ont accès et peuvent cibler des protéines membranaires ou exportées. Des cibles intéressantes sont, par exemple, les enzymes impliquées dans la biosynthèse du peptidoglycane, les pompes d'efflux et protéines impliquées dans les méchanismes de résistance aux antibiotiques, les transporteurs membranaires ABC,.... Pour permettre l'identification des peptides inhibiteurs de ces cibles dans les banques, des méthodes originales de sélection ou de criblage à haut débit devront être mises au point.

Une première méthode qui est actuellement développée au laboratoire est de co-exprimer dans *E. coli* les banques de peptides avec une protéine recombinante périplasmique soluble ou membranaire dont l'activité peut être liée à une toxicité. Les peptides inhibiteurs capables d'annihiler la toxicité permettront à la bactérie d'avoir une croissance normale et d'être sélectionnée *in vivo* (thèse S.

Garcia). La protéine cible peut potentiellement provenir de tout organisme d'intérêt procaryote ou eucaryote pour autant que son expression de façon fonctionnelle soit possible dans *E. coli*.

Une autre méthode qui est également utilisée au laboratoire actuellement est de rechercher des peptides cycliques disrupteurs d'interaction protéine-protéine (mémoire A. Martinez Carratalia). En effet, en liant cette intéraction à un signal de fluorescence grâce au système "BIFC" (Bimolecular fluorescence complementation, Figure 73) et en triant au FACS les bactéries de la banque en fonction de la fluorescence, les clones exprimant des peptides disrupteurs de l'interaction pourront être récupérés pour leur propriété à faire disparaitre le signal. Le système BICF est basé sur la fusion des fragments N- et C-terminaux d'une protéine fluorescente sur une protéine A et B respectivement. En l'absence d'interaction, le fluorophore reste non fonctionnel. L'interaction de ces protéines apporte les fragments fluorescents à proximité, permettant à la protéine rapporteur de prendre sa structure tridimensionnelle native et émettre son signal fluorescent. La même méthode pourrait également être utilisée en liant l'intéraction à un signal FRET.

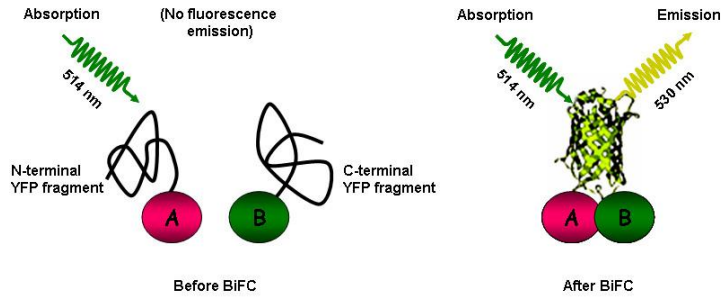


Figure 73 : Représentation schématique du système BIFC (Bimolecular fluorescence complementation). Les protéines A et B sont fusionnées aux fragments N et C terminaux de la YFP respectivement. En l'absence d'interaction entre A et B, le fluorophore reste non fonctionnel. À la suite de l'interaction entre A et B, un fluorophore fonctionnelle est reconstitué qui émet de la fluorescence après excitation avec une longueur d'onde appropriée.

Une dernière méthode, qui est actuellement étudiée au laboratoire et qui est inspirée des méthodes sélections anti-darwiniennes décrites précédemment dans la thèse, permettrait de sélectionner des peptides inhibiteurs de protéines impliqués dans la résistance aux antibiotiques. Elle consiste à sélectionner les peptides qui provoquent la mort bactérienne en combinaison avec un

antibiotique parce qu'ils agissent en inhibant la protéine responsable de la résistance. La stratégie de sélection basée sur la lyse bactérienne dont l'efficacité a été démontrée par J. M. Léonard est actuellemnet utilisée par M. Doumit en combinaison avec l'antibiotique oxacyline pour identifier des peptides inhibiteurs de la pompe d'efflux TolC. Des expériences de sélection ont également été réalisées en combinaison avec l'ampicilline pour tenter d'indentifier des inhibiteurs de la β-lactamase (thèse J.M. Léonard). A long terme, ces peptides inhibiteurs pourraient être utilisés comme complément aux antibiotiques pour lutter contre le phénomène de résistance.

L'exportation de peptides cycliques dans le périplasme représente la première étape vers la sécrétion de composés biosynthétiques dans le milieu extracellulaire. Il serait donc interessant dans la suite du projet d'étudier les différentes stratégies d'exportartion de protéine recombinante dans E. coli qui ont été développées par ingénierie et qui sont décrites dans la litérature[326] pour tenter d'exporter les peptides dans le milieu extracellulaire. Ces stratégies sont basées, par exemple, sur l'ingénierie de systèmes d'exportation existant naturellement dans d'*E. coli* , ou sur la modification de la perméabilité de la membrane externe ou la co-expression de protéine induisant la lyse spécifique. Une piste actuellement évaluée au laboratoire est d'utiliser la protéine pIV du phage filamenteux. En effet, cette protéine forment des multimères et constituent un canal dans la membrane externe qui permet l'évacuation des phages dans le milieu extracellulaire[327]. De plus, une étude réalisée en 1999 par l'équipe de S. Sanford a permis d'isoler un mutant fonctionnel de la pIV qui rend *E. coli* sensible à la vancomycine, un antibiotique hydrophile qui normalement ne traverse pas la membrane externe[328]. Les premières expériences de co-expression de ce pore muté avec un peptide cyclique périplasmique ont permis de détecter la présence du peptide dans le milieu extracellulaire par spectrométrie de masse (mémoire de graduat B. Michiels). Des stratégies d'identification de peptides antibiotiques exprimés dans *E. coli* qui auraient une activité en trans sur un autre organisme vont pouvoir être envisagées. Il faudrait dans ce cas réaliser des co-cultures en milieu solide d'*E. coli* transformées par les banques de peptides avec l'organisme cible et observer si sa croissance est perturbée en condition d'expression des peptides. La cible pourrait être tout organisme capable d'être co-cultivé avec E. *coli*, une bactérie pathogène, une levure, un mycète…

La dernière stratégie, sur laquelle nous avons travaillé dans le cadre de la thèse, a été de mettre au point une méthode de sélection "anti-darwinienne" pour identifier, dans les collections de peptides cycliques cytoplasmiques et périplasmiques créées au laboratoire, ceux qui ont un effet bactéricide contre leur hôte. Des résultats encourageants avec des facteurs d'enrichissement compris entre 10 et 30 ont été obtenus sur l'expérience de sélection modèle. Cela devrait permettre d'identifier 1 peptide cyclique d'intérêt dans une collection de 1 million après 4 à 6 tours de sélection pour peu que la récupération de l'information génétique soit efficace. Or, nous avons mis en évidence que le nombre de bactéries fluorescentes récoltées après le tri est très faible, ce qui rend la manipulation de récupération de l'ADN très difficile. Des optimisations doivent être réalisées pour permettre la récupération d'un minimum de 10% des cellules contenues dans le tube après le tri. Une autre solution serait de travailler avec un FACS capable de trier les cellules à une vitesse plus élevée. Deux clones intéressants ont été identifiés après deux tours de sélection et des repiquages. Par différentes expériences, nous avons observé que l'expression de ces deux peptides a une influence négative sur la croissance des bactéries qui les expriment.

IV. Matériel et méthodes

Souches bactériennes et milieux de culture

Génotypes des souches d'E. coli

TG1	supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5(rK– mK–) [F´traD36 proAB lacIqZΔM15]
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-
JC8679	F- thr-1 leurB6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his-4 argE3 rpsL31 (Sm^R) tsx-33 supE44 kdgK51recB recC sbcA
JM109	e14–(McrA–) recA1 endA1 gyrA96 thi-1 hsdR17 (rK– mK+) supE44 relA1 Δ(lac-proAB) [F´ traD36 proAB lacIqZΔM15]

Milieux de culture

Des colonies d'E. coli sont obtenues après croissance à 37°C pendant 16h ou à 23°C pendant 48h dans un milieu contenant 20g/l de poudre LB Broth Base (Lennox L Broth Base) de chez Invitrogen. Pour travailler en milieu solide, du LB Agar (Lennox L Agar) de chez Invitrogen est utilisé à une concentration de 32 g/l. En fonction de la bactérie cultivée et de l'expérience réalisée, l'antibiotique approprié est ajouté à la concentration finale donnée dans le tableau ci-dessous.

Solution stock	Préparation	Concentration finale
Ampicilline (50 mg/ml)	Dissoudre 500 mg dans 10 ml d'H_2O Stériliser par filtration Conserver à -20°C	100 µg/ml
Chloramphénicol (50 mg/ml)	Dissoudre 500 mg dans 10 ml d'H_2O Stériliser par filtration Conserver à -20°C	15 µg/ml
Kanamycine (50 mg/ml)	Dissoudre 500 mg dans 10 ml d'H_2O Stériliser par filtration Conserver à -20°C	50 µg/ml
Tétracycline (5 mg/ml)	Dissoudre 50 mg dans 5 ml d'éthanol et 5 ml d'H_2O Conserver à -20°C et à l'abri de la lumière	7,5 µg/ml
Streptomycine (10 mg/ml)	Dissoudre 100 mg dans 10 ml d'H_2O Stériliser par filtration Conserver à -20°C	10 µg/ml
Erythromycine (250 mg/ml)	Dissoudre 500 mg dans 2 ml d'H_2O Stériliser par filtration Conserver à -20°C	250 µg/ml

Tableau 17 : Antibiotiques couramment utilisés

Pour les expériences de détection des peptides cycliques par spectrométrie de masse un milieu CAS (milieu minimum) a été utilisé.

Milieu CAS 1l

6,4 g $Na_2HPO_4.7H_20$
1,5 g KH_2PO4
0,25 g NaCl
5 g NH_4Cl
12 g Acides Casamino (DIFCO)
2 mg L-tryptophane
Porter à 1l avec l'H_2O
Ajouter du $CaCl_2$ 0,1 mM et du $MgCl_2$ 1mM

Préparation de cellules électrocompétentes

Cette opération consiste à débarrasser les cellules d'un maximum de sels afin de pouvoir les soumettre à un choc électrique en vue d'y introduire de l'ADN. En effet, une concentration en sels trop importante dans les cellules provoquerait un arc électrique au moment de l'électroporation et les endommagerait. Toutes les opérations doivent être réalisées de manière stérile en maintenant les solutions ainsi que les récipients sur glace.

- Mettre en culture durant une nuit à 37°C, 10 µl d'une solution stock de cellules à - 80°C dans 10 ml de LB.
- Le lendemain, transférer dans 2 erlenmeyers de 2 l : 500 ml de LB + 5 ml de la pré-culture de cellules.
- Incuber à 37°C et 180 rpm jusqu'à ce que les cellules soient en phase exponentielle de croissance (DO_{600} = 0,6).
- Refroidir sur glace et transférer dans 2 tubes à centrifuger stériles de 500 ml.
- Centrifuger à 4000 rpm durant 10 minutes pour récupérer les cellules (centrifugeuse Beckman J2-21, rotor JA-20).
- Eliminer le surnageant et resuspendre chaque culot dans 400 ml d'eau stérile et froide (4°C)
- Centrifuger à nouveau 10 minutes à 4000 rpm.
- Eliminer le surnageant et resuspendre chaque culot dans 200 ml d'eau stérile et froide (4°C)
- Centrifuger 10 minutes à 4000 rpm.
- Eliminer le surnageant et resuspendre chaque culot dans 20 ml de glycérol 10 % stérile et froid (4°C).
- Transférer dans un tube à centrifuger stérile de 40 ml.
- Centrifuger 10 minutes à 4000 rpm (centrifugeuse Beckman J2-21, rotor JA-17)

- Enlever le surnageant et resuspendre délicatement le culot dans 1 ml de glycérol 10 % stérile et froid (4°C)
- Transférer des aliquots de 100 µl dans des eppendorfs stériles refroidis dans un bain carboglace-éthanol.
- Placer rapidement les cellules à -80°C où elles seront conservées

La compétence des cellules est estimée par électoporation d'un plasmide de référence (pUC18, qui confère à la bactérie la résistance à l'ampicilline). Des compétences différentes sont obtenues suivant la souche utilisée et également d'une préparation à l'autre, elle est souvent supérieure à 1.10^8 transformants par µg d'ADN de pUC18.

Transformation des cellules par électroporation

Le principe de cette méthode est de soumettre des bactéries électrocompétentes à une très haute tension électrique en présence d'ADN de plasmide, de phagemide ou de phage. Ce choc électrique forme des pores dans la membrane bactérienne, permettant ainsi à l'ADN d'entrer dans la bactérie. L'appareil utilisé est le « Cell Jet Basic », commercialisé par Eurogentec et réglé sur 2500V. Il est essentiel d'effectuer la manipulation sur glace et à proximité d'un bec Bunsen pour éviter un réchauffement trop rapide des cellules compétentes et pour travailler sous champ stérile.

- Introduire dans une cellule d'électroporation, 100 µl de cellules compétentes et l'ADN du vecteur à électroporer (idéalement 100 à 200 ng)
- Appliquer ensuite la décharge et ajouter 900 µl de milieu SOC stérile
- Récupérer et transférer les cellules dans un falcon de 15 ml.
- Incuber immédiatement à 37°C et 180rpm durant 1h30.
- Etaler ensuite des dilutions de ces cultures sur des boîtes de Pétri contenant le milieu de sélection approprié selon le vecteur introduit.
- Incuber ces boîtes durant une nuit à 37°C. Chaque bactérie transformée donne une colonie sur boîte.

