

Université du Québec  
INRS-Institut Armand-Frappier

**Développement de vecteurs et d'hôtes bactériens pour l'expression de protéines recombinantes et l'analyse métagénomique**

Par

François-Xavier Lussier

Thèse présentée pour l'obtention du grade de  
Philosophiae Doctor (Ph.D.) en biologie

Jury d'évaluation

Président du jury et examinateur interne : Richard Villemur (Ph.D.)  
INRS-Institut Armand-Frappier

Examinateur externe : Ryszard Brzezinski (Ph.D.)  
Université de Sherbrooke

Examinateur externe : Denis Groleau (Ph.D.)  
Institut de recherche en  
biotechnologie du CNRC

Directeur de recherche : François Shareck (Ph.D.)  
INRS-Institut Armand-Frappier



## RÉSUMÉ

Les bactéries sont de formidables outils biotechnologiques, permettant notamment la production de protéines recombinantes à des fins diverses. Les systèmes d'expression bactériens sont essentiels à de nombreux domaines de recherche, ainsi qu'à l'industrie des biotechnologies. Les deux composantes d'un système d'expression sont le vecteur et l'hôte cellulaire. La bactérie Gram-négatif *Escherichia coli* est l'hôte d'expression le plus couramment utilisé et est bien établie dans de nombreux laboratoires. L'utilisation d'*E. coli* n'est cependant pas un gage de succès et il peut être nécessaire de faire appel à des hôtes alternatifs pour obtenir une production adéquate. *Streptomyces lividans* et *Bacillus subtilis* sont deux bactéries Gram-positif présentant des qualités particulièrement intéressantes pour la production de protéines recombinantes.

Pour la découverte de nouvelles enzymes, l'approche métagénomique permet d'accéder à la diversité génétique des microorganismes non-cultivables. Pour ce faire, l'ADN génomique des microorganismes présents dans un échantillon donné est extrait, cloné dans un vecteur et introduit dans un hôte hétérologue pour ainsi créer une banque métagénomique. Cette banque peut ensuite être criblée pour identifier des gènes d'intérêt. L'approche métagénomique requiert donc un vecteur pour accueillir l'ADN génomique et un hôte permettant l'expression des gènes. Comme pour l'expression de protéines recombinantes, *E. coli* est l'hôte le plus utilisé. Ceci s'explique entre autres par le manque de systèmes alternatifs.

Les travaux présentés dans cette thèse ont été réalisés au sein d'une étude métagénomique visant l'identification de nouvelles enzymes d'intérêt industriel. Cette thèse présente le développement de nouveaux systèmes d'expression utilisant *S. lividans* et *B. subtilis* comme hôtes. Deux vecteurs multifonctionnels et deux souches d'expression ont été construits. Leur fonctionnalité a été démontrée par la production d'enzymes sécrétées. Le système d'expression de *S. lividans* a aussi été utilisé pour la construction et le criblage d'une banque métagénomique,

permettant ainsi l'identification d'une nouvelle enzyme lipolytique possédant une application industrielle potentielle.



Directeur de recherche



Étudiant

## REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche, François Shareck, pour m'avoir accueilli dans son laboratoire pour toutes mes études graduées. Merci pour l'encadrement et de m'avoir donné les bases nécessaires pour poursuivre dans le domaine de la recherche. Je désire également remercier François Denis pour ses nombreux conseils et sa passion contagieuse, il fut mon codirecteur pendant une bonne partie de mon doctorat.

J'aimerais aussi remercier Liette Biron, qui a grandement contribué à mon apprentissage des différentes techniques de laboratoire. Merci aussi à Raymonde Jetté pour sa grande gentillesse et son support quotidien. Je remercie également Amélie Côté, Catherine Meilleur, Guillaume Brault et Olivier Chambenoit pour les discussions enrichissantes et l'ambiance dynamique créée dans le laboratoire.

Merci à mes parents pour leur appui inconditionnel, je leur serai éternellement reconnaissant. Finalement, merci à mon épouse Alexia pour son support et ses encouragements, et d'être tout simplement une partenaire de vie extraordinaire.

# TABLE DES MATIÈRES

RÉSUMÉ .....	ii
REMERCIEMENTS .....	iv
TABLE DES MATIÈRES .....	v
LISTE DES FIGURES .....	ix
LISTE DES TABLEAUX.....	xi
LISTES DES ABRÉVIATIONS .....	xii
INTRODUCTION .....	13
PARTIE I: PRODUCTION DE PROTÉINES RECOMBINANTES .....	15
1. REVUE DE LITTÉRATURE.....	16
1.1 Introduction.....	17
1.2 <i>Escherichia coli</i> .....	17
1.2.1 Souches d'expression.....	18
1.2.2 Réplicons.....	21
1.2.3 Promoteurs .....	21
1.3 Système d'expression T7 .....	24
1.3.1 ARN polymérase du phage T7.....	24
1.3.2 Souches .....	26
1.3.3 Vecteurs .....	26
1.3.4 Expression basale.....	27
1.4 <i>Streptomyces lividans</i> .....	28
1.4.1 Souches .....	29
1.4.2 Plasmides .....	32
1.4.3 Promoteurs .....	33
1.4.4 Traduction .....	34
1.4.5 Sécrétion .....	34
1.5 <i>Bacillus subtilis</i> .....	35
1.5.1 Souches .....	37
1.5.2 Vecteurs d'expression.....	37
1.5.3 Promoteurs .....	41

1.5.4 RBS .....	42
1.5.5 Sécrétion .....	43
1.5.6 Système d'expression T7 de <i>B. subtilis</i> .....	43
2. PRÉSENTATION DE L'ARTICLE 1 .....	45
2.1. Mise en contexte de l'article 1 .....	46
2.2. Contribution de l'étudiant à l'article 1 .....	46
2.3. Article 1: Adaptation of the highly productive T7 expression system to <i>Streptomyces lividans</i> .....	47
2.4 Résumé.....	48
2.5 Abstract .....	49
2.6 Introduction.....	49
2.7 T7 RNA polymerase production in <i>S. lividans</i> .....	50
2.8 <i>S. lividans</i> T7 expression strain.....	52
2.9 Bifunctional T7 expression vector .....	53
2.10 T7 RNA polymerase-directed xylanase production.....	56
2.11 Conclusion .....	58
2.12 Acknowledgments.....	58
2.13 Supplemental material.....	58
2.13.1 T7 RNA polymerase gene mutation .....	58
2.13.2 Construction of pFXPtipAT7.....	59
2.13.3 Construction of the pFX583.....	59
3. DÉVELOPPEMENT D'UN SYSTÈME D'EXPRESSION INDUCTIBLE CHEZ <i>BACILLUS SUBTILIS</i> .....	64
3.1. Mise en contexte .....	65
3.2 Contribution de l'étudiant .....	65
3.3 Résumé.....	66
3.4 Abstract .....	66
3.5 Introduction .....	67
3.6 Materials and methods .....	69
3.6.1 Bacterial strains, plasmids and oligonucleotides .....	69
3.6.2 Culture media, growth conditions and recombinant gene expression .	71

3.6.3 DNA preparation and manipulation.....	71
3.6.4 Construction of the <i>B. subtilis</i> T7-expression host .....	72
3.6.5 Construction of the T7-expression vector.....	72
3.6.6 T7 RNA polymerase-directed gene expression .....	75
3.6.7 Analysis of recombinant protein production.....	75
3.7 Results.....	76
3.7.1 <i>B. subtilis</i> T7-expression host .....	76
3.7.2 Shuttle T7-expression vector .....	76
3.7.3 Extracellular protein production .....	77
3.8 Discussion .....	79
3.9 Acknowledgements.....	83
PARTIE II: .....	84
CONSTRUCTION ET CRIBLAGE D'UNE BANQUE MÉTAGÉNOMIQUE .	84
4. REVUE DE LITTÉRATURE .....	85
4.1 Métagénomique.....	86
4.2 Construction de banques métagénomiques .....	88
4.2.1 Source d'ADN .....	88
4.2.2 Isolement d'ADN.....	91
4.2.3 Types de banques.....	91
4.3 Criblage .....	92
4.3.1 Criblage dépendant de la séquence .....	92
4.3.2 Criblage fonctionnel.....	93
4.4 Hôtes hétérologues .....	94
5. PRÉSENTATION DE L'ARTICLE 3 .....	96
5.1 Mise en contexte de l'article 3 .....	97
5.2 Contribution de l'étudiant à l'article 3 .....	97
5.3 Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector.....	98
5.4 Résumé .....	99
5.5 Abstract .....	99
5.6 Introduction.....	100

5.7 Materials and Methods.....	102
5.7.1 Bacterial strains, culture media and growth conditions .....	102
5.7.2 Sequencing Fed-Batch Reactor enrichment.....	103
5.7.3 Library construction.....	104
5.7.4 Functional-based screening and gene isolation.....	104
5.7.5 Partial enzymatic characterization .....	104
5.8 Results.....	105
5.8.1 Metagenomic library construction .....	105
5.8.2 Function-based screening.....	105
5.8.3 Subcloning and sequence analysis .....	106
5.8.4 Partial biochemical characterization .....	108
5.9 Discussion .....	112
5.10 Acknowledgments.....	115
6. DISCUSSION .....	116
6.1 Systèmes d'expression .....	117
6.1.1 Souches .....	117
6.1.2 Vecteurs .....	119
6.1.3 Performances des systèmes d'expression .....	119
6.2 Construction et criblage d'une banque métagénomique .....	121
6.2.1 Vecteurs pFX583 et pFX64 .....	122
6.2.2 Hôtes de criblage alternatifs.....	123
7. CONCLUSION .....	124
RÉFÉRENCES .....	125

## LISTE DES FIGURES

<b>Figure 1.</b> Représentation du système d'expression T7.....	26
<b>Figure 2.</b> Vecteurs d'expression T7 pET-3a.....	29
<b>Figure 3.</b> Cycle de vie des streptomycètes.....	31
<b>Figure 4.</b> Intégration chromosomique par simple (A) et double (B) recombinaison homologue .....	42
<b>Figure 5.</b> Analysis of the T7 RNAPol production by (A) Coomassie-stained SDS-PAGE and (B) Western blot.....	53
<b>Figure 6.</b> Structure of the integrative plasmid pFXPtipAT7.....	55
<b>Figure 7.</b> Structure of the bifunctional T7-expression vector pFX583.....	57
<b>Figure 8.</b> (A) Coomassie-stained SDS-PAGE showing the extracellular production of XlnA2 in non-induced and induced cultures. (B) Xylanase activity in non-induced and induced cultures.....	59
<b>Figure 9.</b> Xylanase activity vs. concentration of thiostrepton used for induction.....	64
<b>Figure 10.</b> Xylanase activity from cultured induced at different times.....	65
<b>Figure 11.</b> Xylanase production in <i>E. coli</i> BL21(DE3)/pFX583xlnA2. ....	66
<b>Figure 12.</b> Structure of the integrative plasmid pAXT7.....	76
<b>Figure 13.</b> Structure of the expression vector pFX64.....	77
<b>Figure 14.</b> Agar plate assay for the evaluation of the $\alpha$ -amylase production.....	81