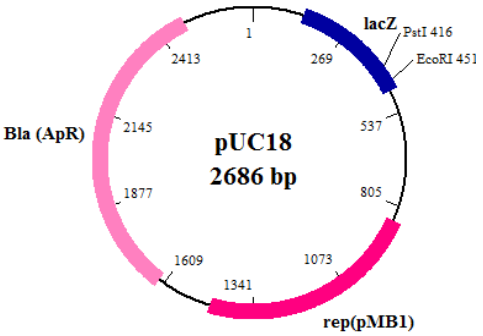
Milieu SOC 50 ml
1 g LB Broth Base (Invitrogen)
Porter à 50 ml avec l'H_2O et autoclaver
Ajouter 500 µl de $MgCl_2$ 2M et 500 µl de glucose 2M

Les vecteurs de clonage

Les plasmides : pGIV012, pUC18Ery, pAP02 et pBADTet

pGIV012 et pUC18Ery

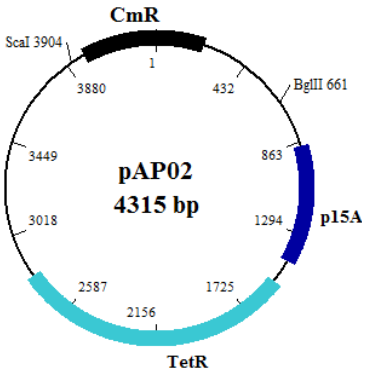
Ces plasmides (reçu du Prof. B. Hallet) ont été utilisés pour l'expérience modèle de sélection "Anti-Darwinienne" au FACS. Ces deux plasmides dérivent du plasmide pUC18 (2686 pb), qui porte une origine de réplication pMBI et le gène de résistance à l'ampicilline.



Pour obtenir le plasmide pGIV012 (3926 pb), le gène de résistance à la kanamycine a été cloné dans le site de restriction PstI. Pour le plasmide pUC18Ery (3763 pb), le gène de résistance à l'érythromycine a été inséré entre les sites de restriction EcoRI et PstI.

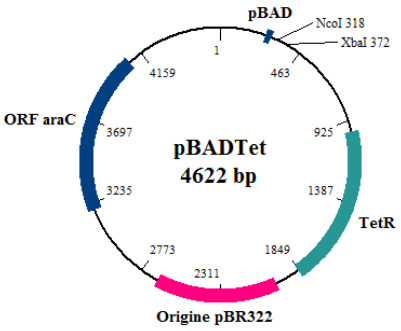
pAP02

Ce plasmide (reçu du Prof. B. Hallet) a été utilisé pour construire le vecteur p15I$_C$PX et p15CX. Il a une taille de 4315 pb et dérive du plasmide pACYC184. Ce plasmide possède des gènes de résistance à la tétracycline ainsi qu'au chloramphénicol et porte une origine de réplication p15a. Deux sites multiples de clonages ont été clonés de part et d'autre du gène de résistance au chloramphénicol. Les sites de restriction uniques ScaI et BglII ont servi au clonage de l'insert désiré.



pBADTet

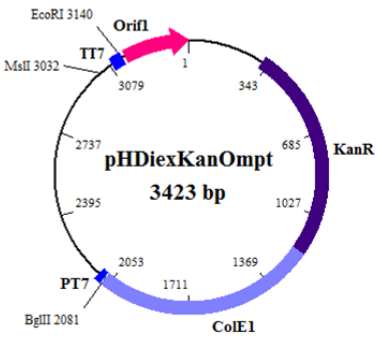
Ce plasmide a été utilisé pour produire la β-lactamase cyclique et les banques de peptides cycliques. Il a une taille de 4622 pb et dérive du vecteur d'expression pBAD/myc-HisB (Invitrogen) dans lequel le marqueur de résistance à l'ampicilline a été remplacé par un gène de résistance à la tétracycline.



Le vecteur pBAD porte le promoteur de l'opéron arabinose araBAD et le gène araC qui code le régulateur négatif et positif de ce promoteur. L'expression peut-être induite en ajoutant de l'arabinose 0,2 % (p/v) et réprimée en ajoutant du glucose 1 % (p/v). Les sites de restriction uniques NcoI et XbaI ont servi au clonage de l'insert en aval du promoteur P_{BAD}.

Le phagemide : pHDiexKanOmpt

Ce phagemide a été utilisé pour construire les vecteurs $pColYI_N$ et pColYm. Il a une taille de 3423 pb et dérive du plasmide pBluescript. Il possède un gène de résistance à la kanamycine, une origine de réplication double brin (ColE1) et une simple brin (ori f1).



Ce vecteur a été modifié pour enlever toute la partie se trouvant entre les sites de restriction BglII et MslI pour obtenir un vecteur de 2472 pb. Le site de restriction unique EcoRI a été utilisé pour cloner le gène désiré devant le terminateur T7.

Culture, purification et titrage des phagemides

Encapsidation et culture des phagemides

Les phagemides se comportent tout à fait comme des plasmides. Cependant, l'infection par des « phages auxiliaire de type fd » (comme le phage r408) d'une culture d'un transformant va fournir en *trans* la machinerie permettant la réplication comme simple brin du vecteur et son encapsidation dans des particules filamenteuses qui seront excrétées.

- Mettre en culture durant une nuit à 37°C, 10 µl d'une solution stock à -80°C de cellules contenant le phagemide dans 10 ml de LB contenant la kanamycine
- Le lendemain, transférer dans 1 erlenmeyer de 100 ml : 50 ml de LB + 5 ml de la pré-culture de cellules.
- Incuber à 37°C jusqu'à une OD de 0,6
- Incuber 20 minutes à 37°C à 80 rpm pour que les pili se développent bien.
- Ajouter les phages auxiliaire r408 à la culture tel qu'il y ait 20 phages par cellule
- Incuber 1h à 37°C et 180 rpm
- Transvaser le tout dans un erlenmeyer de 1 l contenant 200 ml de LB et la kanamycine en concentration appropriée.
- Incuber toute la nuit à 37°C

Purification des phagemides

- Centrifuger la culture de phages de 250 ml à 10000 tours/min pendant 15 minutes à 4°C (centrifugeuse Beckman J2-21, rotor JA-20)
- Récupérer le surnageant et y ajouter 50 ml de PEG-NaCl
- Laisser précipiter les phages durant 1h à 4°C
- Centrifuger à 10000 tours/min pendant 15 minutes à 4°C (centrifugeuse Beckman J2-21, rotor JA-20)
- Eliminer le surnageant et resuspendre le culot dans 10 ml de TBS
- Filtrer la solution (filtre Millex-HV 0,45 µm Filter Unit de chez Millipore, préfiltre Millex-AP20 de chez Whatman)
- Ajouter 2 ml de PEG-NaCl
- Laisser précipiter les phages sur glace pendant 20 min
- Centrifuger à 10000 tours/min pendant 15 minutes à 4°C (centrifugeuse Beckman J2-21, rotor JA-17)
- Eliminer le surnageant et resuspendre le culot dans 500 µl de TBS et ajouter NaN_3 0,02 % pour longue conservation

En suivant ce protocole, 500 µl de solution contenant à la fois le phagemide et le phage auxiliaire R408 est obtenu. Conserver ce mélange à 4°C.

PEG 20% 1l:
200 g de Polyéthylèneglycol 6000
146 g de NaCl (2,5 M)
Porter à 1 litre avec de l'eau $milli_Q$
Autoclaver 20 min

PBS 1l (10X):
80 g de NaCl (1,4 M)
2 g de KCl (0,03 M)
11,5 g de $Na_2HPO_4.7H_2O$
2 g de KH_2PO_4
Porter à 1 litre avec de l'eau $milli_Q$
Ajuster le pH à 7,4
Autoclaver 20 min et conserver à 4°C

Titrage des phagemides

Le titrage des bactériophages est une manière de mesurer leur infectivité.

- Incuber à 37°C et 180 rpm 10 µl de cellules glycérolées TG1 dans un erlenmeyer stérile contenant 25 ml de LB. Eventuellement faire une préculture sur une nuit.
- Quand les cellules sont en phase exponentielle (OD entre 0,6 et 1), mettre 90 µl de culture dans 10 eppendorfs.
- Ajouter 10 µl de phages au premier eppendorf et diluer par 10 dans chacun des 9 eppendorfs suivant.
- Incuber les bactéries infectées sans agiter à 37°C pendant 30 minutes.
- Faire des spots de 10 µl de chaque dilution sur une boîte de Pétri contenant le milieu adapté.
- Incuber une nuit à 37°C.

Le comptage du nombre de colonies nous permet de déterminer le nombre de phages infectieux présents en solution. Cette méthode de comptage des colonies est rendue possible par la présence d'un marqueur de résistance inséré dans l'ADN du phagemide, et les cellules infectées sont rendues résistantes à l'antibiotique en question.

Techniques de manipulation de l'ADN

Préparation d'ADN plasmidique d'E. coli

Les extractions d'ADN plasmidique ont été réalisées à l'aide du kit "High Pure Plasmid Isolation Kit" de chez Roche en suivant le protocole fourni par la firme.

Electrophorèse sur gel d'agarose

Cette technique permet de séparer, selon leur taille, des fragments d'ADN double brin. L'échantillon, préalablement mélangé à un tampon de chargement, est déposé sur un gel d'agarose 1 à 2% (p/v) en fonction de la taille de l'ADN attendu. Le gel est placé dans le tampon de migration et l'électrophorèse est réalisée à 70 mA et 120 V. Le gel et le tampon contiennent du bromure d'éthidium. Ce dernier est un agent intercalant qui s'insère entre les paires de bases de la double hélice de l'ADN et permet de le visualiser par fluorescence sous rayons ultra-violets. La taille des fragments d'ADN peut être estimée grâce à leur position sur le gel par rapport à un marqueur de poids moléculaire dont la taille des différents fragments est connue. La quantité d'ADN présente dans chaque bande peut être calculée en comparant son intensité par rapport à l'intensité de la bande du marqueur de poids moléculaire correspondante.

Gel d'agarose 1%

0,7 g d'agarose
Dissoudre à chaud dans 70 mL de TAE 1X.
Refroidir sous agitation
Ajouter 2 µL bromure d'éthidium à 10 mg/mL
Couler le gel dans le moule

Tampon de chargement (6X)
1300 µL d'eau milliQ
200 µL de glycérol
Quelques grains de bleu de bromophénol

Tampon de migration
30 µL bromure d'éthidium par litre de TAE 1X

TAE 1 l (50X)
242 g de Tris-base
57,1 mL d'acide acétique glacial
100 mL d'EDTA 0,5 M

Porter à 1 l avec de l'eau milliQ

Marqueur de poids moléculaire
2-Log DNA Ladder 0.1–10.0 kb (Biolabs)

Extraction d'un fragment d'ADN sur gel d'agarose

Cette étape est utilisée après une restriction double d'un vecteur de clonage afin d'éliminer le fragment de restriction qui pourrait rentrer en compétition avec l'insert lors de la ligation. L'ADN d'intérêt est récupéré en découpant la bande correspondante sur un gel d'agarose "Low Melting Point" 1% de chez Invitrogen et en le purifiant à l'aide du kit "QIAquick Gel Extraction Kit" de chez Qiagen en respectant les indications fournies par la firme.

Séquencage

Les séquençages ont été réalisés par Bénédicte Purnelle (Unité FYSA, UCL) par la méthode de Sanger.

Préparation des échantillons

- Mettre dans un Eppendorf 100 ng d'ADN double brin
- Ajouter 1 µl d'amorce (10 µM)
- Porter à 12 µl avec de l'H_2O

PCR (Polymerase Chain Reaction)

La PCR permet de synthétiser *in vitro* un très grand nombre de copies totales ou partielles d'une matrice d'ADN présente en quantité très limitée au départ. Deux amorces sont utilisées comprenant chacune une séquence complémentaire de l'une des extrémités du fragment à amplifier. La première étape de la PCR consiste à dénaturer l'ADN par la chaleur. La température est ensuite abaissée afin de permettre l'hybridation spécifique des amorces sur la matrice d'ADN. Durant la troisième étape, la température est ajustée pour une activité optimale de la polymérase choisie. Celle-ci synthétise un brin d'ADN complémentaire à la matrice à partir des oligonucléotides. Ce cycle de trois étapes est répété plusieurs fois, l'ADN synthétisé au cours de chacun des cycles servant de matrice lors du cycle suivant. L'amplification est donc de type exponentiel. Cette technique a été utilisée dans ce travail pour deux applications différentes : le criblage et le

clonage. Pour les criblages, la "Go Taq DNA polymerase" de chez Promega a été utilisée. Pour les clonages, le "FastStart High Fidelity PCR System" de chez Roche a été employé pour obtenir une amplification avec la meilleure fidélité possible. Un protocole standard est décrit ci-dessous pour les fragments inférieurs à 3kb (Tableau 18). Cependant, pour chaque PCR, la température d'hybridation doit être optimisée selon la composition et la taille du segment d'ADN à amplifier, ainsi que celle des amorces utilisées.

	Criblage	**Clonage**
Appareil	Mastercycler Gradient (Eppendorf)	Mastercycler Gradient (Eppendorf)
Enzyme	Go Taq DNA polymerase (Promega)	FastStart High Fidelity Enzyme (Roche)
Réaction	10 µl de tampon (5x) 1 µl amorce forward (20 µM) 1 µl amorce reverse (20 µM) 1 µl mélange dNTP (10 mM) 0,5 µl polymérase (5 U/µl) 100 pg-10 ng d'ADN x µl d'H_20 pour obtenir un volume final de 50 µl	5 µl de tampon (10x) 1 µl amorce forward (20 µM) 1 µl amorce reverse (20 µM) 1 µl mélange dNTP (10 mM) 0,5 µl polymérase (5 U/µl) 100 pg-10 ng d'ADN x µl d'H_20 pour obtenir un volume final de 50 µl
Cycles (25X)	Dénaturation initiale : 2 min 95°C Dénaturation : 30 sec 95°C Hybridation : 30 sec 50°C Elongation : 1 min/kb 72°C Elongation finale : 5 min 72°C	Dénaturation initiale : 2 min 95°C Dénaturation : 30 sec 95°C Hybridation : 30 sec 57°C Elongation : 1 min/kb 72°C Elongation finale : 3 min 72°C

Tableau 18 : Protocole standard pour réaliser une PCR dans le cadre du criblage ou du clonage pour les fragments inférieur à 3 kb.

Restriction

Les enzymes de restriction clivent les deux brins des duplex d'ADN au niveau de sites spécifiques. Elles permettent donc d'obtenir des fragments d'ADN délimité par des bouts droits ou collants spécifiques qui vont pouvoir être ligués dans un vecteur de clonage restreint par les mêmes enzymes. Les enzymes de restriction de chez Biolabs ont été utilisées avec leur tampon de restriction 10x. Les restrictions ont été réalisées dans les conditions optimales spécifiées par la firme concernant le temps et la température d'incubation, le tampon et la quantité d'enzyme à utiliser. Il est possible de restreindre un fragment d'ADN par deux enzymes différentes en même temps, à conditions que les deux

enzymes soient actives dans le même tampon. Certaines enzymes peuvent être inactivées par traitement à la chaleur (20 minutes à 65°C).