<b>Figure 15.</b> Coomassie-stained SDS-PAGE showing the extracellular production of $\alpha$ -amylase.....	82
<b>Figure 16.</b> Étapes générales d'une étude métagénomique.....	92
<b>Figure 17.</b> Multiple amino acid sequence alignment performed with ClustalW2.....	108
<b>Figure 18.</b> Effect of the chain length on the activity of LipF5-11. ....	110
<b>Figure 19.</b> Effect of pH on the activity of LipF5-11. ....	112
<b>Figure 20.</b> Thermostability assay.....	113

## LISTE DES TABLEAUX

<b>Tableau I.</b> Souches d'expression.....	22
<b>Tableau II.</b> Protéines produites chez <i>S. lividans</i> .....	33
<b>Tableau III.</b> Réplicons utilisés dans les vecteurs d'expression de <i>S. lividans</i> ....	35
<b>Tableau IV.</b> Protéines recombinantes produites chez <i>B. subtilis</i> .....	38
<b>Tableau V.</b> Souches de <i>B. subtilis</i> possédant une activité protéolytique réduite..	39
<b>Tableau VI.</b> Réplicons utilisés chez <i>B. subtilis</i> .....	41
<b>Table VII.</b> Oligonucleotides used in this work.....	63
<b>Table VIII.</b> Strains and plasmids used in this work.....	73
<b>Table IX.</b> Oligonucleotides used in this work.....	74
<b>Tableau X.</b> Enzymes découvertes grâce à l'approche métagénomique.....	89
<b>Table XI.</b> Strains and plasmids used in this work.....	104

## LISTES DES ABRÉVIATIONS

ADN	Acide désoxyribonucléique
AMPc	Adénosine monophosphate cyclique
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
ARNpol T7	ARN polymérase du phage T7
ARNr	Acide ribonucléique ribosomal
BAC	Chromosome artificiel bactérien
CR	Cercle roulant
CRE	Élément de réponse à l'AMPc
GRAS	Generally recognized as safe
IPTG	Isopropyl- $\beta$ -D-1-thiogalactopyranoside
OD	Densité optique
<i>ori</i> DB	Origine de réPLICATION double brin
<i>ori</i> SB	Origine de réPLICATION simple brin
<i>oriT</i>	Origine de transfert
RBS	Site de fixation des ribosomes
SD	Shine-Dalgarno
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel
SFBR	Sequencing Fed-Batch Reactor
T7 RNAPol	T7 RNA polymerase
$\Delta G$	Variation d'enthalpie libre

## INTRODUCTION

Les enzymes sont de remarquables catalyseurs agissant dans une infinité de réactions moléculaires. Elles trouvent des applications dans une grande diversité d'industries et de domaines de recherche. Dans le secteur industriel, les enzymes sont de plus en plus utilisées comme biocatalyseurs et remplacent les procédés chimiques traditionnels. Le marché des enzymes industrielles atteignait 2 milliards de dollars au début de la décennie et serait d'environ 2,5 milliard aujourd'hui (Demain et Vaishnav 2009). Les protéases occupent la plus grande part du marché. Viennent ensuite les amylases, les xylose isomérasées, les lactases, les lipases, les cellulases et les xylanases. Environ la moitié de ces enzymes proviennent de champignons et de levures, et le tiers sont d'origine bactérienne (Demain et Vaishnav 2009). L'avenue de la technologie de l'ADN recombinant a rendu possible l'expression de gènes d'intérêt dans un hôte hétérologue. Aujourd'hui, la majorité des enzymes industrielles sont produites grâce à cette approche (Drepper *et al.* 2006; Soetaert et Vandamme 2006). L'accès à des systèmes d'expression performants est donc essentiel à l'utilisation d'enzymes industrielles. La découverte de nouveaux biocatalyseurs est aussi indispensable à l'expansion de cette biotechnologie.

Les bactéries sont présentes dans pratiquement tous les environnements terrestres et aquatiques, incluant les plus extrêmes. Un seul gramme de terre pourrait contenir jusqu'à 18000 génomes procaryotes distincts (Torsvik *et al.* 1996; Torsvik *et al.* 1998; Doolittle 1999; Torsvik et Ovreas 2002). Les bactéries sont donc un réservoir extraordinaire de diversité génétique. Traditionnellement, la recherche de nouvelles enzymes se faisait à partir de microorganismes isolés en cultures pures. Cependant, une très faible proportion des bactéries existantes pourraient être cultivées en laboratoire. La grande majorité des gènes bactériens seraient donc inaccessibles par l'approche classique. Pour accéder à ce matériel génétique, l'approche métagénomique a émergé vers la fin des années 1990 (Handelsman *et al.* 1998). Cette approche consiste à extraire l'ensemble des génomes microbiens d'un environnement donné pour ensuite les introduire dans

un hôte hétérologue. La banque métagénomique ainsi obtenue peut ensuite être criblée pour identifier des gènes d'intérêt. Comme pour la production de protéines recombinantes, la construction et le criblage de banques métagénomiques requièrent l'utilisation de vecteurs et d'hôtes d'expression.

Les travaux de recherche présentés dans cette thèse ont été réalisés dans le cadre général d'une étude métagénomique visant à identifier de nouvelles enzymes d'intérêt industriel. Le principal objectif était de développer de nouveaux systèmes d'expression multifonctionnels pouvant être utilisés à la fois pour la production de protéines recombinantes et l'étude métagénomique. Cette thèse est présentée sous forme d'articles scientifiques et se divise en deux grandes parties. La partie I traite du développement des systèmes d'expression, alors que la partie II démontre l'utilité de ces systèmes dans la découverte de nouvelles enzymes par l'approche métagénomique. Chacune des parties est précédée par une revue de littérature. La thèse se termine par une discussion générale englobant chacune des sections.

## **PARTIE I: PRODUCTION DE PROTÉINES RECOMBINANTES**

## **1. REVUE DE LITTÉRATURE**

## **1.1 Introduction**

Les protéines forment la classe la plus nombreuse et la plus diversifiée des biomolécules. Elles jouent un rôle dans pratiquement tous les aspects fonctionnels et structuraux de la cellule: biocatalyse, régulation, signalisation, transport, motilité et formation du cytosquelette. Avec l'essor du génie génétique du début des années 70 est apparue la possibilité de construire des molécules d'ADN composées de séquences provenant de différents organismes. Cette nouvelle technologie de l'ADN recombinant a rendu possible le clonage de fragments d'ADN d'intérêt, dont ceux codant pour des protéines. La production de protéines recombinantes est aujourd'hui essentielle à de nombreux domaines de recherche, aussi bien au niveau fondamental qu'appliqué, ainsi qu'à l'industrie des biotechnologies: pharmaceutique, agroalimentaire et industrielle (Demain et Vaishnav 2009).

La production de protéines recombinantes requiert que le gène soit cloné dans un vecteur d'expression en aval d'un promoteur, pour en permettre la transcription et la traduction. La machinerie moléculaire nécessaire à l'expression est fournie par la cellule hôte ou peut provenir d'un extrait cellulaire. Une grande variété de systèmes d'expression sont disponibles, basés sur des hôtes bactériens, des levures, des champignons, des cellules de mammifères, d'insectes et de plantes (Schmidt 2004). Il existe aussi différents systèmes acellulaires, mais leur coût d'utilisation demeure encore prohibitif, particulièrement pour la production de grandes quantités de protéines (Kim et Kim 2009). Parce que chaque protéine est différente, il ne peut y avoir de système d'expression unique. Les facteurs à considérer dans le choix d'un système d'expression incluent le rendement, la vitesse de production, la fonctionnalité de la protéine et l'utilisation visée.

## **1.2 *Escherichia coli***

*Escherichia coli* est une  $\gamma$ -protéobactérie Gram-négatif dont la génétique, la physiologie et la biochimie sont très bien caractérisées. Parmi les nombreux hôtes

disponibles pour l'expression de gènes recombinants, *E. coli* demeure le plus couramment utilisé. Ceci peut entre autres s'expliquer par sa capacité à croître rapidement à haute densité dans des milieux peu coûteux, par sa facilité d'utilisation et par la grande diversité de vecteurs et de souches disponibles (Baneyx 1999).

### 1.2.1 Souches d'expression

Les souches d'*E. coli* aujourd'hui utilisées en laboratoire dérivent très majoritairement des souches K-12 et B, issues respectivement de travaux précurseurs en génétique et sur les bactériophages (Gray et Tatum 1944; Delbrück 1945; Demerec et Fano 1945; Tatum 1945; Tatum et Lederberg 1947). Les génomes d'*E. coli* K-12 et B sont identiques à environ 99 %, ne différant principalement que par quelques délétions et insertions (Studier *et al.* 2009). Une différence digne de mention est l'incapacité d'*E. coli* B à produire la protéase ATP-dépendante Lon, dû à l'insertion de l'élément IS186 dans le promoteur du gène (saiSree *et al.* 2001). Une autre différence provient d'une délétion de huit gènes (*gspD-K*) chez *E. coli* K-12 (Jeong *et al.* 2009). Ce groupe de gènes pourrait être impliqué dans le mécanisme de sécrétion de type II (Studier *et al.* 2009). *E. coli* B pourrait ainsi posséder une capacité de sécrétion absente chez *E. coli* K-12.

En tant qu'outil biomoléculaire, *E. coli* est utilisée pour le clonage et l'expression de gènes. À ces fins, des souches aux propriétés distinctes ont été développées. Les souches de clonage dérivent généralement d'*E. coli* K-12 et possèdent deux caractéristiques communes, soit les mutations *endA1* et *recA1*. Ces mutations confèrent respectivement une activité endonucléase et une capacité de recombinaison réduites (Willetts *et al.* 1969; Wright 1971). La mutation *endA1* limite la dégradation de l'ADN lors des extractions de plasmides et la mutation *recA1* stabilise les constructions (Vapnek *et al.* 1976; Taylor *et al.* 1993). Ces souches sont toutefois inappropriées pour la production de protéines. De par leur capacité réduite à réparer l'ADN endommagé et une fréquence réduite de

réPLICATION de leur chromosome (Whittier et Chase 1981; Skarstad et Boye 1988), les souches *recA1* ont une croissance plus lente et un taux de survie réduit.

Les souches développées pour l'expression de gènes dérivent principalement d'*E. coli* B. Une des premières souches d'expression décrite dans la littérature, *E. coli* BL21, découle de la transduction de la souche B834 par le phage P1 afin de rétablir le phénotype Met<sup>+</sup> (Studier et Moffatt 1986; Daegelen *et al.* 2009). Dérivant d'*E. coli* B, la souche BL21 ne produit pas la protéase Lon. Elle a aussi hérité de la mutation acquise par la souche B834, inactivant la protéase membranaire extracellulaire OmpT (Grodberg et Dunn 1988). Ces deux mutations font d'*E. coli* BL21 une souche particulièrement adaptée pour l'expression de gènes hétérologues. Les protéases Lon et OmpT sont en effet connues pour dégrader les protéines cytoplasmiques et sécrétées (Buell *et al.* 1985; White et Richardson 1987; Grodberg et Dunn 1988; Faraldo *et al.* 1991; Kelsall et Kulski 1995). Un peu moins répandus, *E. coli* K-12 et ses dérivés sont tout de même couramment utilisés pour l'expression routinière de gènes hétérologues. Le tableau I présente les souches d'expression les plus connues. Ces dernières sont toutes disponibles commercialement et sont adaptées pour différentes applications.

**Tableau I.** Souches d'expression.

Souche	Origine	Caractéristiques
B834	B	Auxotrophe pour la méthionine, permet l'étiquetage $^{35}\text{S}$ -met
BL21	B834	Ne produit pas les protéases Lon et OmpT
BLR	BL21	Mutant <i>recA</i> , stabilise certaines constructions; ne produit pas les protéases Lon et OmpT
C41 (C43)	BL21	Mutants adaptés pour la production de protéines membranaires
Origami	K-12	Mutant <i>trxb</i> et <i>gor</i> , facilite la formation de ponts disulfures dans le cytoplasme
Origami B	BL21	Mutant <i>trxb</i> et <i>gor</i> ; facilite la formation de ponts disulfures dans le cytoplasme; ne produit pas les protéases Lon et OmpT
Rosetta	BL21	Favorise l'expression de gènes contenant des codons rarement utilisés par <i>E. coli</i> ; ne produit pas les protéases Lon et OmpT
Rosetta-gami	Origami (K12)	Favorise l'expression de gènes contenant des codons rarement utilisés par <i>E. coli</i> ; mutant <i>trxb</i> et <i>gor</i> , facilite la formation de ponts disulfures dans le cytoplasme
Rosetta-gami B	Origami B	Favorise l'expression de gènes contenant des codons rarement utilisés par <i>E. coli</i> ; ne produit pas les protéases Lon et OmpT; mutant <i>trxb</i> et <i>gor</i> , facilite la formation de ponts disulfures dans le cytoplasme
Tuner	BL21	Mutant $\Delta lacZY$ , permet un contrôle précis du niveau d'expression induit par l'IPTG

### **1.2.2 Réplicons**

Pour surexprimer un gène, la séquence codante est généralement clonée dans un vecteur plasmidique multi-copies. Le réplicon le plus utilisé dans les vecteurs d'expression provient du plasmide pMB1 isolé d'*E. coli* (Bolivar 1979). Outre l'origine de réPLICATION, la séquence nécessaire à la réPLICATION comprend deux régions transcrites, générant les ARN partiellement complémentaires I et II. L'extrémité 3' de l'ARN II s'hybride à une région près de l'origine de réPLICATION et permet l'initiation de la réPLICATION (Kues et Stahl 1989). La réPLICATION progresse de façon unidirectionnelle selon le mécanisme de type thêta (del Solar *et al.* 1998). L'initiation de la réPLICATION est régulée par l'entremise de l'ARN I, qui en s'hybridant à l'ARN II, empêche la formation de l'hybride ADN-ARN II (Lacatena et Cesareni 1983). La fréquence d'initiation est aussi influencée par la protéine Rop, qui stabilise l'hybridation non-productive entre l'ARN I et l'ARN II (Lacatena *et al.* 1984). Rop n'est pas essentielle à la réPLICATION et la délétion de son gène a pour effet d'augmenter le nombre de copies par cellule du plasmide (Atlung *et al.* 1999). Basé sur le réplicon pMB1, le plasmide pBR322 a longtemps été le vecteur le plus utilisé pour le clonage et l'expression de gènes recombinants (Itakura *et al.* 1977; Goeddel *et al.* 1979). Ce plasmide présente un nombre de copies par cellule variant entre 15 et 20 (Lee *et al.* 2006a). Plusieurs vecteurs d'expression utilisés aujourd'hui dérivent de pBR322, notamment la série de vecteurs pET basés sur le promoteur T7 (Novagen) (Rosenberg *et al.* 1987).

### **1.2.3 Promoteurs**

Un grand nombre de promoteurs ont été décrits dans la littérature comme pouvant être utilisés pour la production de protéines recombinantes, mais seul un petit nombre d'entre eux sont couramment utilisés. Des promoteurs constitutifs ont quelquefois été utilisés avec succès pour la production de protéines hétérologues (Chauhan *et al.* 2001; Haryoung Poo 2002). La production constitutive de protéines peut toutefois entraîner une diminution du taux de croissance de la

cellule hôte et par le fait même la productivité du système. Dans le cas de protéines toxiques, cela peut mener à d'importants défauts de croissance et à la mort de la cellule (Saida *et al.* 2006). Par conséquent, les promoteurs inductibles sont généralement mieux adaptés à la production de protéines, permettant l'expression au moment où la densité cellulaire optimale est atteinte. Idéalement, un promoteur doit permettre un haut taux de transcription, avoir un faible niveau d'expression basale et son induction doit être simple et peu coûteuse.

L'opéron lactose (*lac*), impliqué dans le métabolisme du lactose, fut pendant de nombreuses années le paradigme de la régulation génique chez les procaryotes. Il est composé de trois gènes structuraux, *lacZYA*, codant respectivement pour une  $\beta$ -galactosidase, une perméase et une transacétylase (Wilson *et al.* 2007). L'opéron *lac* est contrôlé par le répresseur LacI, qui inhibe la transcription en l'absence de lactose (Riggs *et al.* 1970). Le promoteur *lac* (*P<sub>lac</sub>*) est aussi sujet à la répression catabolique en présence de glucose (Loomis et Magasanik 1966; Kolb *et al.* 1993). Le promoteur *lac* a été utilisé pour la première fois dans l'expression de gènes recombinants il y a plus de 30 ans (Polisky *et al.* 1976). Depuis, il a servi dans un grand nombre de systèmes d'expression (Sweet 2003). Le promoteur *lac* est toutefois relativement faible et possède une expression basale non-négligeable (Makoff et Oxer 1991). La fuite d'expression en absence d'inducteur peut toutefois être réduite en ajoutant du glucose au milieu de culture (Grossman *et al.* 1998). Une version de *P<sub>lac</sub>* moins sensible à l'AMPc, *P<sub>lacUV5</sub>*, permet d'obtenir un niveau de transcription légèrement plus élevé, demeurant constant pendant les différents stades de croissance (Wanner *et al.* 1977; Wanner *et al.* 1978).

Dans le but d'obtenir des niveaux de transcription plus élevés, des promoteurs synthétiques hybrides ont été développés à partir de *P<sub>lac</sub>*. Les promoteurs synthétiques *tac* et *trc* sont composés de la région -35 du promoteur *trp* et de la région -10 de *P<sub>lac</sub>* (de Boer *et al.* 1983; Brosius *et al.* 1985). Le promoteur *tac* est au moins cinq fois plus performant que *P<sub>lacUV5</sub>*, permettant une accumulation jusqu'à 15-30% des protéines totales. Ayant la même expression basale que *P<sub>lac</sub>*, cette plus grande activité peut être problématique pour

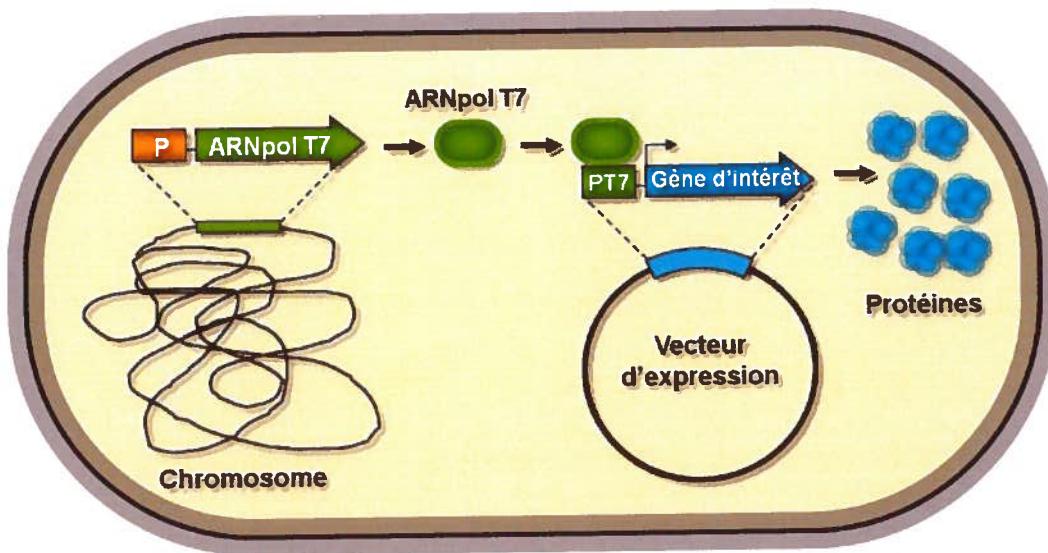
l'expression de gènes dont le produit est toxique. Le promoteur *lac* et ses dérivés, dont la régulation est médiée par LacI, peuvent être induits par un analogue non-hydrolysable du lactose, l'isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG).

Un autre promoteur utilisé dans l'expression de gènes recombinants provient de l'opéron arabinose (*ara*), lequel est composé des gènes codant pour une isomérase (*araA*), une ribulokinase (*araB*) et une épimérase (*araD*) (Greenfield *et al.* 1978). La régulation de la transcription se fait par l'entremise du produit du gène *araC* (Hirsh et Schleif 1977). Le mécanisme de régulation de l'opéron *ara* est légèrement plus complexe que celui de l'opéron *lac*. Le régulateur AraC agit en effet comme répresseur en absence d'arabinose et comme activateur en sa présence (Ogden *et al.* 1980). Le promoteur de l'opéron *ara* (PBAD) permet donc une expression finement régulée, de faible à forte selon la dose d'inducteur. Comme pour le promoteur *lac*, PBAD est sensible à la répression catabolique et l'ajout de glucose dans le milieu de culture permet de réduire les fuites d'expression. Avec une souche dont les gènes *araBAD* sont inactifs, il est possible d'obtenir une expression maximale avec aussi peu que 0,001 % d'arabinose. Par contre, avec une souche capable de métaboliser l'arabinose, 1 % d'inducteur est requis pour obtenir une induction complète (Mayer 1995). D'autres promoteurs moins populaires incluent celui du gène *tetA* et de l'opéron *rhaA*, induits respectivement par la tétracycline et le rhamnose.