- Mettre dans un Eppendorf 100 ng à 1µg d'ADN
- Ajouter 1 à 2 unités d'enzyme de restriction
- Ajouter 2 µl de tampon de restriction 10x
- Porter à 20 µl avec de l'H_2O
- Incuber pendant 1 à 3h à 37°C

Déphosphorylation

Cette étape permet d'enlever les groupements 5'-phosphates de l'ADN et est utile lors des clonages pour éviter que le vecteur restreint une fois ne se referme sur lui-même. L'Antarctic Phosphatase (5 U/µl) de chez Biolabs a été utilisée avec son tampon 10x en suivant les indications de la firme.

- Mettre dans un Eppendorf 100 ng à 1µg d'ADN
- Ajouter 1µl d'Antarctic Phosphatase (5 U/µl)
- Ajouter 2 µl de tampon 10x concentré
- Porter à 20 µl avec de l'H_2O
- Incuber pendant 1h à 37°C
- Désactiver l'enzyme à 65°C 5 min

Polymérisation

Cette technique permet de liguer un vecteur sur lui-même alors qu'il a été restreint par deux enzymes qui génèrent des bouts collants non compatibles. Après polymérisation, les extrémités à bouts collants du vecteur sont transformées en bouts droits qui vont pouvoir être liguées. La T4 DNA polymerase (3 U/µl) de chez Biolbas a été utilisée avec son tampon 10x en suivant le protocole de la firme.

- Mettre dans un Eppendorf 100 ng à 1µg d'ADN
- Ajouter 1µl T4 DNA polymerase (3 U/µl)
- Ajouter 2 µl tampon 10x concentré
- Ajouter 1 µl mélange dNTP (10 mM chacun)
- Porter à 20 µl avec de l'H_2O
- Incuber pendant 1h à 37°C
- Désactiver l'enzyme à 75°C 20 min

Ligation

La ligation permet de relier deux fragments d'ADN entre eux par la formation d'un lien phosphodiester entre l'extrémité 5′ phosphorylée et l'extrémité 3′ OH. Elle est particulièrement utilisée pour effectuer l'introduction d'un insert dans un vecteur de clonage. La T4 DNA ligase de chez Biolabs a été utilisé avec son tampon 10x conformément aux instructions fournies par la firme. Les quantités d'insert et de vecteur utilisés ont été calculées avec l'équation suivante :

$$\frac{Qt\ de\ vecteur\ (ng)\ x\ Taille\ de\ l^{'}insert(kb)}{Taille\ du\ vecteur\ (ng)} \times N = Qt\ d^{'}insert(ng)$$

Dans cette équation, N est le rapport du nombre de moles de vecteur et d'insert. Le rapport généralement utilisé est 3/1.

- Mélanger dans un eppendorf le vecteur restreint avec une quantité adéquate d'insert
- Ajouter 1 µL de l'enzyme T4 DNA ligase 4U/µl (Biolabs)
- Ajouter 2 µL tampon de ligation 10x concentré
- Porter à 20 µL avec de l'H_2O
- Incuber pendant une nuit à 16°C
- Inactiver l'enzyme en incubant 10 min à 65°C

Purification d'ADN en solution

Cette étape est réalisée après une PCR, une restriction enzymatique ou une ligation pour éliminer les enzymes ainsi que les sels qui ont été utilisés. Ceci permet d'éviter qu'ils n'interfèrent lors des étapes ultérieures. Le kit de purification « High pure PCR Product Purification Kit » de chez Roche a été utilisé en suivant les instructions de la firme.

TA-cloning

Le passage par cette étape lors des clonages est d'une grande importance car il permet de faciliter la digestion des sites qui se trouvent en bordure des fragments PCR par les enzymes de restriction. Les fragments PCR ont été clonés dans le vecteur pGEM-T à l'aide du kit "pGEM-T and pGEM-T Easy Vector Systems" de chez Promega en suivant le protocole fourni par la firme. Cette technique profite du fait que lors de l'amplification des fragments d'ADN par la

PCR « *in vitro* », la polymérase ajoute une désoxyadénosine à chaque extrémité 5' des fragments et que les vecteurs pGEM-T possèdent une désoxythymidine à chacune de leur extrémité 3'. En présence de ligase, les fragments obtenus par PCR vont donc pouvoir s'introduire dans les vecteurs. Une propriété importante de ces vecteurs est qu'ils permettent d'identifier les clones recombinants par simple criblage de la couleur bleu / blanc. Cela est rendu possible par la présence dans le vecteur de la région codant pour la β-galactosidase qui est entrecoupé par le site de clonage. En conséquence, en présence du substrat chromogénique, X-Gal, l'enzyme catalyse la formation d'un produit coloré en bleu qui peut être détecté au niveau de la colonie. Donc, le vecteur original donnera lieu à une coloration bleue de la colonie et le vecteur recombinant à une coloration blanche

- Mélanger dans un eppendorf 1 µl de vecteur pGEM-T (50 ng/µl) avec une quantité adéquate de produit PCR (ratio 1 : 3)
- Ajouter 1 µL de l'enzyme T4 DNA ligase (3 U/µl)
- Ajouter 5 µL tampon de ligation 2x concentré
- Porter à 10 µL avec de l'H_2O
- Incuber pendant une nuit à 16°C
- Electroporer 2 µl du mélange dans 100 µl de cellules TOP10
- Etaler sur des boîtes de Pétri contenant de l'ampicilline (50 µg/ml), de l'IPTG 0,5 mM et de l'X-Gal 80 µg/ml.

Mutagenèse dirigée

Le kit "QuickChange Site-Directed Mutagenesis Kit" de chez Stratagene permet d'introduire des mutations ponctuelles à un endroit précis du gène. Il a été utilisé dans le cadre de la thèse en suivant les instructions fournies par la firme. Deux amorces sens et anti-sens, portant la mutation à insérer, ont été crées. Une PCR a été réalisée sur le plasmide contenant le gène à muter avec ces amorces pour synthétiser un nouveau brin d'ADN ouvert et portant la mutation désirée. L'enzyme DpnI a été ajoutée et les échantillons ont été incubés 1h à 37°C pour digérer les brins parentaux méthylés. Les nouveaux brins mutés ont été électroporés dans des cellules compétentes TOP10. L'ADN est alors réparé pas les cellules et les colonies poussant sur boîte de Pétri ont été analysées par séquençage pour isoler les clones mutants.

Analyse des protéines

Préparation d'extraits périplasmiques

Les extraits périplasmiques ont été préparés à partir des culots bactériens par la méthode du choc osmotique.

- Centrifuger 10 ml de culture bactérienne 10 min à 4000 rpm (Centrifugeuse ALC pK120, rotor O-E24)
- Resuspendre le culot bactérien dans 330 µL de sucrose 20 %, tris-HCL 20 mM pH 8
- Ajouter 6,6 µl EDTA (5M pH8)
- Incuber à température ambiante 10 min et centrifuger les cellules 10 min à 6000 rpm (Centrifugeuse Beckman, microfuge 18)
- Ecarter le surnageant et réaliser le choc osmotique en ajoutant 500 µL de $MgSO_4$ 5 mM (froid) sur le culot bactérien
- Incuber la suspension 10 min à 4°c
- Centrifuger 10 min à 10000 rpm pour obtenir l'extrait périplasmique dans le surnageant (Centrifugeuse Beckman, microfuge 18)

Préparation d'extraits cytoplasmiques

L'extrait cytoplasmique a été réalisé sur le culot de cellules obtenu après l'extrait périplasmique.

- Resuspendre le culot dans 200 µl de tampon de lyse (tris-HCl 50 mM pH 8, 1 mM EDTA et 100 mM NaCl)
- Conserver toute la nuit à -20°c
- Dégeler l'échantillon en incubant à 37°c
- Ajouter 8 µL de lysozyme (10 mg/ml) et incuber 20 min à température ambiante et ensuite 45 min à 37°c
- Ajouter 16 µL de DNase (1 mg/ml) et incuber 15 min à température ambiante
- Centrifuger 15 min à 10000 rpm pour obtenir l'extrait cytoplasmique dans le surnageant (Centrifugeuse Beckman, microfuge 18)

Le culot final peut être resuspendu dans 500 µl de PBS 1x pour donner la fraction insoluble.

PBS 1l (10X):
80 g de NaCl (1,4 M)
2 g de KCl (0,03 M)
11,5 g de $Na_2HPO_4.7H_2O$
2 g de KH_2PO_4

Porter à 1 litre avec de l'eau milli$_Q$
Ajuster le pH à 7,4
Autoclaver 20 min et conserver à 4°C

Préparation des fractions membranaires

- Centrifuger 10 ml de culture bactérienne 10 min à 4000 rpm (Centrifugeuse ALC pK120, rotor O-E24)
- Ecarter le surnageant et conserver le culot à -20°C toute la nuit
- Resuspendre le culot dans 1 ml de tampon de broyage
- Ajouter 10 µl de lysozyme (10 mg/ml) et incuber 10 min à température ambiante
- Soniquer dans des falcons de 15 ml pendant 5 min (30s-30s) dans le sonicateur bioruptor (Diagenode)
- Centrifuger 5 min à 4000 rpm (Centrifugeuse ALC pK120, rotor O-E24)
- Transférer le surnageant dans un nouveau tube (Beckman/Centrifuge tube) et centrifuger 15 min à 54000 rpm (Centrifugeuse Beckman Coulter, Optima max, rotor TLA-55)
- Ecarter le surnageant
- Resuspendre le culot dans 40 µL de tampon de suspension par sonication (Sonicateur VibraCell 75022)

Milieu de broyage 50 mL
0,1702 g Imidazole
0,0292 g EDTA
2,277 g Sorbitol
Porter à 50 ml avec H_2O
Aliquoter en 10 ml et conserver à 4°C
Avant utilisation, ajouter 0,3 g de DTT, 2,5 µl de chaque inhibiteur de protéase (leupeptine, aprotinine, antipaine, pepstatine, chymostatine) et 100 µl de PMFS par tube de 10 ml

Milieu de suspension 50 ml
0,1702 g Imidazole
0,0292 g EDTA
Porter à 50 ml avec H_2O
Aliquoter en 10 ml et conserver à 4°C
Avant utilisation, ajouter 2,5 µl de chaque inhibiteur de protéase (leupeptine, aprotinine, antipaine, pepstatine, chymostatine)

Dosage de protéines (méthode de Bradford)

La méthode de Bradford est un dosage colorimétrique, basé sur le changement d'absorbance du bleu de Commassie après liaison avec les acides aminés aromatiques (tryptophane, tyrosine et phénylalanine) et hydrophobes présent

dans les protéines. La forme libre du colorant est rouge et verte, absorbant à 465-470 nm. La forme liée du colorant est bleue, et possède un spectre d'absorption maximal à 595 nm. Le changement d'absorbance est proportionnel à la quantité de colorant lié et permet donc de déterminer la concentration en protéines dans l'échantillon.

- Transférer un volume connu de protéines dans un tube Eppendorf de 1,5 ml contenant 100 µL de NaOH 0,1 M
- Ajuster le volume avec de l'H_2O pour arriver à 200 µL
- Ajouter 800 µl de la solution de Bradford
- Transférer 300 µl du mélange dans un puits d'une microplaque 96 puits
- Lire la $D0_{595nm}$ (BIO-Rad model 550 Microplate Reader)
- Réaliser la droite de calibration en utilisant des concentrations croissantes (0, 2, 4, 6, 8 et 10 µg) en une solution standard de BSA (1µg/ml)

Solution de Bradford 1 l
0,1 g bleu de Coomassie G250 (Fluka)
47,5 ml éthanol
85 ml acide orthophosphorique
Porter à 1 l avec H_2O
Mélanger pendant 16 h a 4°C dans le noir
Filtrer sur du papier filtre Whatman 2
Conserver à 4°C dans le noir

Analyses SDS-PAGE

La technique du gel d'électrophorèse en conditions dénaturantes ("sodium dodecyl sulfate polyacrylamide gel electrophoresis" ou SDS-PAGE) décrite par Ulrich Laemmli en 1970 a été utilisée pour détecter les protéines d'intérêt dans nos échantillons. En effet, elle permet de séparer des protéines uniquement selon leur poids moléculaire. Les protéines sont tout d'abord condensées dans un gel de concentration (couche supérieure) puis séparées dans un gel de séparation (couche inférieure). La protéine d'intérêt peut être détectée sur le gel en comparant les bandes avec un marqueur de poids moléculaire dont la taille des différents fragments est connue.

Détection de la β-lactamase (30 kDa) et des fragments de l'intéine (5 et 12 kDa)

Préparation pour deux gels 10%

Solution stock	Gel de séparation	Gel de concentration
Acrylamide/bisacrylamide 30%	5 ml	2 ml
Tris-Cl/SDS, pH8,45	5 ml	3 ml
H2O	3 ml	8 ml
Glycérol	1,6 ml	/
Ammonium persulfate 10%	100 µl	100 µl
TEMED	7,5 µl	15 µl

Tableau 19 : Recette pour la préparation de deux gels d'acrylamide 10% pour la détection de protéines de faible poids moléculaire

- Mélanger la solution d'acrylamide, de Tris-Cl/SDS et d'H_20 dans un falcon de 50 ml pour le gel de séparation et de 15 ml pour le gel de concentration.
- Ajouter le glycérol pour le gel de séparation.
- Couler sans attendre le gel de séparation entre les deux plaques de verre après avoir ajouté l'ammonium persulfate et le TEMED.
- Recouvrir la surface d'H_2O_Q et laisser le gel polymériser pendant une heure.
- Enlever l'eau et couler le gel de concentration après avoir ajouté l'ammonium persulfate et le TEMED
- Placer les peignes qui formeront les puits.
- Mélanger l'échantillon (10 µl) avec le bleu de charge (10 µl) et chauffer 10 min à 95°C avant de charger sur le gel.
- Réaliser l'électrophorèse dans un cuve contenant les tampons cathode et anode à un voltage constant de 100-150 V jusqu'à ce que le front de colorant atteigne le bas du gel.
- Pour colorer le gel, ajouter la solution de coloration et passer 10 sec au micro-onde.
- Pour décolorer le gel, ajouter la solution de décoloration et passer à nouveau 10 sec au micro-onde.