Parmi l'ensemble des promoteurs utilisés pour l'expression de gènes recombinants, le promoteur du phage T7 est sans aucun doute le plus populaire (Terpe 2006). À la différence des promoteurs décrits précédemment, le promoteur T7 n'est reconnu que par l'ARN polymérase du phage T7. Pour que ce promoteur puisse être utilisé dans un vecteur d'expression, l'hôte doit avoir la capacité de produire l'ARN polymérase du phage T7. Pour ce faire, le gène codant pour l'ARN polymérase T7 peut être intégré dans le chromosome de l'hôte ou être porté par un plasmide.

## 1.3 Système d'expression T7

L'un des systèmes d'expression le plus utilisé est celui basé sur l'ARN polymérase du phage T7 (système T7), développé chez *Escherichia coli* par Studier et Moffatt (1986). Dans le système T7, le gène d'intérêt est placé sous le contrôle d'un promoteur tardif du phage et la transcription est effectuée grâce à l'ARN polymérase phagique (ARNpol T7) produite par la cellule hôte (Figure 1).



**Figure 1.** Représentation du système d'expression T7. Le gène codant pour l'ARNpol T7 est placé sous le contrôle d'un promoteur inductible (P) dans le chromosome de la cellule hôte. En présence de l'inducteur, l'ARNpol T7 est produite et transcrit spécifiquement le gène d'intérêt cloné dans le vecteur d'expression, en aval d'un promoteur T7 (PT7).

### 1.3.1 ARN polymérase du phage T7

Chez les cellules procaryotes et eucaryotes, la transcription s'effectue grâce à des ARN polymérases composées de plusieurs sous-unités. Par contre, les ARN

polymérases de certains phages, comme T7, T3, SP6 et K11, ont une structure beaucoup plus simple et ne sont composées que d'une seule unité (McAllister et Raskin 1993; Cermakian *et al.* 1997). Ces ARN polymérases ont la capacité d'effectuer un cycle complet de transcription sans l'intervention de facteurs protéiques supplémentaires. Elles sont aussi caractérisées par leur activité de synthèse élevée et leur grande spécificité envers leurs promoteurs.

L'ARN polymérase la plus étudiée et la mieux caractérisée est celle du bactériophage T7. L'ARNpol T7 a été isolée pour la première fois chez des cellules d'*Escherichia coli* infectées par le phage T7 (Chamberlin *et al.* 1970). Cette enzyme est composée de 883 acides aminés et possède une masse de 98 kDa (Moffatt *et al.* 1984). L'ARNpol T7 est produite durant la phase précoce de l'infection et joue un rôle clé dans la régulation de l'expression des gènes du bactériophage T7. La spécificité de l'ARNpol T7 envers les promoteurs T7 tardifs est centrale dans la stratégie d'infection du phage, lui permettant de transcrire uniquement ses propres gènes (McAllister et Carter 1980). Les promoteurs T7 tardifs contiennent tous une séquence nucléotidique conservée de 23 nucléotides, s'étendant de - 17 à + 6 relativement au point de départ de transcription (Dunn et Studier 1983). Cette séquence de 23 nucléotides requise pour être reconnue par l'ARNpol T7 fait en sorte qu'il est très peu probable qu'une telle séquence soit présente dans l'ADN de l'hôte (Studier et Moffatt 1986). L'ARNpol T7 initie la synthèse d'ARN de façon très efficace et effectue l'elongation environ cinq fois plus rapidement que l'ARN polymérase d'*E. coli* (Golomb et Chamberlin 1974). De plus, contrairement aux ARN polymérases bactériennes, l'ARNpol T7 n'est pas inactivée par la rifampicine (Chamberlin et Ring 1973). Lorsque l'ARNpol T7 rencontre un terminateur de transcription du phage T7 ( $T\phi$ ), l'elongation est en grande partie arrêtée (Carter *et al.* 1981). Par contre, les terminateurs de transcription provenant d'autres sources sont beaucoup moins efficaces et permettent la synthèse d'ARN de plusieurs dizaines de milliers de bases (McAllister *et al.* 1981). Ainsi, l'ARNpol T7 a la capacité de générer des transcrits complets de pratiquement tout ADN lié à un promoteur T7.

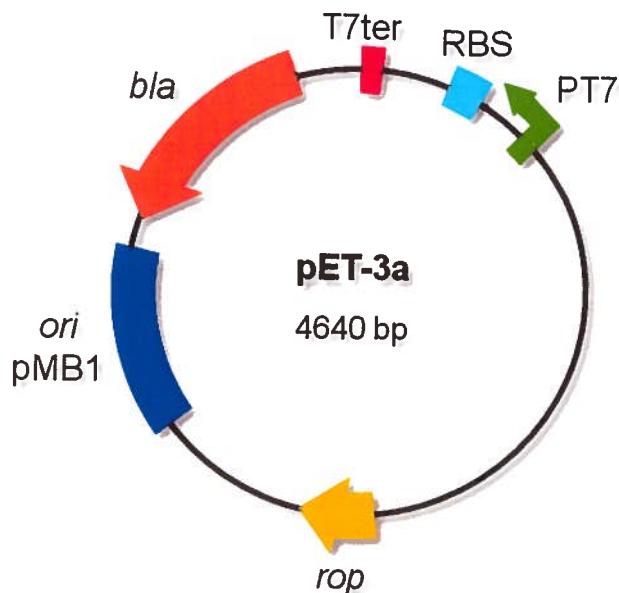
### **1.3.2 Souches**

Dans le système T7 d'*E. coli*, les souches d'expression utilisées contiennent une copie chromosomique du gène codant pour l'ARNpol T7 (gène *I*). La grande majorité des souches d'expression T7 sont des lysogènes du phage DE3, dérivé du vecteur *lambda* D69 comprenant le gène *I* sous le contrôle du promoteur *lacUV5*. Le segment *lacUV5*-gène *I* est inséré dans le gène *int* du vecteur *lambda* D69, ce qui stabilise l'intégration du phage DE3 (Studier et Moffatt 1986). Ainsi, l'ajout d'IPTG au milieu de culture induit la production d'ARNpol T7, qui à son tour permet la transcription du gène d'intérêt présent dans le vecteur d'expression. Il existe aussi d'autres souches dans lesquelles la production d'ARNpol T7 est induite par l'ajout de NaCl. Dans ces souches, le gène de l'ARNpol T7 est sous le contrôle du promoteur de l'opéron *proU* d'*E. coli* (Gowrishankar et Manna 1996). Ce promoteur permet de contrôler la production d'ARNpol T7 de façon plus efficace que le promoteur *lacUV5*, idéal pour la production de protéines toxiques (Bhandari et Gowrishankar 1997). La première souche d'*E. coli* lysogène DE3 créée par Studier et Moffatt (1986) a été générée à partir de la souche BL21. Cette dernière est spécialement adaptée pour la production de protéines recombinantes, par son activité protéolytique réduite. Depuis la création de la première souche lysogène DE3, de nombreuses autres souches ont été générées pour différentes applications (Novagen). Plus récemment, des souches d'expression T7 ont été développées en intégrant directement le gène *I* dans le chromosome, sans l'utilisation de phages lysogéniques (New England Biolabs).

### **1.3.3 Vecteurs**

Les premiers vecteurs développés spécifiquement pour l'expression dirigée par l'ARNpol T7 ont été construits en introduisant dans le plasmide pBR322 le promoteur du gène *I0* du phage T7, ainsi que le site de liaison des ribosomes (RBS) et le terminateur de transcription du même gène (Rosenberg *et al.* 1987) (Figure 2). Depuis, plusieurs autres vecteurs ont été mis au point, présentant des

fonctionnalités améliorées pour le clonage, la détection et la purification des protéines (Novagen). Ainsi, certains vecteurs permettent de fusionner la protéine d'intérêt à une étiquette peptidique pouvant être utilisée à différentes fins. Par exemple, les étiquettes poly-histidines permettent la purification de la protéine d'intérêt par chromatographie d'affinité et la détection par immunobuvardage de type Western. D'autres étiquettes, telles que NUS-tag et Trx-tag, peuvent faciliter le repliement de la protéine et augmenter sa solubilité (Davis *et al.* 1999).



**Figure 2.** Vecteur d'expression T7 pET-3a. PT7, promoteur du gène *10* du phage T7; T7ter, terminateur de transcription; RBS, site de liaison des ribosomes; *ori* pMB1, réplicon pMB1 provenant du plasmide pBR322; *rop*, gène codant pour la protéine ROP impliquée dans la régulation du nombre de copies du plasmide.

### 1.3.4 Expression basale

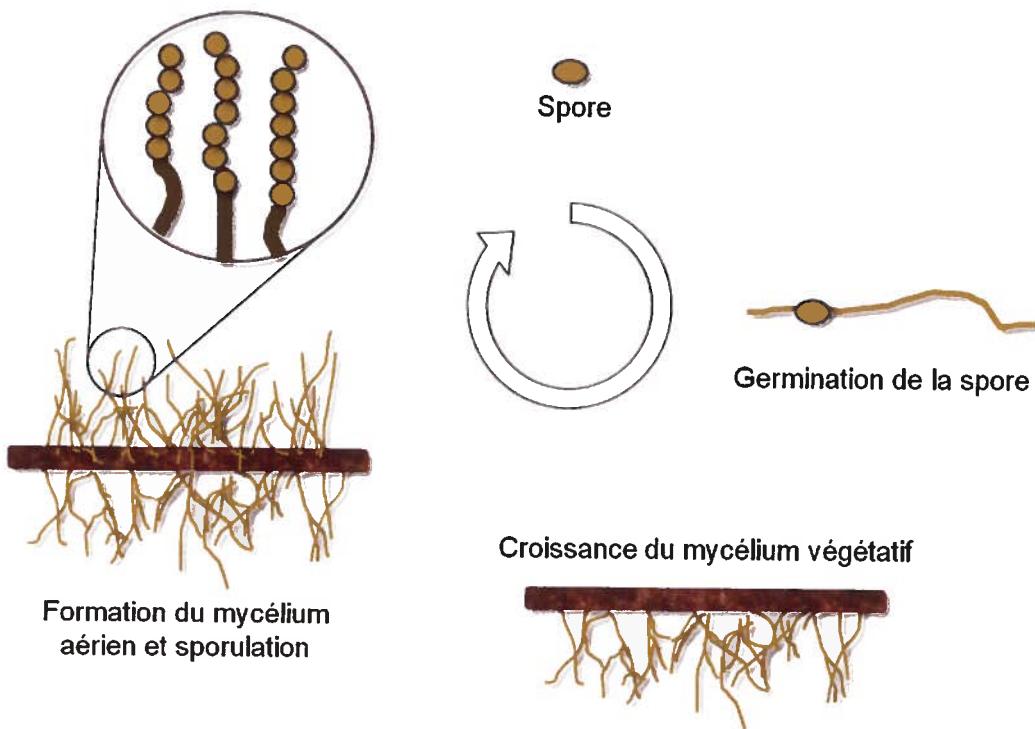
Les souches d'expression T7 présentent un niveau basal de production d'ARNpol T7 qui se traduit par une légère expression du gène d'intérêt en absence d'inducteur. Si la protéine produite est suffisamment毒ique pour la cellule, elle