Tris-Cl/SDS 500 mL
182 g de Tris base
1,5 g de SDS
Ajuster le pH à 8,45 avec HCl
Porter à 500 ml avec H_2O et conserver à 4°C.

Tampon Cathode 1l (1x)
12,11 g Tris-base
17,92 g Tricine
1 g SDS

Porter à 1l avec H_2O
Conserver à 4°C

Tampon Anode 1l (5x)
121,1 g Tris-base
500 mL H2O
Ajouter le pH à 8,9 avec de l'HCl
Porter à 1 l avec de l'H_2O
Conserver à 4°C

Bleu de charge 10 ml (2x)
1 ml 1M Tris-Cl pH6,8
0,4 g SDS
2 ml glycerol
0,02 g bleu de bromophénol
Porter à 10 ml avec H_20
Aliquoter en 1 ml et conserver à température ambiante
Avant utilisation, ajouter 31 mg de DTT par tube de 1 ml et 1 µl de chaque inhibiteur de protéase (leupeptine, aprotinine, antipaine, pepstatine, chymostatine)

Marqueur de poids moléculaire
PageRuler™ Prestained Protein Ladder, Fermentas

Solution de coloration 500 ml
1g bleu de Coomassie
180 ml méthanol
40 ml acide acétique
180 ml d'H_2O

Solution de décoloration 1l
200 ml méthanol
25 ml acide acétique
775 ml H_2O

Western Blot

Les protéines ainsi séparées par électrophorèse peuvent être transférées sur une membrane PVDF pour ensuite être exposées à un anticorps spécifique.

- Couper un morceau de la membrane PVDF (Immobilon-P Transfert Membrane, 0,45 µm, Millipore) et deux papiers filtres (Extra Thick Filter Paper, Biorad) de la même taille que le gel
- Placer les deux papiers filtres dans un bain contenant du tampon de transfert
- Récupérer le gel, enlever le gel de concentration et le plonger dans le tampon de transfert 4 min

- Inbiber la membrane PVDF dans du méthanol 1 seconde puis laisser 5 min dans le tampon de transfert
- Placer la membrane et le gel entre les deux papiers filtres et mettre le tout dans le Trans-Blot Semi-Dry Transfert Cell (Biorad)
- Lancer le transfert pendant 60 min à 22 volts
- Saturer la membrane pendant 16h à 4°C sous agitation dans le tampon de saturation
- Après un rinçage rapide, laver la membrane pendant 10 min dans 30 ml de tampon de lavage à température ambiante
- Incuber la membrane en présence de l'anticorps primaire dilué dans le tampon de lavage pendant 1h
- Après un rinçage rapide, réaliser 3 lavages de 10 min à température ambiante
- Incuber la membrane en présence de l'anticorps secondaire dilué dans le tampon de lavage pendant 1h
- Après un rinçage rapide, réaliser 3 lavages de 10 min à température ambiante
- Pour la révélation chimiluminescente, mélanger 5 ml de réactif A (luminal et 4-iodophenol) et 50 µl de réactif B (peroxide) (Kit Roche)
- Mettre la membrane en contact de cette solution pendant 1 min
- Exposer la membrane sur film Biomax-MF (Kodak) et développer avec le processeur Kodak M35 X-OMAT

Tampon de transfert semi-dry 2,5 L
14,55 g Tris-Cl
7,325 g Glycine
9,375 g SDS
250 ml Méthanol
Porter à 2,5 L avec H_2O

TBS 1l (10X)
24,2 g Tris-Cl
80 g NaCl
Ajouster à pH7,6 avec HCl
Porter à 1l avec H_2O

Tampon de saturation 200 ml
20 ml TBS 10x
1 ml Tween 20
6g lait en poudre
Porter à 200 ml avec H_2O

Tampon de lavage 2 L
200 ml TBS 10x
2 ml Tween 20
10 g lait en poudre
Porter à 2 L avec H_2O

Analyses par spectrométrie de masse (MALDI-TOF-TOF)

Des analyses par spectrométrie de masse (MADLI-TOF-TOF) ont été réalisées pour identifier la β-lactamase et les peptides dans le périplasme et déterminer s'ils sont sous la forme cyclique ou linéaire. Les extraits périplasmiques, préparés à partir de cellules induites, ont été concentrés 50 fois et traités de deux manières différentes avant d'être analysés sur le spectromètre Applied Biosystems MALDI-TOF-TOF AB4800. Sans purification, 0,5 µl d'une solution contenant 10 mg/ml d'acide alpha-cyano-4-hydroxycinnamic (matrice MALDI alpha-cyano) dans 50% (v/v) d'acétonitrile et 0,1% (v/v) de TFA a été ajouté à 0,5 µl d'échantillon. 0,5 µl de ce mélange a été déposé sur la cible. Avec purification, la solution contenant les peptides a été acidifiée avec 1 µl de TFA 10 % (v/v) et passée dans un ZipTip MiliporeC18 (Bedford MA, USA). 0,5 µl d'une solution contenant 10 mg/ml d'acide alpha-cyano-4-hydroxycinnamic (matrice MALDI alpha-cyano) dans 50% (v/v) d'acétonitrile et 0,1% (v/v) de TFA a été ajouté à 0,5 µl d'échantillon purifié. 0,5 µl de ce mélange a été déposé sur la cible. L'ionisation est réalisée à l'aide un laser en phase solide à 200 Hz qui opère à une longueur d'onde de λ = 355 nm. Les spectres MS ont été obtenus dans une gamme de m/z entre 500 et 4000. Les précurseurs automatiquement sélectionnés ont été fragmentés à une énergie de collision de 1kV avec ou sans une collision gaz air à une pression d'environ $1x10^{6}$ Torr. Les données ont été collectées avec le logiciel "Applied Biosystems 4000 Series Explorer™" et les spectres MS/MS ont été analysés avec le logiciel "Applied Biosystems GPS Explorer TM – De Novo Explorer Version 3.6".

Analyse des bactéries marquées à l'IP au microscope à épifluoresence.

Le pourcentage de bactéries mortes a été estimé au moyen du kit LIVE/DEAD bactlight™ Bacterial Viability L-7012 de Molecular Probes (Molecular Probes, Eugene, Oregon, USA) et d'un microscope à épifluorescence (Polyvar Reichert-Jung).

Le kit contient deux fluorochromes qui marquent les acides nucléiques, le SYTO-9 de fluorescence verte et l'iodure de propidium de fluorescence rouge. Ils différent par leurs caractéristiques spectrales et par leur capacité à pénétrer dans les cellules bactériennes. Les maxima d'excitation/émission sont de 480nm/500nm pour le SYTO-9 et de 490nm/635nm pour l'iodure de propidium. Le SYTO-9 marque généralement toute la population bactérienne, aussi bien les cellules dont la membrane est intacte que celles dont la membrane est altérée. Par contre, l'iodure de propidium pénètre seulement dans les bactéries dont la membrane est endommagée, c'est à dire les mortes. Lorsque les deux colorants sont présents dans une même cellule, la fluorescence du SYTO-9 est réduite car elle est captée par l'Iodure de Propidium. Ce phénomène est connu sous le nom de FRET (*Fluorescence Resonnance Energy Transfert*). En utilisant un mélange de ces deux colorants en proportion adéquate, les bactéries dont les membranes sont intactes (les vivantes) sont colorées en vert alors que les bactéries dont les membranes sont endommagées (les mortes) sont rouges.

Les concentrations finales utilisées pour les fluorochromes sont les suivantes : SYTO-9 5 µM (solution stock à 3,34 mM dans du DMSO) ; IP 30 µM (solution stock à 20 mM dans du DMSO) dans 50 µL de suspension cellulaire. Afin d'obtenir ces concentrations, les solutions stock des colorants sont diluées 13 fois et un mélange 1 :1 est réalisé le jour même. Ces solutions sont conservées sur glace et à l'abri de la lumière jusqu'à leur utilisation.

Les suspensions cellulaires marquées sont incubées 15 min à l'abri de la lumière. Les lames d'observation microscopique sont ensuite préparées (Menzel-Glazer). On dépose successivement sur la lame : 1µL de suspension cellulaire marquée, 2 µL d'une solution de Mowiol, une lamelle. Les lames sont placées à l'abri de la lumière le plus rapidement possible afin d'éviter la dégradation des fluorochromes.

Les lames sont observées au moyen d'un microscope à épifluorescence (Polyvar Reichert-jung) muni des filtres B2 (FITC) (901124, filtre d'excitation en LR bleu BP450-495, filtre d'arrêt sélectif BP520-560) et G2 (901220, filtre d'excitation en LR vert BP520-560, filtre d'arrêt sélectif LP590). Ces deux filtres ne permettent pas de visualiser les deux colorants à la fois mais la saisie de photos avec chacun des deux filtres permet d'obtenir des clichés des bactéries marquées avec chacun des deux colorants. Grâce à l'immobilisation des bactéries, la comparaison des clichés permet de déterminer la proportion de bactéries mortes et vivantes dans l'échantillon. Trois photos sont prises avec chaque filtre. Les photos ont été prises à l'aide d'un appareil digital Fujifilm digital camera FinePix S1 Pro et du programme informatique Camera shooting sftware for FinePix S1 Pro. Le nombre de cellules marquées en vert ou en rouge est comptabilisé manuellement pour un nombre total minimum de 300 cellules. Ensuite, le pourcentage de cellules mortes par rapport au nombre total de cellules est calculé.

Le Mowiol (Mowiol 4.88, Calbiochem catalog number 475904) est un milieu de montage adapté à l'utilisation de fluorochromes. Il permet l'immobilisation des bactéries. Il est préparé comme suit :

- Mettre 6g de glycérol dans un tube de centrifugation (30 ml) et ajouter un petit barreau magnétique.
- Ajouter 2,4 g de Mowiol et agiter pour mélanger.
- Ajouter 6 ml d'eau distillée et stérile et laisser agiter 2h à température ambiante
- Ajouter 12 ml de 0,2 M Tris (pH 8,5).
- Incuber le tube dans de l'eau chaude (50-60°c) pendant 10 minutes pour dissoudre le Mowiol. Cette étape peut être répétée plusieurs fois si nécessaire.
- Une fois que le Mowiol est dissous, centrifuger à 5000 x g pendant 15 minutes pour enlever les solides non dissous.
- Aliquoter dans des eppendorfs de 1 mL et congeler à long terme à -20°c ou réfrigérer à 4°c pour conserver maximum 2 semaines.

Cytométrie en flux : FACS

La cytométrie en flux est une technologie qui permet la mesure simultanée de plusieurs caractéristiques physiques d'une cellule (sa taille, sa complexité interne et son intensité de fluorescence) et d'analyser une population à la vitesse de plusieurs milliers d'évènements par seconde (Figure 74). Elle est basée sur l'injection des cellules à analyser dans un flux liquide qui va les forcer à adopter une progression en file indienne afin d'être irradiées une par une avec un faisceau lumineux. Chaque cellule va émettre des signaux optiques qui vont être captés par une optique appropriée, transmis à des détecteurs photoélectriques puis à un ordinateur qui mémorise toutes les données individuelles et dans l'ordre de passage de chaque cellule. Finalement, l'ordinateur va calculer les données statistiques associées aux distributions des paramètres mesurés et les représenter sous la forme d'histogramme (1 paramètre) ou de dot plot (2 paramètres). Les cytomètres les plus évolués possèdent une fonction de tri qui permet de trier les cellules en fonction de leurs propriétés mesurées.

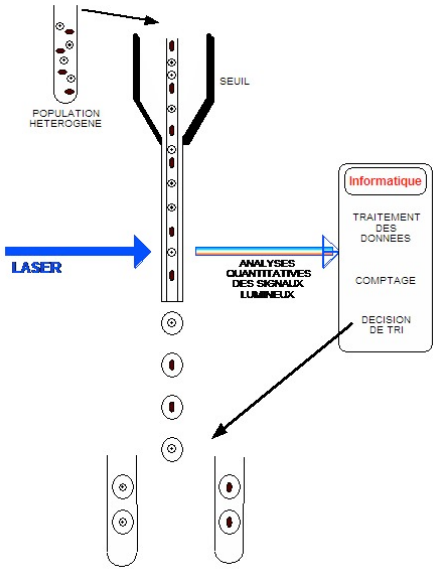


Figure 74 : Schéma simplifié d'un cytomètre en flux.

Les signaux sont émis par les cellules dans toutes les directions de l'espace et sont essentiellement le résultat de diffusion lumineuse et de fluorescence obtenues par des marquages spécifiques. Les signaux de diffusion sont captés

dans deux directions privilégiées, dans l'axe de la propagation laser et à 90° de cet axe. La fluorescence, 1000 fois plus faible qu'un signal de diffusion, ne peut être mesurée qu'à 90° de l'axe sous peine d'être "noyée" par le faisceau lumineux. Si la diffusion de la lumière est mesurée quasi dans l'axe du rayon incident, l'intensité du signal peut être corrélée avec la taille cellulaire, on parle de "Forward Scatter". Sous un angle de 90°, la mesure correspond à la structure intracellulaire de la cellule (réfringence du cytoplasme, morphologie, rapport nucléo-cytoplamsique), on parle de "Side Scatter".

Le tri cellulaire est une des applications de la cytométrie en flux. Dans ce cas, après l'analyse et dans un délai parfaitement connu par la machine, le flux liquide est fragmenté de façon à emprisonner chaque cellule dans une goutte. Les gouttes sont chargées négativement ou positivement en fonction des critères désignés par le chercheur et sont ensuite déviées de leur chute initiale par un champ électrique constant pour arriver dans des flacons récolteurs.

Le cytomètre utilisé dans nos expériences est un FACS Vantage, capable d'analyser et trier 20000 événements/s. Les cellules ont été triées en fonction de leur fluorescence dans le rouge en utilisant l'IP. C'est un agent d'intercalation qui se lie aux bases de l'ADN avec peu ou pas de spécificité de séquence et lorsqu'il est liée, émet une fluorescence 20 à 30 fois plus élevée qu'une molécule libre en solution. Son maximum d'excitation lorsqu'il est lié à l'ADN est de 535 nm et son maximum d'émission est 617 nm (Figure 75). L'IP peut donc être excité par le laser du FACS Vantage qui émet à 488 nm. Grâce à sa relative hydrophobie, l'IP a été utilisé comme marqueur de viabilité cellulaire. En effet, les cellules vivantes possèdent une intégrité membranaire qui empêche l'IP de pénétrer, tandis que les cellules mortes ont des membranes perméabilisées qui le laissent entrer.