peut nuire à la croissance et rendre le vecteur d'expression instable (Studier *et al.* 1990). Une des approches pour réduire cette fuite d'expression a été de fusionner l'opérateur *lac* au promoteur T7 dans le vecteur (Studier *et al.* 1990; Dubendorff et Studier 1991). Ainsi, en absence d'inducteur, même si une petite quantité d'ARNpol T7 est produite, l'opérateur *lac* permet de réduire la transcription du gène cloné grâce au répresseur LacI. Une autre approche pour réduire la fuite d'expression chez le système T7 d'*E. coli*, est d'utiliser une souche qui produit une petite quantité de lysozyme T7 à partir d'un plasmide compatible avec le vecteur d'expression. Le lysozyme T7 est un inhibiteur naturel de l'ARNpol T7 et réduit ainsi l'expression du gène d'intérêt en absence d'inducteur (Chamberlin et Ring 1973). Lorsque l'inducteur est ajouté, la quantité d'ARNpol T7 active est suffisamment grande pour permettre la transcription adéquate du gène d'intérêt (Studier 1991).

#### 1.4 *Streptomyces lividans*

Parmi la classe *Actinobacteria*, le genre *Streptomyces* comporte plus de 600 espèces et sous-espèces répertoriées (German Collection of Microorganisms and Cell Cultures; <http://www.dsmz.de/index.htm>). Ces bactéries Gram-positif sont strictement aérobies et possèdent un génome riche en G+C (69-78 %) (Glauert et Hopwood 1961; Korn-Wendisch et Kutzner 1992; Wright et Bibb 1992). De type saprophyte, ces bactéries sont présentes naturellement dans le sol et assurent leur croissance à partir de la dégradation des matières organiques grâce à la production de nombreuses enzymes hydrolytiques extracellulaires (Hodgson 2000). Les streptomycètes se caractérisent par leur forme filamenteuse, se développant sous forme d'hyphes coenocytiques et présentant un cycle de vie complexe (Figure 3) (Claessen *et al.* 2006). Les streptomycètes sont parmi les bactéries possédant les plus gros génomes, leur chromosome linéaire variant entre 8,5 et 12 Mb (Ikeda *et al.* 2003; Borodina *et al.* 2005; Ohnishi *et al.* 2008; Wang *et al.* 2010b). Une importante proportion des gènes est consacrée à la production de métabolites secondaires, dont plusieurs possèdent des propriétés antibiotiques (Hopwood

1999). De ce fait, les streptomycètes sont l'une des principales sources d'antibiotiques, tels le chloramphénicol, la néomycine, la streptomycine et la tétracycline (Distler *et al.* 1987; Dairi *et al.* 1995; Fan *et al.* 2008; Podzelinska *et al.* 2010).



**Figure 3.** Cycle de vie des streptomycètes. Dans les conditions propices, la spore germe pour former un mycélium végétatif qui colonise le substrat. Par la suite, un mycélium aérien émerge du substrat pour éventuellement se segmenter et former des spores.

#### 1.4.1 Souches

Parmi le genre *Streptomyces*, *S. coelicolor* est l'espèce la plus étudiée et la mieux caractérisée génétiquement (Hopwood 1999; Bentley *et al.* 2002). *S. coelicolor* est

en effet étudiée depuis plus de 50 ans et est devenue le paradigme des actinomycètes. Par contre, pour la production de protéines recombinantes, *S. lividans* est pratiquement la seule espèce utilisée. Ceci s'explique en partie par le fait que *S. coelicolor*, comme beaucoup d'autres *Streptomyces*, possède un système de restriction très actif qui limite l'utilisation de vecteurs dérivés d'*E. coli* (Kieser *et al.* 2000). *S. lividans* est connue pour sa capacité à sécréter de grandes quantités d'enzymes (Gilbert *et al.* 1995). *S. lividans* permet de produire les protéines d'intérêt directement dans le milieu de culture, ce qui peut être avantageux pour un grand nombre d'applications. En effet, ceci permet d'obtenir la protéine d'intérêt sous sa forme mature et active. La sécrétion des protéines facilite aussi grandement la purification de la protéine d'intérêt en l'isolant des nombreuses protéines cytoplasmiques. La protéine produite étant directement accessible dans le milieu de culture, il est aisément d'en évaluer l'activité. De plus, *S. lividans* est reconnue pour son activité protéolytique extracellulaire réduite, limitant la dégradation des protéines recombinantes (Binnie *et al.* 1997; Kieser *et al.* 2000). Ces dernières caractéristiques font de *S. lividans* un hôte alternatif particulièrement intéressant pour la production de protéines homologues et hétérologues (Tableau II).

Depuis que *S. lividans* est utilisée comme hôte d'expression, peu de modifications génétiques ont été apportées pour tenter d'améliorer cet organisme. La souche sauvage, ainsi que des dérivés comportant des marqueurs de sélection, sont le plus souvent utilisés pour la production de protéines. La souche TK24, l'une des plus populaires, ne contient pas les plasmides SLP2 et SLP3 présents chez la souche sauvage, et possède un marqueur de sélection lui conférant une résistance à la streptomycine (Kieser *et al.* 2000). TK64, un dérivé auxotrophe de TK24, est aussi utilisé pour la production de protéines. La souche 10-164, bien que moins répandue, a été largement utilisée pour la production de différentes enzymes impliquées dans la dégradation du xylane et de la cellulose (Dupont *et al.* 1996; Sulzenbacher *et al.* 1997; Vincent *et al.* 1997; Nisole *et al.* 2006). La souche 10-164 possède une mutation ponctuelle au niveau du gène *msiK* lui conférant un phénotype xylanase/cellulase négatif (Hurtubise *et al.* 1995).

**Tableau II.** Protéines produites chez *S. lividans*.

Protéine	Source	Productivité (mg/l)	Référence(s)
Interleukine-6	<i>Homo sapiens</i>	0,61	(Zhu <i>et al.</i> 2011)
L-Asparaginase	<i>Streptomyces thermoluteus</i>	0,88	(Hatanaka <i>et al.</i> 2010)
Transglutaminase	<i>Streptoverticillium cinnamoneum</i>	0.23	(Noda <i>et al.</i> 2010)
Lipases	Banque métagénomique	-----	(Meilleur <i>et al.</i> 2009; Côté et Shareck 2010)
Laccase	<i>Streptomyces coelicolor</i>	350	(Dubé <i>et al.</i> 2008)
Streptokinase	<i>Streptococcus equisimilis</i>	15	(Pimienta <i>et al.</i> 2007)
$\alpha$ Intégrine (domaine A)	<i>Rattus norvegicus</i>	8	(Ayadi <i>et al.</i> 2007)
Protéine APA	<i>Mycobacterium tuberculosis</i>	80	(Vallin <i>et al.</i> 2006)
Acétyle xylane estérase	<i>Streptomyces lividans</i>	388	(Nisole <i>et al.</i> 2006)
Xyloglucanase	<i>Jonesia</i> sp.	100-150	(Sianidis <i>et al.</i> 2006)

### 1.4.2 Plasmides

La grande majorité des vecteurs d'expression utilisés chez *S. lividans* sont basés sur des réplicons utilisant le mode de réPLICATION par cercle roulant (CR) (Khan 2005) (Tableau III). Un des plus utilisés est celui provenant du plasmide circulaire multi-copies pIJ101, initialement isolé chez *S. lividans* ISP5434 (Kieser *et al.* 1982). Les fonctions nécessaires à la réPLICATION de pIJ101 sont incluses dans un segment de 2,1 kb comprenant le gène codant pour la protéine Rep et l'origine de réPLICATION double-brin (*ori* DB) (Kendall et Cohen 1988). La protéine Rep est requise pour l'initiation de la réPLICATION, elle se fixe à l'*ori* DB et effectue une coupure d'un des brins d'ADN (Khan 2005). Les plasmides se répliquent par le mécanisme CR se caractérisent par la formation d'intermédiaires simple-brin. Le plasmide est ensuite reconstitué grâce à la présence d'une origine de réPLICATION simple-brin (*ori* SB) (del Solar *et al.* 1998). Bon nombre de vecteurs d'expression ne possèdent pas d'*ori* SB à proprement dit, mais sont tout de même maintenus dans l'hôte. Cette fonction serait remplie plus ou moins efficacement par des séquences alternatives présentes dans le vecteur (Zaman *et al.* 1993). Un autre réplicon de type CR est celui du plasmide pJV1 isolé chez *S. phaeochromogenes* (Bailey *et al.* 1986). Les réplicons pJV1 et pIJ101 sont très similaires mais demeurent tout de même compatibles (Servin-Gonzalez 1993). Les réplicons pSG5 et pSAM2 utilisent aussi le mécanisme CR, mais donnent des plasmides avec un nombre de copies un peu moins élevé (Muth *et al.* 1988; Smokvina *et al.* 1991). SCP2\* est le seul plasmide se répliquant par le mécanisme thêta (Haug *et al.* 2003). Il permet d'obtenir un nombre de copies par chromosomes variant entre 1 et 5, idéal pour les études fonctionnelles. Bien que moins utilisés que les vecteurs réPLICATIFS, les vecteurs intégratifs permettent de maintenir et d'exprimer un gène d'intérêt de façon très stable. Ces vecteurs s'intègrent dans le chromosome de la cellule hôte par recombinaison site-spécifique, grâce à une intégrase fournie par le vecteur (Van Mellaert *et al.* 1998; Thorpe *et al.* 2000; Raynal *et al.* 2002). Basé sur l'intégrase du phage tempéré phiC31, le vecteur pSET152 est le plus couramment utilisé (Bierman *et al.* 1992).