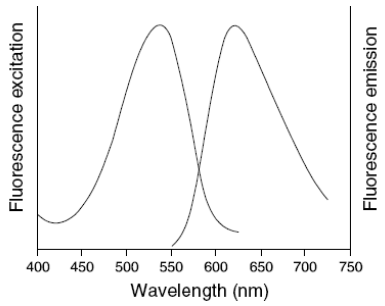


Figure 75 : Profil d'excitation et d'émission de l'iodure de propidium lié à l'ADN

- Centrifuger 10 min à 4000 rpm 25 ml de culture de cellules ($0D_{600}$= 0,6) à analyser au FACS Vantage (centrifugeuse Beckman J2-21, rotor JA-17)
- Resuspendre dans 10 ml de PBS pour réhydrater les cellules
- Attendre 15 min à RT
- Centrifuger 10 min à 4000 rpm (centrifugeuse Beckman J2-21, rotor JA-17)
- Resuspendre dans 10 ml de Tris Staining Buffer
- Ajuster la densité optique de l'échantillon à OD_{600} = 1
- Ajouter 10 µl d'IP (1 mg/ml) dans 500 µl d'échantillon
- Faire passer l'échantillon dans le FACS Vantage

Tris Staining Buffer 500 ml (1X)

6g Tris-base, pH7,4 (100 mM)
4,4 g NaCl (150 mM)
0,55 g mM $CaCL_2$(1 mM)
0,24 g $MgCl_2$ (0,5 mM)

V. Références bibliographiques

1. Klekota, B. & Miller, B.L. Dynamic diversity and small-molecule evolution: a new paradigm for ligand identification. *Trends Biotechnol* **17**, 205-209(1999).
2. McGregor, D.P. Discovering and improving novel peptide therapeutics. *Current Opinion in Pharmacology* **8**, 616-619(2008).
3. Loffet, A. Peptides as drugs: is there a market? *J. Pept. Sci* **8**, 1-7(2002).
4. Hummel, G., Reineke, U. & Reimer, U. Translating peptides into small molecules. *Mol Biosyst* **2**, 499-508(2006).
5. Hruby, V.J. Conformational restrictions of biologically active peptides via amino acid side chain groups. *Life Sci* **31**, 189-199(1982).
6. Hruby, V.J., al-Obeidi, F. & Kazmierski, W. Emerging approaches in the molecular design of receptor-selective peptide ligands: conformational, topographical and dynamic considerations. *Biochem. J* **268**, 249-262(1990).
7. Kessler, H. Conformation and Biological Activity of Cyclic Peptides. *Angewandte Chemie International Edition in English* **21**, 512-523(1982).
8. Rizo, J. & Gierasch, L.M. Constrained peptides: models of bioactive peptides and protein substructures. *Annu. Rev. Biochem* **61**, 387-418(1992).
9. Shao, Y., Lu, W. & Kent, S.B.H. A novel method to synthesize cyclic peptides. *Tetrahedron Letters* **39**, 3911-3914(1998).
10. Zhang, L. & Tam, J.P. Synthesis and Application of Unprotected Cyclic Peptides as Building Blocks for Peptide Dendrimers. *Journal of the American Chemical Society* **119**, 2363-2370(1997).
11. Camarero, J.A. & Muir, T.W. Chemoselective backbone cyclization of unprotected peptides. *Chem. Commun.* 1369-1370(1997).
12. Camarero, J.A., Pavel, J. & Muir, T.W. Chemical Synthesis of a Circular Protein Domain: Evidence for Folding-Assisted Cyclization. *Angewandte Chemie International Edition* **37**, 347-349(1998).
13. Sancheti, H. & Camarero, J.A. "Splicing up" drug discovery.: Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Advanced Drug Delivery Reviews* **61**, 908-917(2009).
14. Kohli, R.M., Walsh, C.T. & Burkart, M.D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658-661(2002).
15. Walsh, C.T. Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **303**, 1805-1810(2004).
16. Camarero, J.A. et al. Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. *Bioorganic & Medicinal Chemistry* **9**, 2479-2484(2001).
17. Scott, C.P. et al. Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A.* **96**, 13638–13643(1999).
18. Evans, T.C., Benner, J. & Xu, M. The Cyclization and Polymerization of Bacterially Expressed Proteins Using Modified Self-splicing Inteins. *J. Biol. Chem.* **274**, 18359-18363(1999).
19. Iwai, H., Lingel, A. & Pluckthun, A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus. *J. Biol. Chem* **276**, 16548-16554(2001).
20. Saïda, F. et al. Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Curr Protein Pept Sci* **7**, 47-56(2006).

21. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **72**, 211-222(2006).
22. Saïda, F. Overview on the expression of toxic gene products in Escherichia coli. *Curr Protoc Protein Sci* **Chapter 5**, Unit 5.19(2007).
23. Anthony, L.C., Suzuki, H. & Filutowicz, M. Tightly regulated vectors for the cloning and expression of toxic genes. *Journal of Microbiological Methods* **58**, 243-250(2004).
24. Müller, J., Oehler, S. & Müller-Hill, B. Repression oflacPromoter as a Function of Distance, Phase and Quality of an AuxiliarylacOperator. *Journal of Molecular Biology* **257**, 21-29(1996).
25. Glascock, C.B. & J. Weickert, M. Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in Escherichia coli. *Gene* **223**, 221-231(1998).
26. Krebber, A., Burmester, J. & Plückthun, A. Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p. *Gene* **178**, 71-74(1996).
27. Brown, W.C. & Campbell, J.L. A new cloning vector and expression strategy for genes encoding proteins toxic to Escherichia coli. *Gene* **127**, 99-103(1993).
28. Studier, F.W. & Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-30(1986).
29. Grossman, T.H. et al. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* **209**, 95-103(1998).
30. Pan, S.H. & Malcolm, B.A. Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *Biotechniques* **29**, 1234-8(2000).
31. Moffatt, B.A. & Studier, F.W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221-227(1987).
32. Dubendorff, J.W. & Studier, F.W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219**, 45-59(1991).
33. Calos, M.P. DNA sequence for a low-level promoter of the lac repressor gene and an 'up' promoter mutation. *Nature* **274**, 762-5(1978).
34. Doherty, A.J., Connolly, B.A. & Worrall, A.F. Overproduction of the toxic protein, bovine pancreatic DNaseI, in Escherichia coli using a tightly controlled T7-promoter-based vector. *Gene* **136**, 337-40(1993).
35. Bouet, J.Y. et al. The effects on Escherichia coli of expression of the cloned bacteriophage T4 nucleoid disruption (ndd) gene. *Mol Microbiol* **20**, 519-28(1996).
36. Bouet, J., Krisch, H.M. & Louarn, J. Ndd, the Bacteriophage T4 Protein That Disrupts the Escherichia coli Nucleoid, Has a DNA Binding Activity. *J Bacteriol.* **180**, 5227–5230(1998).

37. Carra, J.H. & Schleif, R.F. Variation of half-site organization and DNA looping by AraC protein. *EMBO J* **12**, 35-44(1993).
38. Ogden, S. et al. The Escherichia coli L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proc Natl Acad Sci U S A* **77**, 3346-50(1980).
39. Lobell, R.B. & Schleif, R.F. DNA looping and unlooping by AraC protein. *Science* **250**, 528-32(1990).
40. Miyada, C.G., Stoltzfus, L. & Wilcox, G. Regulation of the araC gene of Escherichia coli: catabolite repression, autoregulation, and effect on araBAD expression. *Proc Nat Acad Sci U S A* **81**, 4120-4124(1984).
41. Guzman, L.M. et al. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-30(1995).
42. Roux, A., Beloin, C. & Ghigo, J. Combined inactivation and expression strategy to study gene function under physiological conditions: application to identification of new Escherichia coli adhesins. *J Bacteriol* **187**, 1001-13(2005).
43. Boomershine, W.P. et al. Preparation of uniformly labeled NMR samples in Escherichia coli under the tight control of the araBAD promoter: expression of an archaeal homolog of the RNase P Rpp29 protein. *Protein Expression and Purification* **28**, 246-251(2003).
44. Hall, T.A. & Brown, J.W. Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins. *RNA* **8**, 296-306(2002).
45. Wycuff, D.R. & Matthews, K.S. Generation of an AraC-araBAD Promoter-Regulated T7 Expression System. *Analytical Biochemistry* **277**, 67-73(2000).
46. Novick, A. & Weiner, M. ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proc. Natl. Acad. Sci. U.S.A* **43**, 553-566(1957).
47. Carrier, T.A. & Keasling, J.D. Investigating autocatalytic gene expression systems through mechanistic modeling. *J. Theor. Biol* **201**, 25-36(1999).
48. Siegele, D.A. & Hu, J.C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. U.S.A* **94**, 8168-8172(1997).
49. Khlebnikov, A. et al. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol* **182**, 7029-7034(2000).
50. Goswitz, V.C. & Brooker, R.J. Isolation of lactose permease mutants which recognize arabinose. *Membr Biochem* **10**, 61-70(1993).
51. Morgan-Kiss, R.M., Wadler, C. & Cronan, J.E. Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. U.S.A* **99**, 7373-7377(2002).
52. Mizuno, T., Chou, M.Y. & Inouye, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A.* **81**, 1966–1970(1984).

53. Pestka, S. et al. Anti-mRNA: specific inhibition of translation of single mRNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 7525-7528(1984).
54. O'Connor, C.D. & Timmis, K.N. Highly repressible expression system for cloning genes that specify potentially toxic proteins. *J. Bacteriol.* **169**, 4457-4462(1987).
55. Worrall, A. & Connolly, B. The chemical synthesis of a gene coding for bovine pancreatic DNase I and its cloning and expression in Escherichia coli. *J. Biol. Chem.* **265**, 21889-21895(1990).
56. Ward, D.F. & Murray, N.E. Convergent transcription in bacteriophage lambda: interference with gene expression. *J Mol Biol* **133**, 249-66(1979).
57. Podhajska, A.J., Hasan, N. & Szybalski, W. Control of cloned gene expression by promoter inversion in vivo: construction of the heat-pulse-activated att-nutLp-att-N. *Gene* **43**, 325(1986).
58. Hasan, N. & Szybalski, W. Control of cloned gene expression by promoter inversion in vivo: construction of improved vectors with a multiple cloning site and the ptac promoter. *Gene* **56**, 145-151(1987).
59. Wülfing, C. & Plückthun, A. A versatile and highly repressible Escherichia coli expression system based on invertible promoters: expression of a gene encoding a toxic product. *Gene* **136**, 199-203(1993).
60. Rose, R.E. The nucleotide sequence of pACYC177. *Nucleic Acids Res* **16**, 356(1988).
61. Komai, T. et al. Development of HIV-1 protease expression methods using the T7 phage promoter system. *Applied Microbiology and Biotechnology* **47**, 241-245(1997).
62. Lanzer, M. & Bujard, H. Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* **85**, 8973-7(1988).
63. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucl. Acids Res.* **25**, 1203-1210(1997).
64. Hillen, W. & Berens, C. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol* **48**, 345-69(1994).
65. Sadowski, P. Site-specific genetic recombination: hops, flips, and flops. *FASEB J.* **7**, 760-767(1993).
66. Morgan, T.H. An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in Drosophila. *Journal of Experimental Zoology* **11**, 365-413(1911).
67. McCLINTOCK, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol* **16**, 13-47(1951).
68. Campbell, A. Episomes in evolution. *Brookhaven Symp Biol* **23**, 534-62(1972).
69. Kuzminov, A. Recombinational Repair of DNA Damage in Escherichia coli and Bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* **63**, 751-813(1999).
70. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88(2002).

71. Churchill, J.J., Anderson, D.G. & Kowalczykowski, S.C. The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of χ, resulting in constitutive recombination activation. *Genes Dev.* **13**, 901–911(1999).
72. Anderson, D.G. & Kowalczykowski, S.C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77-86(1997).
73. Umezu, K. & Kolodner, R.D. Protein interactions in genetic recombination in Escherichia coli. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. *J Biol Chem* **269**, 30005-13(1994).
74. Webb, B.L., Cox, M.M. & Inman, R.B. Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. *Cell* **91**, 347-56(1997).
75. Shan, Q. et al. RecA protein filaments: end-dependent dissociation from ssDNA and stabilization by RecO and RecR proteins. *J Mol Biol* **265**, 519-40(1997).
76. Wang, T.V. & Smith, K.C. recF-dependent and recF recB-independent DNA gap-filling repair processes transfer dimer-containing parental strands to daughter strands in Escherichia coli K-12 uvrB. *J Bacteriol.* **158**, 727–729(1984).
77. West, S.C. Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* **31**, 213-44(1997).
78. Cox, M.M. & Lehman, I.R. recA protein of Escherichia coli promotes branch migration, a kinetically distinct phase of DNA strand exchange. *Proc Nat Acad Sci U S A* **78**, 3433-3437(1981).
79. DasGupta, C. & Radding, C.M. Polar branch migration promoted by recA protein: effect of mismatched base pairs. *Proc. Natl. Acad. Sci. U.S.A* **79**, 762-766(1982).
80. Shen, P. & Huang, H.V. Homologous Recombination in ESCHERICHIA COLI: Dependence on Substrate Length and Homology. *Genetics.* **112**, 441–457(1986).
81. Ayares, D. et al. Sequence homology requirements for intermolecular recombination in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A* **83**, 5199-5203(1986).
82. Lovett, S.T. et al. Crossing over between regions of limited homology in Escherichia coli. RecA-dependent and RecA-independent pathways. *Genetics.* **160**, 851–859(2002).
83. Artz, S. et al. Use of M13mp phages to study gene regulation, structure and function: cloning and recombinational analysis of genes of the Salmonella typhimurium histidine operon. *Gene* **26**, 147-58(1983).
84. Blum, P. et al. Gene replacement and retrieval with recombinant M13mp bacteriophages. *J Bacteriol.* **171**, 538–546(1989).
85. Zeef, L.A.H. & Bosch, L. A technique for targeted mutagenesis of the EF-Tu chromosomal gene by M13-mediated gene replacement. *Molecular and General Genetics MGG* **238**, 252-260(1993).

86. Schaaper, R.M., Danforth, B.N. & Glickman, B.W. Rapid repeated cloning of mutant lac repressor genes. *Gene* **39**, 181-9(1985).
87. Arthur, A. & Sherratt, D. Dissection of the transposition process: a transposon-encoded site-specific recombination system. *Mol Gen Genet* **175**, 267-74(1979).
88. Reed, R.R. & Grindley, N.D. Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* **25**, 721-8(1981).
89. KAMP, D. et al. Inversion of the G DNA segment of phage Mu controls phage infectivity. *Nature* **271**, 577-580(1978).
90. van de Putte, P., Cramer, S. & Giphart-Gassler, M. Invertible DNA determines host specificity of bacteriophage Mu. *Nature* **286**, 218-222(1980).
91. Austin, S., Ziese, M. & Sternberg, N. A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell* **25**, 729-36(1981).
92. Hallet, B. & Sherratt, D.J. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev* **21**, 157-78(1997).
93. Grindley, N.D.F., Whiteson, K.L. & Rice, P.A. Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605(2006).
94. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem* **58**, 913-949(1989).
95. Hallet, B., Arciszewska, L.K. & Sherratt, D.J. Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination. *Mol Cell* **4**, 949-59(1999).
96. Biswas, T. et al. A structural basis for allosteric control of DNA recombination by [lambda] integrase. *Nature* **435**, 1059-1066(2005).
97. Nunes-Düby, S.E., Matsumoto, L. & Landy, A. Site-specific recombination intermediates trapped with suicide substrates. *Cell* **50**, 779-88(1987).
98. Kitts, P.A. & Nash, H.A. Homology-dependent interactions in phage lambda site-specific recombination. *Nature* **329**, 346-8
99. Heffron, F.*Mobile Genetic Elements (ed. Shapiro, J.)* 223-260(1983).
100. Argos, P. et al. The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**, 433–440(1986).
101. Mahillon, J. & Lereclus, D. Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate-resolution process. *EMBO J* **7**, 1515-26(1988).
102. Vanhooff, V. et al. Self-control in DNA site-specific recombination mediated by the tyrosine recombinase TnpI. *Mol Microbiol* **60**, 617-29(2006).
103. Stewart, A. F., Zhang, Y. & Hallet, B. A new tyrosine recombinase for genetic engineering. *US Patent Application #09/895.435.* (2001).
104. Clardy, J., Fischbach, M.A. & Walsh, C.T. New antibiotics from bacterial natural products. *Nat Biotech* **24**, 1541-1550(2006).
105. Newman, D.J. & Cragg, G.M. Natural Products as Sources of New Drugs over the Last 25 Years⊥. *Journal of Natural Products* **70**, 461-477(2007).
106. Von Döhren, H. Antibiotics: Actions, origins, resistance, by C. Walsh. 2003. Washington, DC: ASM Press. 345 pp. $99.95 (hardcover). *Protein Sci* **13**, 3059-3060(2004).

107. Nolan, E.M. & Walsh, C.T. How nature morphs peptide scaffolds into antibiotics. *Chembiochem* **10**, 34-53(2009).
108. Lipmann, F. Attempts to map a process evolution of peptide biosynthesis. *Science* **173**, 875-884(1971).
109. Hopwood, D.A. Genetic Contributions to Understanding Polyketide Synthases. *Chemical Reviews* **97**, 2465-2498(1997).
110. Schwarzer, D. & Marahiel, M.A. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* **88**, 93-101(2001).
111. Mootz, H.D., Schwarzer, D. & Marahiel, M.A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3**, 490-504(2002).
112. Walsh, C.T. et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* **5**, 525-534(2001).
113. Hubbard, B.K. & Walsh, C.T. Vancomycin Assembly: Nature's Way. *Angewandte Chemie International Edition* **42**, 730-765(2003).
114. Miao, V. et al. Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507-1523(2005).
115. Hunter, P.A. Coping with the rising tide of resistance to antimicrobial agents. *Drug News Perspect* **14**, 309-317(2001).
116. Keating, T.A. & Walsh, C.T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology* **3**, 598-606(1999).
117. Bernhardt, P. & O'Connor, S.E. Opportunities for enzyme engineering in natural product biosynthesis. *Current Opinion in Chemical Biology* **13**, 35-42(2009).
118. Stachelhaus, T., Schneider, A. & Marahiel, M.A. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69-72(1995).
119. Schwarzer, D., Finking, R. & Marahiel, M.A. Nonribosomal peptides: from genes to products. *Nat Prod Rep* **20**, 275-287(2003).
120. Finking, R. & Marahiel, M.A. Biosynthesis of nonribosomal peptides1. *Annu. Rev. Microbiol* **58**, 453-488(2004).
121. Sieber, S.A. & Marahiel, M.A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev* **105**, 715-738(2005).
122. Nguyen, K.T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A.* **103**, 17462–17467(2006).
123. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A* **97**, 6640-6645(2000).
124. Nguyen, K.T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A.* **103**, 17462–17467(2006).
125. Donia, M.S., Ravel, J. & Schmidt, E.W. A global assembly line for cyanobactins. *Nat. Chem. Biol* **4**, 341-343(2008).

126. Ziemert, N. et al. Ribosomal Synthesis of Tricyclic Depsipeptides in Bloom-Forming Cyanobacteria13. *Angewandte Chemie* **120**, 7870-7873(2008).
127. Willey, J.M. & van der Donk, W.A. Lantibiotics: peptides of diverse structure and function. *Annu. Rev. Microbiol* **61**, 477-501(2007).
128. Levengood, M.R. et al. In Vitro Mutasynthesis of Lantibiotic Analogues Containing Nonproteinogenic Amino Acids. *Journal of the American Chemical Society* **131**, 12024-12025(2009).
129. Acker, M.G., Bowers, A.A. & Walsh, C.T. Generation of thiocillin variants by prepeptide gene replacement and in vivo processing by Bacillus cereus. *J. Am. Chem. Soc* **131**, 17563-17565(2009).
130. Moore, B.S. Extending the biosynthetic repertoire in ribosomal peptide assembly. *Angew. Chem. Int. Ed. Engl* **47**, 9386-9388(2008).
131. Donia, M.S. et al. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol* **2**, 729-735(2006).
132. Chatterjee, C. et al. Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. *J. Am. Chem. Soc* **127**, 15332-15333(2005).
133. Scott, C.P. et al. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem Biol* **8**, 801-15(2001).
134. Perler, F.B. et al. Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res* **22**, 1125-1127(1994).
135. Perler, F.B. Protein splicing mechanisms and applications. *IUBMB Life* **58**, 63(2006).
136. Xu, M.Q. & Perler, F.B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**, 5146–5153(1996).
137. Perler, F.B. InBase: the Intein Database. *Nucleic Acids Res* **30**, 383-384(2002).
138. Pietrokovski, S. Intein spread and extinction in evolution. *Trends Genet* **17**, 465-472(2001).
139. Kane, P. et al. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**, 651-657(1990).
140. Cooper, A.A. et al. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* **12**, 2575-2583(1993).
141. Liu, X.Q. Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu. Rev. Genet* **34**, 61-76(2000).
142. Dalgaard, J.Z. et al. Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. *J. Comput. Biol* **4**, 193-214(1997).
143. Gogarten, J.P. et al. Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**, 263-87(2002).
144. Yang, J., Meng, Q. & Liu, X. Intein harbouring large tandem repeats in replicative DNA helicase of Trichodesmium erythraeum. *Mol. Microbiol* **51**, 1185-1192(2004).
145. Southworth, M.W. et al. Control of protein splicing by intein fragment reassembly. *EMBO J.* **17**, 918–926(1998).
146. Belfort, M. & Roberts, R.J. Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**, 3379-3388(1997).

147. Gimble, F.S. & Thorner, J. Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae. *Nature* **357**, 301-306(1992).
148. Derbyshire, V. & Belfort, M. Lightning Strikes Twice: Intron-Intein Coincidence. *Proc Nat Acad Sci U S A* **95**, 1356-1357(1998).
149. Gimble, F.S. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett* **185**, 99-107(2000).
150. Chong, S. & Xu, M. Protein Splicing of the Saccharomyces cerevisiae VMA Intein without the Endonuclease Motifs. *J. Biol. Chem.* **272**, 15587-15590(1997).
151. Derbyshire, V. et al. Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Proc Nat Acad Sci U S A* **94**, 11466-11471(1997).
152. Telenti, A. et al. The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol.* **179**, 6378–6382(1997).
153. Evans, T.C. et al. Protein trans-Splicing and Cyclization by a Naturally Split Intein from the dnaE Gene of Synechocystis Species PCC6803. *J. Biol. Chem.* **275**, 9091-9094(2000).
154. Wu, H., Xu, M. & Liu, X. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1387**, 422-432(1998).
155. Shingledecker, K., Jiang, S. & Paulus, H. Molecular dissection of the Mycobacterium tuberculosis RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments. *Gene* **207**, 187-195(1998).
156. Sun, W., Yang, J. & Liu, X. Synthetic Two-piece and Three-piece Split Inteins for Protein trans-Splicing. *J. Biol. Chem.* **279**, 35281-35286(2004).
157. Ding, Y. et al. Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem* **278**, 39133-39142(2003).
158. Perler, F.B. et al. Intervening sequences in an Archaea DNA polymerase gene. *Proc Nat Acad Sci U S A* **89**, 5577-5581(1992).
159. Perler, F.B., Xu, M.Q. & Paulus, H. Protein splicing and autoproteolysis mechanisms. *Curr Opin Chem Biol* **1**, 292-299(1997).
160. Paulus, H. The chemical basis of protein splicing. *Chem. Soc. Rev.* **27**, 375-386(1998).
161. Hirata, R. & Anraku, Y. Mutations at the putative junction sites of the yeast VMA1 protein, the catalytic subunit of the vacuolar membrane H(+)-ATPase, inhibit its processing by protein splicing. *Biochem. Biophys. Res. Commun* **188**, 40-47(1992).
162. Shao, Y., Xu, M.Q. & Paulus, H. Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry* **34**, 10844-10850(1995).
163. Xu, M.Q. et al. In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell* **75**, 1371-1377(1993).

164. Kawasaki, M. et al. Protein splicing in the yeast Vma1 protozyme: evidence for an intramolecular reaction. *FEBS Lett* **412**, 518-520(1997).
165. Romanelli, A. et al. Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein–intein junction. *Proc Nat Acad Sci USA* **101**, 6397-6402(2004).
166. Saleh, L. & Perler, F.B. Protein splicing in cis and in trans. *Chem Rec* **6**, 183-193(2006).
167. Chong, S. et al. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271-281(1997).
168. Chong, S. et al. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucl. Acids Res.* **26**, 5109-5115(1998).
169. Yu, H. Extending the size limit of protein nuclear magnetic resonance. *Proc Nat Acad Sci USA* **96**, 332-334(1999).
170. Dawson, P.E. et al. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779(1994).
171. Tam, J.P. et al. Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc. Natl. Acad. Sci. U.S.A* **92**, 12485-12489(1995).
172. Muir, T.W., Sondhi, D. & Cole, P.A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A* **95**, 6705-6710(1998).
173. Hofmann, R.M. & Muir, T.W. Recent advances in the application of expressed protein ligation to protein engineering. *Curr. Opin. Biotechnol* **13**, 297-303(2002).
174. Severinov, K. & Muir, T.W. Expressed Protein Ligation, a Novel Method for Studying Protein-Protein Interactions in Transcription. *J. Biol. Chem.* **273**, 16205-16209(1998).
175. Ayers, B. et al. Introduction of unnatural amino acids into proteins using expressed protein ligation. *Peptide Science* **51**, 343-354(1999).
176. Evans, T.C., Benner, J. & Xu, M.Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256-2264(1998).
177. Xu, R. et al. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. U.S.A* **96**, 388-393(1999).
178. Cotton, G.J. et al. Insertion of a Synthetic Peptide into a Recombinant Protein Framework: A Protein Biosensor. *Journal of the American Chemical Society* **121**, 1100-1101(1999).
179. Cotton, G.J. & Muir, T.W. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol* **7**, 253-261(2000).
180. Evans, T.C., Benner, J. & Xu, M.Q. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem* **274**, 18359-18363(1999).

181. Camarero, J.A. & Muir, T.W. Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity. *Journal of the American Chemical Society* **121**, 5597-5598(1999).
182. Baker, R.T. Protein expression using ubiquitin fusion and cleavage. *Curr. Opin. Biotechnol* **7**, 541-546(1996).
183. Mills, K.V. et al. Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc. Natl. Acad. Sci. U.S.A* **95**, 3543-3548(1998).
184. Wu, H., Xu, M. & Liu, X. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1387**, 422-432(1998).
185. Southworth, M.W. et al. Control of protein splicing by intein fragment reassembly. *EMBO J* **17**, 918-926(1998).
186. Lew, B.M., Mills, K.V. & Paulus, H. Protein splicing in vitro with a semisynthetic two-component minimal intein. *J. Biol. Chem* **273**, 15887-15890(1998).
187. Wu, H., Hu, Z. & Liu, X.Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A* **95**, 9226-9231(1998).
188. Martin, D.D., Xu, M.Q. & Evans, T.C. Characterization of a naturally occurring trans-splicing intein from Synechocystis sp. PCC6803. *Biochemistry* **40**, 1393-1402(2001).
189. Perler, F.B. Protein splicing mechanisms and applications. *IUBMB Life* **58**, 63(2006).
190. Mootz, H.D. et al. Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc* **125**, 10561-10569(2003).
191. Kwon, Y., Matthew A. Coleman & Julio A. Camarero Selective Immobilization of Proteins onto Solid Supports through Split-Intein-Mediated Protein Trans-Splicing13. *Angewandte Chemie* **118**, 1758-1761(2006).
192. Naumann, T.A., Savinov, S.N. & Benkovic, S.J. Engineering an affinity tag for genetically encoded cyclic peptides. *Biotechnol. Bioeng* **92**, 820-830(2005).
193. Williams, N.K. et al. In Vivo Protein Cyclization Promoted by a Circularly Permuted Synechocystis sp. PCC6803 DnaB Mini-intein. *J. Biol. Chem.* **277**, 7790-7798(2002).
194. Iwai, H., Lingel, A. & Plueckthun, A. Cyclic green fluorescent protein produced in vivo using artificially split PI-PfuI intein from Pyrococcus furiosus. *J. Biol. Chem.*
195. Tavassoli, A. & Benkovic, S.J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat. Protocols* **2**, 1126-1133(2007).
196. Troitskaya, L.A. & Kodadek, T. Peptides as modulators of enzymes and regulatory proteins. *Methods* **32**, 406-415(2004).
197. Trabi, M. & Craik, D.J. Circular proteins — no end in sight. *Trends in Biochemical Sciences* **27**, 132-138(2002).