**Tableau III.** Réplicons utilisés dans les vecteurs d'expression de *S. lividans*.

Réplicon	Mécanisme de réPLICATION	Nombre de copies (par chromosome)	Source
pIJ101	Cercle roulant	40-300	<i>S. lividans</i>
pJV1	Cercle roulant	150	<i>S. phaeochromogenes</i>
pSG5	Cercle roulant	20-50	<i>S. ghanaensis</i>
pSAM2	Cercle roulant	10	<i>S. ambofaciens</i>
SCP2*	Thêta	1-5	<i>S. coelicolor</i>

#### 1.4.3 Promoteurs

Certains promoteurs de streptomycètes sont semblables à ceux reconnus par le facteur  $\sigma$ 70 d'*E. coli*. Cependant, la majorité ne présente pas les régions typiques -10 et -35 et ne sont pas fonctionnels chez *E. coli* (Strohl et Connors 1992). La plupart des vecteurs d'expression utilisés chez *S. lividans* sont basés sur des promoteurs constitutifs forts. Une grande diversité de promoteurs sont utilisés et varient souvent d'un groupe de recherche à un autre. Parmi les plus connus, on retrouve celui provenant de l'opéron mélanine (*Pmel*) de *Streptomyces antibioticus* (Katz *et al.* 1983) et celui du gène de résistance à l'érythromycine (*PermE*) de *Saccharopolyspora erythraea* (Bibb *et al.* 1985). Bien que l'utilisation de promoteurs constitutifs soit souvent efficace et simple d'utilisation, cette approche peut s'avérer problématique lorsque le produit d'un gène affecte la croissance ou est toxique. Dans ces cas, il peut être préférable d'utiliser un promoteur inducible, permettant d'exprimer le gène d'intérêt au moment voulu. L'un des promoteurs inducibles les plus utilisés est celui du gène codant pour le régulateur transcription TipA de *S. lividans* (Murakami *et al.* 1989). Ce promoteur est induit par le thiostrepton, un antibiotique couramment utilisé chez *S. lividans* (Thompson *et al.* 1980). Son utilisation requiert toutefois que l'hôte d'expression possède le gène *tipA*, car son produit est requis pour l'auto-induction du promoteur (Chiu *et al.* 1999). Plus récemment, le promoteur de la nitrilase A

(*PnitA*) de *Rhodococcus rhodochrous* a été utilisé pour l'expression régulée de gènes recombinants chez diverses *Streptomyces*, dont *S. lividans* (Herai *et al.* 2004). *PnitA* est induit par le  $\epsilon$ -caprolactame par l'entremise du régulateur transcriptionnel positif NitR (Komeda *et al.* 1996).

#### 1.4.4 Traduction

La séquence Shine-Dalgarno (SD) consensus de *S. lividans* est la même que celle d'*E. coli*, soit GGAGG (Shine et Dalgarno 1974; Anne et Van Mellaert 1993; Chen *et al.* 1994). La distance entre la séquence SD et le codon d'initiation varie entre 5 et 12 nucléotides chez *S. lividans* et généralement entre 5 et 9 nucléotides chez *E. coli* (Anne et Van Mellaert 1993; Chen *et al.* 1994). De par son génome riche en G+C, le codon d'initiation GTG est plus souvent utilisé qu'ATG et le codon stop TGA est le plus répandu. Comme bon nombre d'organismes, *S. lividans* présente un biais dans l'usage des codons. En effet, les codons comportant plus de deux A ou C sont très rares (Codon Usage Database, <http://www.kazusa.or.jp/codon/>). Chez les *Streptomyces*, incluant *S. lividans*, le codon TTA<sub>leu</sub> est le moins utilisé et est uniquement présent dans les gènes impliqués dans le métabolisme secondaire (Chater et Chandra 2008). Le codon TTA<sub>leu</sub> n'est efficacement transcrit que par un seul ARNt codé par *bldA* (Lawlor *et al.* 1987; Ueda *et al.* 1993). La forme mature de l'ARNt<sub>leu</sub> ne serait présente qu'à la fin de la phase de croissance logarithmique, lorsque le métabolisme secondaire est activé (Leskiw *et al.* 1993; Chater et Chandra 2008). Pour pouvoir être efficacement exprimés chez *S. lividans*, les gènes comportant un ou des codons TTA<sub>leu</sub> doivent donc être mutés.

#### 1.4.5 Sécrétion

Les protéines destinées à être sécrétées sont produites sous forme de précurseurs comportant un peptide signal à l'extrémité N-terminal (Anne et Van Mellaert 1993). Lors du processus de translocation au travers de la membrane

cytoplasmique, le peptide signal est clivé par une peptidase (Dalbey *et al.* 1997). Un peptide signal se compose d'une région N-terminal riche en acides aminés chargés, d'une région hydrophobique centrale et se termine par une région hydrophile comportant le site de clivage (von Heijne 1985; von Heijne 1986). Mis à part ces caractéristiques communes, les séquences des peptides signaux varient grandement (von Heijne 1984; Abrahmsen *et al.* 1985). La longueur moyenne des peptides signaux chez les *Streptomyces* est de 35 acides aminés (Gilbert *et al.* 1995; Morosoli *et al.* 1997), tandis qu'elle est de 24 chez *E. coli* (von Heijne et Abrahmsen 1989).

Chez les bactéries, la sécrétion des protéines à travers la membrane cytoplasmique s'effectue *via* le *general secretion pathway* (système Sec) (Wickner *et al.* 1991) et le *twin arginine translocation pathway* (système Tat) (Berks *et al.* 2000; Berks *et al.* 2005). Le système Sec est impliqué dans la sécrétion de protéines non-repliées à travers la membrane cytoplasmique et dans l'insertion de protéines membranaires (Osborne *et al.* 2005). Le système Tat est utilisé pour la sécrétion de protéines déjà repliées et est caractérisé par des peptides signaux contenant le motif consensus Arg-Arg-X-φ-φ ( $\phi$  = résidus hydrophobe) (Stanley *et al.* 2000; Natale *et al.* 2008). Dans les systèmes d'expression bactériens utilisés pour la production de protéines recombinantes sécrétées, le système Sec est presque exclusivement utilisé. Le système Tat demeure tout de même une alternative intéressante et de plus en plus de travaux sont réalisés sur le sujet (Schaerlaekens *et al.* 2004; Gauthier *et al.* 2005; De Keersmaeker *et al.* 2006; Vrancken *et al.* 2007; Guimond et Morosoli 2008).

## 1.5 *Bacillus subtilis*

*Bacillus subtilis* est une bactérie Gram-positif aérobie stricte formant des endospores. Comme *S. lividans*, *B. subtilis* a la capacité de sécréter une grande diversité d'enzymes hydrolytiques, lui permettant de coloniser une multitude d'environnements (Earl *et al.* 2008). Après *E. coli*, *B. subtilis* est l'hôte procaryote le plus couramment utilisé pour la production de protéines recombinantes

(Tableau IV) (Westers *et al.* 2004). L'intérêt pour *B. subtilis* comme hôte d'expression peut entre autres s'expliquer par sa capacité à sécréter de grandes quantités d'enzymes directement dans le milieu de culture (Li *et al.* 2004). Aussi, contrairement à *E. coli*, *B. subtilis* ne produit pas d'endotoxines et est considérée comme un organisme GRAS (*generally regarded as safe*) par la *Food and Drug Administration*. De plus, *B. subtilis* ne présente pas de biais notables dans l'utilisation des codons et est très bien caractérisée au niveau biochimique, physiologique et génétique (Harwood 1992; Wong 1995; Kobayashi *et al.* 2003; Schallmey *et al.* 2004; Westers *et al.* 2004).

**Tableau IV.** Protéines recombinantes produites chez *B. subtilis*.

Protéine	Source	Productivité (mg/l)	Référence
α-amylase	<i>Bacillus amyloliquefaciens</i>	1000-3000	(Palva 1982)
EGF	Humain	7	(Lam <i>et al.</i> 1998)
Interféron-α2	Humain	0,5-1,0	(Palva <i>et al.</i> 1983)
Lipase A	<i>B. subtilis</i>	600	(Lesuisse <i>et al.</i> 1993)
Pénicilline G acylase	<i>B. subtilis</i>	-----	(Yang <i>et al.</i> 2001)
Dépolymerase	<i>Paucimonas lemoignei</i>	1,9	(Braaz <i>et al.</i> 2002)
ScFv	Synthétique	10-15	(Wu <i>et al.</i> 2002)
Staphylokinase	<i>Staphylococcus aureus</i>	337	(Ye <i>et al.</i> 1999)
Streptavidine	<i>Streptomyces avidinii</i>	35-50	(Wu et Wong 2002)
Thiorédoxine	<i>Aliciclobacillus acidocaldarius</i>	500	(Digilio <i>et al.</i> 2003)
Proinsuline	Humain	1000	(Olmos-Soto et Contreras-Flores 2003)

### 1.5.1 Souches

L'un des principaux problèmes rencontrés chez *B. subtilis* lors de la production de protéines hétérologues, est la présence d'un grand nombre de protéases extracellulaires (Li *et al.* 2004). À ce jour, huit protéases extracellulaires ont été identifiées chez *B. subtilis*, soit une protéase alcaline (AprE) (Haugaard et Haugaard 1955), deux protéases neutres (NprE et NprB) (Yang *et al.* 1984; Tran *et al.* 1991), une métalloprotéase (Mpr) (Rufo *et al.* 1990), trois protéases à sérine (Epr, Bpr et Vpr) (Bruckner *et al.* 1990; Sloma *et al.* 1990; Sloma *et al.* 1991) et une protéase associée à la paroi (WprA) (Margot et Karamata 1996). Bien que les protéines natives de *B. subtilis* résistent généralement à l'action de ces protéases, les protéines hétérologues sont souvent rapidement hydrolysées en leur présence (Li *et al.* 2004). Pour contourner ce problème, différentes souches avec des activités protéolytiques extracellulaires réduites ont été développées (Tableau V).

**Tableau V.** Souches de *B. subtilis* possédant une activité protéolytique réduite.

Souche	Protéases inactivées	Référence
WB600	<i>nprE nprB aprE epr mpr bpr</i>	(Wu <i>et al.</i> 1991)
LB700	<i>nprE nprB aprE epr mpr bpr wprA</i>	(Lee <i>et al.</i> 2000)
WB800	<i>nprE nprB aprE epr mpr bpr vpr wprA</i>	(Wu <i>et al.</i> 2002)
Dpr8	<i>nprE nprB aprE epr mpr bpr vpr wprA</i>	(Kodama <i>et al.</i> 2007a)

### 1.5.2 Vecteurs d'expression

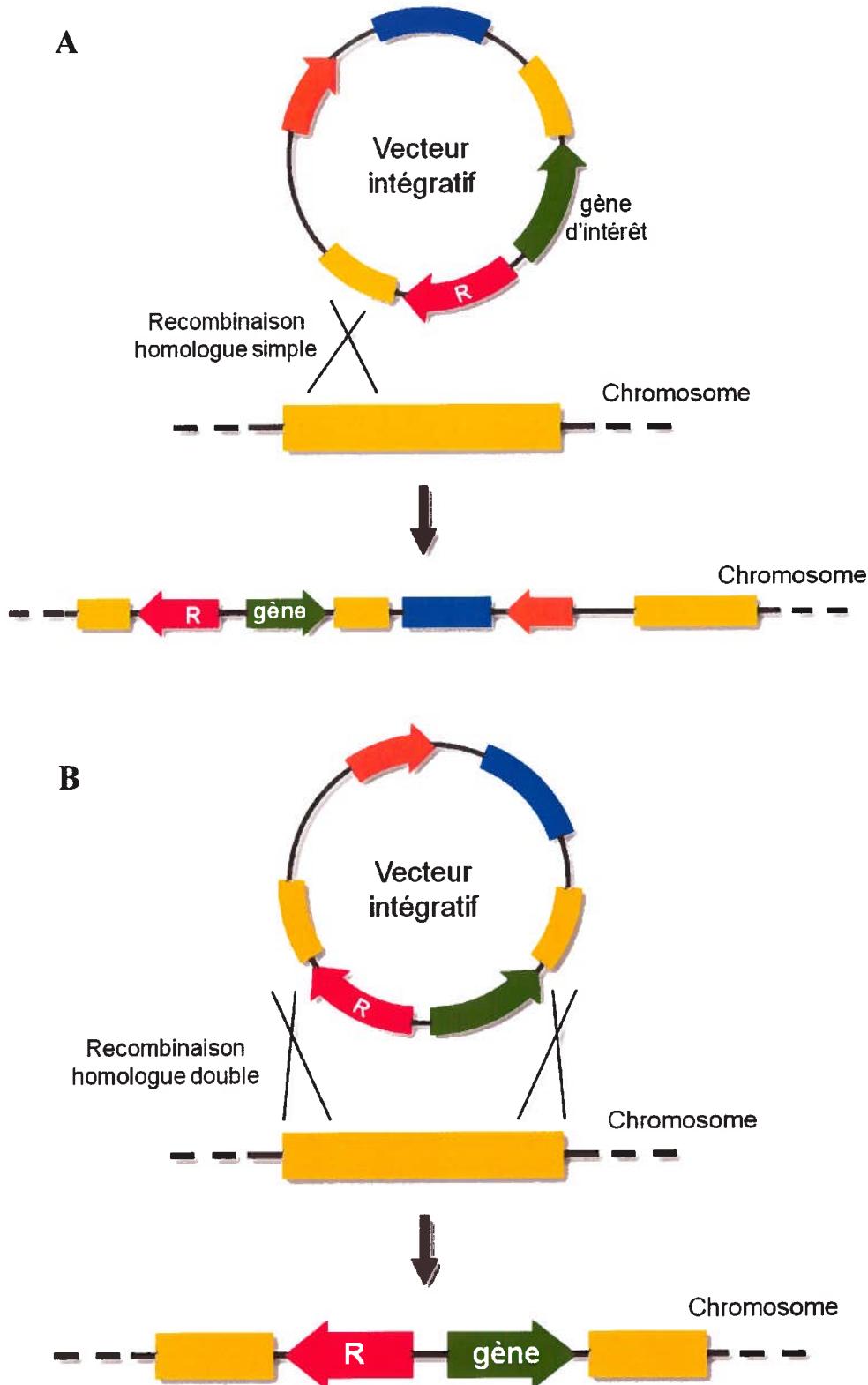
Pour la production de protéines recombinantes chez *B. subtilis*, les principaux systèmes d'expression sont basés sur deux types de vecteurs, soit réplicatifs et

intégratifs. De par leur mode de réPLICATION, les vecteurs plasmidiques peuvent être classés en deux groupes, ceux utilisant le mécanisme de cercle roulant (CR) et ceux utilisant le mécanisme thêta (Tableau VI). Les premiers vecteurs développés étaient basés sur des réplicons isolés de plasmides de types CR provenant de *Staphylococcus aureus*, tels pUB110, pC194, pE194 et pT181 (Gryczan *et al.* 1982; Horinouchi et Weisblum 1982; Kumar et Novick 1985; Alonso 1989). Ces vecteurs, bien que stables en eux-mêmes, peuvent devenir structurellement instables lorsqu'un gène y est introduit (Schumann 2007). L'instabilité structurelle est principalement causée par la recombinaison entre de courtes répétitions directes, favorisée par la formation d'ADN intermédiaire simple-brin durant le processus de réPLICATION (Meijer *et al.* 1995; Lee *et al.* 1998). Ce problème de stabilité fut pendant plusieurs années un important inconvénient à l'utilisation de *B. subtilis* pour la production de protéines recombinantes (Schumann 2007). L'autre groupe, les vecteurs utilisant le mécanisme thêta, ne présente pas cette instabilité structurelle (Lee *et al.* 1998). L'un des réplicons les plus utilisés dérive du plasmide pBS72, isolé chez *B. subtilis* (Titok *et al.* 2003). Le segment d'ADN d'environ 2 kb nécessaire à la réPLICATION comprend un gène codant pour une protéine REP et l'origine de réPLICATION (Titok *et al.* 2006).

Les vecteurs intégratifs permettent l'insertion stable de gènes à l'intérieur du chromosome. Une fois intégré, le gène d'intérêt peut être maintenu même en l'absence de sélection (Vazquez-Cruz *et al.* 1996). Les vecteurs intégratifs comportent généralement un réplicon d'*E. coli* et un gène de résistance à un antibiotique permettant la sélection chez *B. subtilis*. Pour permettre l'intégration chromosomique, les vecteurs contiennent généralement deux séquences homologues à *B. subtilis*, entre lesquelles le marqueur de sélection et le gène d'intérêt sont placés. Une fois transféré chez *B. subtilis*, le vecteur entier peut s'intégrer au chromosome par simple recombinaison homologue (Figure 4A). À une fréquence moins élevée, seul le gène d'intérêt et le marqueur de sélection sont intégrés par double recombinaison homologue (Figure 4B). Ce dernier type de recombinaison est préférable, car plus stable (Schumann 2007).

**Tableau VI.** Réplicons utilisés chez *B. subtilis*.

Réplicon	Mécanisme de réPLICATION	Source	Référence
pUB110	Cercle roulant	<i>Staphylococcus aureus</i>	(Lacey et Chopra 1974)
pC194	Cercle roulant	<i>Staphylococcus aureus</i>	(Iordanescu <i>et al.</i> 1978)
pE194	Cercle roulant	<i>Staphylococcus aureus</i>	(Iordanescu 1976)
pT181	Cercle roulant	<i>Staphylococcus aureus</i>	(Iordanescu 1976)
pTA1060	Cercle roulant	<i>Bacillus subtilis</i>	(Uozumi <i>et al.</i> 1980)
pBS72	Thêta	<i>Bacillus subtilis</i>	(Titok <i>et al.</i> 2003)
pAMβ1	Thêta	<i>Enterococcus faecalis</i>	(Clewel <i>et al.</i> 1974)
pTB19	Thêta	<i>Bacillus stearothermophilus</i>	(Imanaka <i>et al.</i> 1981)



**Figure 4.** Intégration chromosomique par simple (A) et double (B) recombinaison homologue (R, marqueur de sélection).

### 1.5.3 Promoteurs

Comme pour *E. coli*, l'expression de gènes recombinants chez *B. subtilis* s'effectue principalement par l'entremise de promoteurs régulés. La grande majorité des systèmes d'expression sont basés sur des promoteurs inductibles par le xylose ou l'IPTG. Le premier promoteur inductible développé chez *B. subtilis* est une fusion du promoteur du phage SPO-1 avec l'opérateur *lac* d'*E. coli* (Yansura et Henner 1984). Ce promoteur, dénommé *spac*, est induit par l'ajout d'IPTG, qui se lie au répresseur LacI et l'empêche de se fixer à l'opérateur. Le répresseur n'étant pas naturellement produit par *B. subtilis*, le gène *lacI* d'*E. coli* doit être inclus dans le vecteur d'expression ou porté par un plasmide compatible. Le promoteur *spac* est relativement faible et comme le promoteur *lac* d'*E. coli*, n'est pas complètement inhibé en absence d'inducteur (Schumann 2007). Cependant, de nombreux vecteurs d'expression basés sur l'opérateur *lac* ont depuis été développés (Vagner *et al.* 1998; Hartl *et al.* 2001; Nguyen *et al.* 2005; Phan *et al.* 2006). Un autre promoteur couramment utilisé est celui de l'opéron xylose (*xyl*), provenant de *B. subtilis* ou de *B. megaterium* (Rygus *et al.* 1991; Dahl *et al.* 1994). Le promoteur *xyl* est sous le contrôle négatif de XylR et est induit par le xylose (Gartner *et al.* 1988). Il a été utilisé pour la première fois dans un vecteur d'expression par Gärtner *et al.* (1992) et a par la suite été utilisé par plusieurs autres groupes. Le promoteur *xyl* est très bien régulé et l'addition de 0,1 à 2 % de xylose permet d'obtenir un haut taux d'induction (Kim *et al.* 1996). Plusieurs autres promoteurs ont été décrits dans la littérature, induits par la tétracycline (Geissendorfer et Hillen 1990), la carence en phosphate (Lee *et al.* 1991), l'atteinte de la phase de croissance stationnaire (Jan *et al.* 2001), le citrate (Fukushima *et al.* 2002), l'amidon (Ho et Lim 2003), la glycine (Phan et Schumann 2007), la réduction de la température (Thuy Le et Schumann 2007) et la subtiline (Bongers *et al.* 2005).