198. Veber, D.F. et al. A potent cyclic hexapeptide analogue of somatostatin. *Nature* **292**, 55-58(1981).
199. Satoh, T. et al. Synthetic peptides derived from the fourth domain of CD4 antagonize off function and inhibit T cell activation. *Biochem. Biophys. Res. Commun* **224**, 438-443(1996).
200. Horton, D.A., Bourne, G.T. & Smythe, M.L. Exploring privileged structures: the combinatorial synthesis of cyclic peptides. *Mol. Divers* **5**, 289-304(2002).
201. Katz, B.A. et al. Topochemical catalysis achieved by structure-based ligand design. *J. Biol. Chem* **270**, 31210-31218(1995).
202. Horswill, A.R., Savinov, S.N. & Benkovic, S.J. A systematic method for identifying small-molecule modulators of protein–protein interactions. *Proc. Natl. Acad. Sci. U.S.A* **101**, 15591-15596(2004).
203. Horswill, A.R. & Benkovic, S.J. Identifying small-molecule modulators of protein-protein interactions. *Curr Protoc Protein Sci* **Chapter 19**, Unit 19.15(2006).
204. Tavassoli, A. & Benkovic, S.J. Genetically selected cyclic-peptide inhibitors of AICAR transformylase homodimerization. *Angew Chem Int Ed Engl* **44**, 2760-3(2005).
205. Tavassoli, A. et al. Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. *ACS Chem Biol* **3**, 757-64(2008).
206. Cheng, L. et al. Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease. *Protein Sci.* **16**, 1535–1542(2007).
207. Naumann, T.A., Tavassoli, A. & Benkovic, S.J. Genetic selection of cyclic peptide Dam methyltransferase inhibitors. *Chembiochem* **9**, 194-7(2008).
208. Kinsella, T.M. et al. Retrovirally Delivered Random Cyclic Peptide Libraries Yield Inhibitors of Interleukin-4 Signaling in Human B Cells. *J. Biol. Chem.* **277**, 37512-37518(2002).
209. Mori, H. & Ito, K. The Sec protein-translocation pathway. *Trends Microbiol* **9**, 494-500(2001).
210. Manting, E.H. & Driessen, A.J. Escherichia coli translocase: the unravelling of a molecular machine. *Mol. Microbiol* **37**, 226-238(2000).
211. Wickner, W., Driessen, A.J. & Hartl, F.U. The enzymology of protein translocation across the Escherichia coli plasma membrane. *Annu. Rev. Biochem* **60**, 101-124(1991).
212. Driessen, A.J., Manting, E.H. & van der Does, C. The structural basis of protein targeting and translocation in bacteria. *Nat. Struct. Biol* **8**, 492-498(2001).
213. Osborne, A.R., Rapoport, T.A. & van den Berg, B. Protein translocation by the Sec61/SecY channel. *Annu. Rev. Cell Dev. Biol* **21**, 529-550(2005).
214. Berks, B.C., Sargent, F. & Palmer, T. The Tat protein export pathway. *Mol. Microbiol* **35**, 260-274(2000).
215. Sheldon E. Broedel & Sharon M. Papciak ACES TM Signal Sequence and YebF Expression Systems. *Athena Enzyme Systems TM* (2007).

216. de Keyzer, J., van der Does, C. & Driessen, A.J.M. The bacterial translocase: a dynamic protein channel complex. *Cell. Mol. Life Sci* **60**, 2034-2052(2003).
217. Veenendaal, A.K.J., van der Does, C. & Driessen, A.J.M. The protein-conducting channel SecYEG. *Biochim. Biophys. Acta* **1694**, 81-95(2004).
218. Hartmann, E. et al. Evolutionary conservation of components of the protein translocation complex. *Nature* **367**, 654-657(1994).
219. Schatz, G. & Dobberstein, B. Common principles of protein translocation across membranes. *Science* **271**, 1519-1526(1996).
220. Pohlschröder, M. et al. Protein translocation in the three domains of life: variations on a theme. *Cell* **91**, 563-566(1997).
221. Akimaru, J. et al. Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A* **88**, 6545-6549(1991).
222. Brundage, L. et al. The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**, 649-657(1990).
223. Nishiyama, K., Mizushima, S. & Tokuda, H. A novel membrane protein involved in protein translocation across the cytoplasmic membrane of Escherichia coli. *EMBO J* **12**, 3409-3415(1993).
224. Duong, F. & Wickner, W. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *EMBO J* **16**, 2756-2768(1997).
225. Economou, A. et al. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* **83**, 1171-1181(1995).
226. Duong, F. & Wickner, W. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J* **16**, 4871-4879(1997).
227. Kato, Y., Nishiyama, K. & Tokuda, H. Depletion of SecDF-YajC causes a decrease in the level of SecG: implication for their functional interaction. *FEBS Lett* **550**, 114-118(2003).
228. Pogliano, J.A. & Beckwith, J. SecD and SecF facilitate protein export in Escherichia coli. *EMBO J* **13**, 554-561(1994).
229. Nouwen, N. & Driessen, A.J.M. Inactivation of protein translocation by cold-sensitive mutations in the yajC-secDF operon. *J. Bacteriol* **187**, 6852-6855(2005).
230. van der Laan, M. et al. Reconstitution of Sec-dependent membrane protein insertion: nascent FtsQ interacts with YidC in a SecYEG-dependent manner. *EMBO Rep* **2**, 519-523(2001).
231. Scotti, P.A. et al. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* **19**, 542-549(2000).
232. Samuelson, J.C. et al. YidC mediates membrane protein insertion in bacteria. *Nature* **406**, 637-641(2000).
233. Papanikou, E., Karamanou, S. & Economou, A. Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol* **5**, 839-851(2007).

234. Mitra, K., Frank, J. & Driessen, A. Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nat. Struct. Mol. Biol* **13**, 957-964(2006).
235. Koch, H.G. et al. In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein-targeting pathways of Escherichia coli. *Mol. Biol. Cell* **10**, 2163-2173(1999).
236. Ulbrandt, N.D., Newitt, J.A. & Bernstein, H.D. The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* **88**, 187-196(1997).
237. Phillips, G.J. & Silhavy, T.J. The E. coli ffh gene is necessary for viability and efficient protein export. *Nature* **359**, 744-746(1992).
238. Luirink, J. et al. Signal-sequence recognition by an Escherichia coli ribonucleoprotein complex. *Nature* **359**, 741-743(1992).
239. Valent, Q.A. et al. Nascent membrane and presecretory proteins synthesized in Escherichia coli associate with signal recognition particle and trigger factor. *Mol. Microbiol* **25**, 53-64(1997).
240. Walter, P. & Blobel, G. Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol* **91**, 557-561(1981).
241. Pugsley, A.P. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev* **57**, 50-108(1993).
242. Doudna, J.A. & Batey, R.T. Structural insights into the signal recognition particle. *Annu. Rev. Biochem* **73**, 539-557(2004).
243. Shan, S. & Walter, P. Co-translational protein targeting by the signal recognition particle. *FEBS Lett* **579**, 921-926(2005).
244. Siu, F.Y., Spanggord, R.J. & Doudna, J.A. SRP RNA provides the physiologically essential GTPase activation function in cotranslational protein targeting. *RNA* **13**, 240-250(2007).
245. Powers, T. & Walter, P. Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases. *Science* **269**, 1422-1424(1995).
246. Connolly, T. & Gilmore, R. Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. *J. Cell Biol* **103**, 2253-2261(1986).
247. Ng, D.T., Brown, J.D. & Walter, P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol* **134**, 269-278(1996).
248. Huber, D. et al. Use of Thioredoxin as a Reporter To Identify a Subset of Escherichia coli Signal Sequences That Promote Signal Recognition Particle-Dependent Translocation. *J. Bacteriol.* **187**, 2983-2991(2005).
249. Huber, D. et al. A selection for mutants that interfere with folding of Escherichia coli thioredoxin-1 in vivo. *Proc. Natl. Acad. Sci. U.S.A* **102**, 18872-18877(2005).
250. Fekkes, P. & Driessen, A.J. Protein targeting to the bacterial cytoplasmic membrane. *Microbiol. Mol. Biol. Rev* **63**, 161-173(1999).

251. Kumamoto, C.A. & Francetić, O. Highly selective binding of nascent polypeptides by an Escherichia coli chaperone protein in vivo. *J. Bacteriol* **175**, 2184-2188(1993).
252. Kumamoto, C.A. Escherichia coli SecB protein associates with exported protein precursors in vivo. *Proc. Natl. Acad. Sci. U.S.A* **86**, 5320-5324(1989).
253. Hartl, F.U. et al. The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. *Cell* **63**, 269-279(1990).
254. Fekkes, P. et al. Preprotein transfer to the Escherichia coli translocase requires the co-operative binding of SecB and the signal sequence to SecA. *Mol. Microbiol* **29**, 1179-1190(1998).
255. Fekkes, P., van der Does, C. & Driessen, A.J. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *EMBO J* **16**, 6105-6113(1997).
256. Schiebel, E. et al. Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64**, 927-939(1991).
257. van der Wolk, J.P., de Wit, J.G. & Driessen, A.J. The catalytic cycle of the escherichia coli SecA ATPase comprises two distinct preprotein translocation events. *EMBO J* **16**, 7297-7304(1997).
258. Economou, A. & Wickner, W. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**, 835-843(1994).
259. van Dalen, A., Killian, A. & de Kruijff, B. Delta psi stimulates membrane translocation of the C-terminal part of a signal sequence. *J. Biol. Chem* **274**, 19913-19918(1999).
260. Nouwen, N., Tommassen, J. & de Kruijff, B. Requirement for conformational flexibility in the signal sequence of precursor protein. *J. Biol. Chem* **269**, 16029-16033(1994).
261. Rapoport, T.A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663-669(2007).
262. Prickril, B.C. et al. Putative signal peptide on the small subunit of the periplasmic hydrogenase from Desulfovibrio vulgaris. *J. Bacteriol* **167**, 722-725(1986).
263. Robinson, C. & Bolhuis, A. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim. Biophys. Acta* **1694**, 135-147(2004).
264. Santini, C.L. et al. A novel sec-independent periplasmic protein translocation pathway in Escherichia coli. *EMBO J* **17**, 101-112(1998).
265. Settles, A.M. & Martienssen, R. Old and new pathways of protein export in chloroplasts and bacteria. *Trends Cell Biol* **8**, 494-501(1998).
266. Berks, B.C. A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol* **22**, 393-404(1996).
267. Bogsch, E.G. et al. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J. Biol. Chem* **273**, 18003-18006(1998).
268. Sargent, F. et al. Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J* **17**, 3640-3650(1998).

269. Sargent, F. et al. Sec-independent protein translocation in Escherichia coli. A distinct and pivotal role for the TatB protein. *J. Biol. Chem* **274**, 36073-36082(1999).
270. Weiner, J.H. et al. A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell* **93**, 93-101(1998).
271. Zhang, J.W. et al. A role for SlyD in the Escherichia coli hydrogenase biosynthetic pathway. *J. Biol. Chem* **280**, 4360-4366(2005).
272. Jong, W.S.P. et al. Trigger factor interacts with the signal peptide of nascent Tat substrates but does not play a critical role in Tat-mediated export. *Eur. J. Biochem* **271**, 4779-4787(2004).
273. Graubner, W., Schierhorn, A. & Brüser, T. DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone. *J. Biol. Chem* **282**, 7116-7124(2007).
274. Alami, M. et al. Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Mol. Cell* **12**, 937-946(2003).
275. Mori, H. & Cline, K. A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid [Delta]pH/Tat translocase. *J. Cell Biol* **157**, 205-210(2002).
276. Yahr, T.L. & Wickner, W.T. Functional reconstitution of bacterial Tat translocation in vitro. *EMBO J* **20**, 2472-2479(2001).
277. Natale, P., Brüser, T. & Driessen, A.J.M. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. *Biochim. Biophys. Acta* **1778**, 1735-1756(2008).
278. Lee, H.C. & Bernstein, H.D. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc. Natl. Acad. Sci. U.S.A* **98**, 3471-3476(2001).
279. Peterson, J.H., Woolhead, C.A. & Bernstein, H.D. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J. Biol. Chem* **278**, 46155-46162(2003).
280. Cunningham, K. & Wickner, W. Specific recognition of the leader region of precursor proteins is required for the activation of translocation ATPase of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A* **86**, 8630-8634(1989).
281. Perlman, D. & Halvorson, H.O. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol* **167**, 391-409(1983).
282. von Heijne, G. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem* **133**, 17-21(1983).
283. von Heijne, G. Signal sequences. The limits of variation. *J. Mol. Biol* **184**, 99-105(1985).
284. Paetzel, M. & Dalbey, R.E. Catalytic hydroxyl/amine dyads within serine proteases. *Trends in Biochemical Sciences* **22**, 28-31(1997).
285. Dalbey, R.E. et al. The chemistry and enzymology of the type I signal peptidases. *Protein Science* **6**, 1129-1138(1997).
286. Brédo, A. et al. Gene reconstitution using high efficiency homologous recombination between a bacteriophage fd and a plasmid. *Journal of Biotechnology* **126**, 286-290(2006).

287. Auvray, F. et al. Unusual Structure of the attB Site of the Site-Specific Recombination System of Lactobacillus delbrueckii Bacteriophage mv4. *J. Bacteriol.* **181**, 7385-7389(1999).
288. Falciani, C. et al. Bioactive Peptides from Libraries. *Chemistry & Biology* **12**, 417-426(2005).
289. Boldt, G.E. et al. Emerging chemical and biological approaches for the preparation of discovery libraries. *Drug Discovery Today* **11**, 143-148(2006).
290. Mersich, C. & Jungbauer, A. Generation of bioactive peptides by biological libraries. *Journal of Chromatography B* **861**, 160-170(2008).
291. Boer, A.L.D. & Schmidt-Dannert, C. Recent efforts in engineering microbial cells to produce new chemical compounds. *Current Opinion in Chemical Biology* **7**, 273-278(2003).
292. Cane, D.E., Walsh, C.T. & Khosla, C. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **282**, 63-68(1998).
293. Du, L. & Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr Opin Drug Discov Devel* **4**, 215-28(2001).
294. Iwai, H. & Plückthun, A. Circular [beta]-lactamase: stability enhancement by cyclizing the backbone. *FEBS Letters* **459**, 166-172(1999).
295. Mathys, S. et al. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**, 1-13(1999).
296. Ludwig, C., Schwarzer, D. & Mootz, H.D. Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. *J. Biol. Chem* **283**, 25264-25272(2008).
297. Patrick, W.M., Firth, A.E. & Blackburn, J.M. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng* **16**, 451-457(2003).
298. Hughes, M.D. et al. Removing the redundancy from randomised gene libraries. *J. Mol. Biol* **331**, 973-979(2003).
299. Patrick, W.M. & Firth, A.E. Strategies and computational tools for improving randomized protein libraries. *Biomol. Eng* **22**, 105-112(2005).
300. Koshland, D. & Botstein, D. Evidence for posttranslational translocation of beta-lactamase across the bacterial inner membrane. *Cell* **30**, 893-902(1982).
301. Schierle, C.F. et al. The DsbA Signal Sequence Directs Efficient, Cotranslational Export of Passenger Proteins to the Escherichia coli Periplasm via the Signal Recognition Particle Pathway. *J Bacteriol.* **185**, 5706–5713(2003).
302. Powers, T. & Walter, P. Co-translational protein targeting catalyzed by the Escherichia coli signal recognition particle and its receptor. *EMBO J* **16**, 4880-4886(1997).
303. Luirink, J. & Sinning, I. SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1694**, 17-35(2004).

304. Eser, M. et al. Disulfide bond formation by exported glutaredoxin indicates glutathione's presence in the E. coli periplasm. *Proc. Natl. Acad. Sci. U.S.A* **106**, 1572-1577(2009).
305. Depuydt, M. et al. A periplasmic reducing system protects single cysteine residues from oxidation. *Science* **326**, 1109-1111(2009).
306. Shelton, M.D., Chock, P.B. & Mieyal, J.J. Glutaredoxin: role in reversible protein s-glutathionylation and regulation of redox signal transduction and protein translocation. *Antioxid. Redox Signal* **7**, 348-366(2005).
307. Lee, J., Soonsanga, S. & Helmann, J.D. A complex thiolate switch regulates the Bacillus subtilis organic peroxide sensor OhrR. *Proc. Natl. Acad. Sci. U.S.A* **104**, 8743-8748(2007).
308. Gao, X. et al. Methods for analysis of protein glutathionylation and their application to photosynthetic organisms. *Mol Plant* **2**, 218-235(2009).
309. Sun, X.Z. et al. Formation of disulfide bond in p53 correlates with inhibition of DNA binding and tetramerization. *Antioxid. Redox Signal* **5**, 655-665(2003).
310. Khlebnikov, A. et al. Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241-3247(2001).
311. Morgan-Kiss, R.M., Wadler, C. & Cronan, J.E. Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7373-7377(2002).
312. Lee, S.K., Newman, J.D. & Keasling, J.D. Catabolite repression of the propionate catabolic genes in Escherichia coli and Salmonella enterica: evidence for involvement of the cyclic AMP receptor protein. *J. Bacteriol* **187**, 2793-2800(2005).
313. Lee, S.K. & Keasling, J.D. A propionate-inducible expression system for enteric bacteria. *Appl. Environ. Microbiol* **71**, 6856-6862(2005).
314. Lee, S.K. & Keasling, J.D. Propionate-regulated high-yield protein production in Escherichia coli. *Biotechnol. Bioeng* **93**, 912-918(2006).
315. Lee, S.K. & Keasling, J.D. A Salmonella-based, propionate-inducible, expression system for Salmonella enterica. *Gene* **377**, 6-11(2006).
316. Lee, S.K. & Keasling, J.D. Effect of glucose or glycerol as the sole carbon source on gene expression from the Salmonella prpBCDE promoter in Escherichia coli. *Biotechnol. Prog* **22**, 1547-1551(2006).
317. Tao, L. et al. Isolation of chromosomal mutations that affect carotenoid production in Escherichia coli: mutations alter copy number of ColE1-type plasmids. *FEMS Microbiol. Lett* **243**, 227-233(2005).
318. March, J.B. et al. Cloning and characterization of an Escherichia coli gene, pcnB, affecting plasmid copy number. *Mol. Microbiol* **3**, 903-910(1989).
319. Ederth, J., Isaksson, L.A. & Abdulkarim, F. Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the Escherichia coli RNA polymerase. *Mol. Genet. Genomics* **267**, 587-592(2002).

320. Nakae, T. & Nikaido, H. Outer membrane as a diffusion barrier in Salmonella typhimurium. Penetration of oligo- and polysaccharides into isolated outer membrane vesicles and cells with degraded peptidoglycan layer. *J. Biol. Chem.* **250**, 7359-7365(1975).
321. Decad, G.M. & Nikaido, H. Outer membrane of gram-negative bacteria. XII. Molecular-sieving function of cell wall. *J Bacteriol* **128**, 325-36(1976).
322. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781(2000).
323. Shi, L. et al. Limits of propidium iodide as a cell viability indicator for environmental bacteria. *Cytometry Part A* **71A**, 592-598(2007).
324. Aeschbacher, M., Reinhardt, C.A. & Zbinden, G. A rapid cell membrane permeability test using fluorescent dyes and flow cytometry. *Cell Biol. Toxicol* **2**, 247-255(1986).
325. Lehtinen, J., Nuutila, J. & Lilius, E. Green fluorescent protein-propidium iodide (GFP-PI) based assay for flow cytometric measurement of bacterial viability. *Cytometry A* **60**, 165-172(2004).
326. Ni, Y. & Chen, R. Extracellular recombinant protein production from Escherichia coli. *Biotechnol. Lett* **31**, 1661-1670(2009).
327. Russel, M. Mutants at conserved positions in gene IV, a gene required for assembly and secretion of filamentous phages. *Mol. Microbiol* **14**, 357-369(1994).
328. Marciano, D.K., Russel, M. & Simon, S.M. An aqueous channel for filamentous phage export. *Science* **284**, 1516-1519(1999).

VI. Annexe

Séquence en acide aminé et en acide nucléique de la construction permettant de cycliser et exporter la β-lactamase via le peptide signal de Bla.

```
1     ATG GGT ATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG   45
1     Met Gly Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala   15

46    GCA TTT TGC CTT CCT GTT TTT GCC ATG GGA TCA CCA GAA ATA GAA   90
16    Ala Phe Cys Leu Pro Val Phe Ala Met Gly Ser Pro Glu Ile Glu   30

91    AAG TTG TCT CAG AGT GAT ATT TAC TGG GAC TCC ATC GTT TCT ATT   135
31    Lys Leu Ser Gln Ser Asp Ile Tyr Trp Asp Ser Ile Val Ser Ile   45

136   ACG GAG ACT GGA GTC GAA GAG GTT TTT GAT TTG ACT GTG CCA GGA   180
46    Thr Glu Thr Gly Val Glu Glu Val Phe Asp Leu Thr Val Pro Gly   60

181   CCA CAT AAC TTT GTC GCC AAT GAC ATC ATT GTC CAT AAC AGT GCT   225
61    Pro His Asn Phe Val Ala Asn Asp Ile Ile Val His Asn Ser Ala   75

226   AGC GGT GGT GGT CAT CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT   270
76    Ser Gly Gly Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala   90

271   GAA GAT CAG TTG GGT GCT CGA GTG GGT TAC ATC GAA CTG GAT CTC   315
91    Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu   105

316   AAC AGC GGT AAG ATC CTT GAG AGT TTT CGC CCC GAA GAA CGT TTT   360
106   Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe   120

361   CCA ATG ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC GCG GTA TTA   405
121   Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu   135

406   TCC CGT GTT GAC GCC GGG CAA GAG CAA CTC GGT CGC CGC ATA CAC   450
136   Ser Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His   150

451   TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA GAA AAG   495
151   Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys   165

496   CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC   540
166   His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala   180

541   ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG   585
181   Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr   195

586   ATC GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG   630
196   Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly   210

631   GAT CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA   675
211   Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu   225

676   GCC ATA CCA AAC GAC GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG   720
226   Ala Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro Ala Ala Met   240

721   GCA ACA ACG TTG CGC AAA CTA TTA ACT GGC GAA CTA CTT ACT CTA   765
241   Ala Thr Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu Thr Leu   255
```

```
766   GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT   810
256   Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys Val   270

811   GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC TGG TTT ATT   855
271   Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp Phe Ile   285

856   GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT ATC ATT   900
286   Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile   300

901   GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC   945
301   Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile   315

946   TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG   990
316   Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln   330

991   ATC GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG GGA CCC GGG   1035
331   Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp Gly Pro Gly   345

1036  GAA TTC GGA TGC ATC AGT GGA GAT AGT TTG ATC AGC TTG GCG AGC   1080
346   Glu Phe Gly Cys Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser   360

1081  ACA GGA AAA AGA GTT TCT ATT AAA GAT TTG TTA GAT GAA AAA GAT   1125
361   Thr Gly Lys Arg Val Ser Ile Lys Asp Leu Leu Asp Glu Lys Asp   375

1126  TTT GAA ATA TGG GCA ATT AAT GAA CAG ACG ATG AAG CTA GAA TCA   1170
376   Phe Glu Ile Trp Ala Ile Asn Glu Gln Thr Met Lys Leu Glu Ser   390

1171  GCT AAA GTT AGT CGT GTA TTT GCT ACT GGC AAA AAG CTA GTT TAT   1215
391   Ala Lys Val Ser Arg Val Phe Ala Thr Gly Lys Lys Leu Val Tyr   405

1216  ATT TTA AAA ACT CGA CTA GGT AGA ACT ATC AAG GCA ACA GCA AAT   1260
406   Ile Leu Lys Thr Arg Leu Gly Arg Thr Ile Lys Ala Thr Ala Asn   420

1261  CAT AGA TTT TTA ACT ATT GAT GGT TGG AAA AGA TTA GAT GAG CTA   1305
421   His Arg Phe Leu Thr Ile Asp Gly Trp Lys Arg Leu Asp Glu Leu   435

1306  TCT TTA AAA GAG CAT ATT GCT CTA CCC CGT AAA CTA GAA AGC TCC   1350
436   Ser Leu Lys Glu His Ile Ala Leu Pro Arg Lys Leu Glu Ser Ser   450

1351  TCT TTA CAA TTA TAG
451   Ser Leu Gln Leu End
```

En bleu : séquence du peptide signal de Bla, en rouge : séquence de la β-lactamase TEM-1 avec les acides aminés ajoutés aux extrémités N et C-terminales pour la cyclisation, en noir : séquence des fragments $DnaBI_C$ et $DnaBI_N$ de l'intéine, en bleu clair : mutation de la cystéine en alanine dans la séquence de l'intéine $DnaBI_N$

Séquence en acide aminé et en acide nucléique de la construction permettant de cycliser et exporter le peptide SERELG via le peptide signal de DsbA.

```
1     ATG AAA AAG ATT TGG CTG GCG CTG GCT GGT TTA GTT TTA GCG TTT   45
1     Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe   15

46    AGC GCA TCG GCG GCC ATG GGA TCA CCA GAA ATA GAA AAG TTG TCT   90
16    Ser Ala Ser Ala Ala Met Gly Ser Pro Glu Ile Glu Lys Leu Ser   30

91    CAG AGT GAT ATT TAC TGG GAC TCC ATC GTT TCT ATT ACG GAG ACT   135
31    Gln Ser Asp Ile Tyr Trp Asp Ser Ile Val Ser Ile Thr Glu Thr   45

136   GGA GTC GAA GAG GTT TTT GAT TTG ACT GTG CCA GGA CCA CAT AAC   180
46    Gly Val Glu Glu Val Phe Asp Leu Thr Val Pro Gly Pro His Asn   60

181   TTT GTC GCC AAT GAC ATC ATT GTC CAT AAC AGT GAG CGT GAG TTG   225
61    Phe Val Ala Asn Asp Ile Ile Val His Asn Ser Glu Arg Glu Leu   75

226   GGA TGC ATC AGT GGA GAT AGT TTG ATC AGC TTG GCG AGC ACA GGA   270
76    Gly Cys Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser Thr Gly   90

271   AAA AGA GTT TCT ATT AAA GAT TTG TTA GAT GAA AAA GAT TTT GAA   315
91    Lys Arg Val Ser Ile Lys Asp Leu Leu Asp Glu Lys Asp Phe Glu   105

316   ATA TGG GCA ATT AAT GAA CAG ACG ATG AAG CTA GAA TCA GCT AAA   360
106   Ile Trp Ala Ile Asn Glu Gln Thr Met Lys Leu Glu Ser Ala Lys   120

361   GTT AGT CGT GTA TTT GCT ACT GGC AAA AAG CTA GTT TAT ATT TTA   405
121   Val Ser Arg Val Phe Ala Thr Gly Lys Lys Leu Val Tyr Ile Leu   135

406   AAA ACT CGA CTA GGT AGA ACT ATC AAG GCA ACA GCA AAT CAT AGA   450
136   Lys Thr Arg Leu Gly Arg Thr Ile Lys Ala Thr Ala Asn His Arg   150

451   TTT TTA ACT ATT GAT GGT TGG AAA AGA TTA GAT GAG CTA TCT TTA   495
151   Phe Leu Thr Ile Asp Gly Trp Lys Arg Leu Asp Glu Leu Ser Leu   165

496   AAA GAG CAT ATT GCT CTA CCC CGT AAA CTA GAA AGC TCC TCT TTA   540
166   Lys Glu His Ile Ala Leu Pro Arg Lys Leu Glu Ser Ser Ser Leu   180

541   CAA TTA TAG
181   Gln Leu End
```

En bleu : séquence du peptide signal de DsbA, en rouge : séquence du peptide à cycliser, en noir : séquence des fragments $DnaBI_C$ et $DnaBI_N$ de l'intéine, en bleu clair : mutation de la cystéine en alanine dans la séquence de l'intéine $DnaBI_N$

Résultats de spéctrométrie de masse permettant de mettre en évidence la glutathionylation de précurseur du clone SKISRG