### 1.5.4 RBS

Les différentes caractéristiques des sites de liaison des ribosomes (RBS) d'*E. coli* ont été largement analysées (Shepard *et al.* 1982; Gren 1984; Munson *et al.* 1984; de Smit et van Duin 1990), mais sont par contre moins bien connues chez *B. subtilis*. Au moins quatre éléments sont reconnus pour influencer l'efficacité de traduction, soit la séquence de Shine-Dalgarno (SD), le codon d'initiation, la distance entre ces deux éléments et la présence de structures secondaires dans le RBS (Vellanoweth et Rabinowitz 1992; Schumann 2007). En général, les séquences SD de *B. subtilis* ont une variation d'enthalpie libre ( $\Delta G$ ) d'interaction avec l'ARNr 16s plus grande que chez *E. coli* (Hager et Rabinowitz 1985b). Pour être efficacement traduits par *B. subtilis*, les ARNm doivent en effet posséder des séquences SD avec de grandes  $\Delta G$  d'interaction ( $> 50,4 \text{ kJ/mol}$ ) (McLaughlin *et al.* 1981; Band et Henner 1984; Hager et Rabinowitz 1985a). Chez *B. subtilis*, la fréquence d'utilisation des codons d'initiation ATG, TTG et GTG est respectivement de 78%, 13% et 9% (Jan *et al.* 2001). Un certains nombre d'études ont montré que le meilleur taux de traduction était obtenu avec le codon ATG (Ambulos *et al.* 1990; Vellanoweth et Rabinowitz 1992; Yeh *et al.* 1997). Dans une autre étude par contre, aucune augmentation significative de la production d'AprE n'a pu être mesurée en remplaçant le codon GTG par ATG (Jan *et al.* 2001). La distance optimale entre la séquence SD et le codon d'initiation est similaire chez *B. subtilis* et *E. coli*, soit entre 7 et 9 nucléotides. *B. subtilis* serait par contre plus affectée qu'*E. coli* par un espacement sous-optimal (Vellanoweth et Rabinowitz 1992). De la même façon, *B. subtilis* tolère moins bien qu'*E. coli* la présence de structures secondaires à l'intérieur du RBS (Vellanoweth et Rabinowitz 1992). Contrairement à *E. coli*, plusieurs ARNm de *B. subtilis* sont stabilisés par la liaison ou le *stalling* de ribosomes à leur extrémité 5' (Vellanoweth et Rabinowitz 1992; Hu et Switzer 1995; Jurgen *et al.* 1998; Hambraeus *et al.* 2002; Sharp et Bechhofer 2003).

### **1.5.5 Sécrétion**

Chez *B. subtilis*, le transit de la pré-protéine vers la membrane, la translocation et le clivage subséquent du peptide signal sont les principales étapes limitantes de la sécrétion de protéines recombinantes (Li *et al.* 2004). Le peptide signal joue un rôle crucial dans le processus de translocation en interagissant avec la machinerie de sécrétion et la signal peptidase (Tjalsma *et al.* 2000; Harwood et Cranenburgh 2008). Des tentatives d'augmentation de la sécrétion de protéines recombinantes par l'utilisation de peptides signaux optimaux se sont toutefois avérées infructueuses, reflétant le manque de connaissance dans le domaine (Brockmeier *et al.* 2006).

Dans le processus de sécrétion, les chaperonnes intracellulaires, telles que GroEL, DnaK, DnaJ et GprE, aident au repliement, minimisent l'agrégation et maintiennent les protéines nouvellement formées dans une conformation propice à la translocation (Yuan et Wong 1995). La chaperonne extra-cytoplasmique PrsA intervient suite à la translocation en aidant au repliement final des protéines matures (Kontinen *et al.* 1991; Leskela *et al.* 1999). En augmentant le niveau cellulaire de ces chaperonnes, il a été démontré qu'il était possible d'améliorer la sécrétion de protéines recombinantes (Wu *et al.* 1998).

### **1.5.6 Système d'expression T7 de *B. subtilis***

Le système d'expression T7 développé chez *E. coli* a été adapté à *B. subtilis* par Conrad *et al.* (1996). Bien que l'idée de combiner l'ARNpol T7 à *B. subtilis* semblait prometteuse, ce système d'expression ne semble pas avoir été réutilisé dans aucune autre publication. La souche d'expression, capable de produire l'ARNpol T7, a été construite à partir de *B. subtilis* BD170 (*trpC2 thr-5*) (Conrad *et al.* 1992). Le gène codant pour l'ARNpol T7 a été placé sous le contrôle du promoteur *xyl* et intégré au chromosome de *B. subtilis* par l'entremise du vecteur pDH32 (Henner 1990). L'intégration chromosomique a été réalisée par double recombinaison homologue entre le vecteur et le locus *amyE*. La souche ainsi

obtenue permettait de produire l'ARNpol T7 de façon inducible. En absence d'inducteur, le niveau basal d'ARNpol T7 était pratiquement nul. Tout comme le système T7 d'*E. coli*, le vecteur d'expression construit par Conrad *et al.* (1996) permettait le clonage de gènes en aval d'un promoteur T7. Le vecteur d'expression pBET7 a été construit en intégrant le promoteur du gène *10* ( $\Phi 10$ ) et son terminateur ( $T\Phi$ ) au plasmide pPL703 (Ambulos *et al.* 1985). Le vecteur pBET7 comportait un site de clonage multiple entre le promoteur  $\Phi 10$  et le terminateur  $T\Phi$ , mais aucune séquence de Shine-Dalgarno.

Le système d'expression a été évalué pour la production de protéines cytoplasmiques et extracellulaires. Pour la production cytoplasmique, les gènes codant pour la  $\beta$ -galactosidase d'*E. coli* et la  $\beta$ -glucosidase de *Thermoanaerobacter brockii* ont été clonés dans le vecteur pBET7. Les résultats obtenus ont démontré que le système d'expression T7 était fonctionnel chez *B. subtilis*, bien que la quantité de protéines produites était relativement faible. La productivité du système pouvait cependant être accrue par l'ajout de rifampicine suite à l'induction. Cet effet a aussi été observé chez le système d'expression T7 d'*E. coli* (Kuderova *et al.* 1999). La rifampicine inhibe l'ARN polymérase de la cellule hôte, mais laisse active l'ARNpol T7 (Chamberlin et Ring 1973). Au moment d'ajouter la rifampicine, il y aurait suffisamment de protéines actives dans la cellule hôte pour lui permettre de survivre pendant une certaine période de temps. Toutes les ressources de la cellule seraient alors dédiées à la traduction des ARNm produits par l'ARNpol T7 (Studier *et al.* 1990; Kuderova *et al.* 1999). La production de protéines extracellulaires a été évaluée avec l' $\alpha$ -amylase de *Thermoactinomyces vulgaris*, une protéine reconnue pour être efficacement sécrétée chez *B. subtilis* (Hofemeister *et al.* 1994). L'enzyme s'accumulait dans le milieu de culture, pour atteindre 70 mg/l entre 10-20 heures après l'induction. Dans ce cas-ci, l'ajout de rifampicine réduisait de façon considérable la sécrétion de l' $\alpha$ -amylase.

## **2. PRÉSENTATION DE L'ARTICLE 1**

## **2.1. Mise en contexte de l'article 1**

La production de protéines recombinantes est un aspect essentiel de la recherche dans plusieurs laboratoires, entre autres pour la découverte de nouvelles enzymes, leur purification et leur caractérisation biochimique. Pour augmenter les chances de succès il est bon d'avoir accès à une variété de systèmes d'expression efficaces et simples d'utilisation. Le présent article décrit la construction d'un nouveau système d'expression chez *Streptomyces lividans*, basé sur l'ARN polymérase hautement active du phage T7. L'article a été accepté pour publication dans la revue *Applied and Environmental Microbiology* le 6 décembre 2009.

## **2.2. Contribution de l'étudiant à l'article 1**

Les travaux présentés dans cet article ont été entièrement réalisés par l'étudiant sous la supervision de François Denis (Ph.D.) et François Shareck (Ph.D.). L'article a été écrit par l'étudiant et révisé par les co-auteurs.

### **2.3. Article 1: Adaptation of the highly productive T7 expression system to *Streptomyces lividans***

François-Xavier Lussier\*, François Denis and François Shareck

INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval,  
Québec, Canada H7V 1B7

\* Corresponding author. Mailing address: INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval, Québec, Canada H7V 1B7. Phone: (450) 687-5010, ext. 4610. Fax: ( 686-5501. E-mail: [francois-xavier.lussier@iaf.inrs.ca](mailto:francois-xavier.lussier@iaf.inrs.ca)

**Référence:** Lussier, F.X., Denis, F. et Shareck, F. (2010). "Adaptation of the highly productive T7 expression system to *Streptomyces lividans*." *Appl Environ Microbiol* **76**(3): 967–970.

## 2.4 Résumé

La bactérie Gram-négatif *Escherichia coli* est l'organisme le plus couramment utilisé pour l'expression de gènes hétérologues (Terpe 2006). Un grand nombre de souches et de vecteurs sont disponibles commercialement, adaptés pour différents usages. Parmi les systèmes d'expression disponibles chez *E. coli*, le système basé sur l'ARN polymérase du phage T7 (ARNpol T7) est l'un des plus populaires et permet d'obtenir un très haut niveau d'expression (Studier et Moffatt 1986; Schumann et Ferreira 2004). Ce système est aussi utilisé avec succès chez plusieurs autres organismes (Fuerst *et al.* 1986; Barnard *et al.* 2004; Kang *et al.* 2007; Gamer *et al.* 2009; Katzke *et al.* 2010).

Dans les systèmes d'expression utilisant *E. coli* comme hôte, les protéines s'accumulent généralement dans le cytoplasme ou le périplasme (Schmidt 2004). Pour avoir accès aux protéines, les cellules doivent donc être lysées. *Streptomyces lividans* est un hôte d'expression Gram-positif reconnu pour sa grande capacité de sécrétion. Cet hôte permet de produire les protéines d'intérêt directement dans le milieu de culture, ce qui est un avantage notable d'un point de vue biotechnologique. En effet, ceci simplifie la purification de la protéine d'intérêt en la séparant de tous les constituants cellulaires. De plus, les conditions extracellulaires sont plus propices au repliement protéique que celles du cytoplasme, diminuant ainsi la formation de corps d'inclusion (Moks *et al.* 1987).

L'objectif était de combiner les qualités de l'ARNpol T7 à celles de *Streptomyces lividans*. Pour ce faire, le gène codant pour l'ARNpol T7 a été intégré de façon stable dans le génome de *S. lividans*, sous le contrôle d'un promoteur inductible. Pour être efficacement exprimé par *S. lividans*, le gène de l'ARNpol T7 a été modifié pour se conformer à l'usage de codons de l'hôte. Un nouveau vecteur d'expression navette a aussi été développé, pouvant être utilisé chez *S. lividans* et *E. coli*.

## 2.5 Abstract

*Streptomyces lividans* is a Gram-positive bacterium known for its remarkable secretion efficiency and low extracellular protease activity. In the present work, we adapted the highly productive T7 expression system to *S. lividans*. A codon-optimized T7 RNA polymerase gene was chromosomally integrated and a bifunctional T7 expression vector was constructed.

## 2.6 Introduction

The *Escherichia coli* T7 RNA polymerase-based expression system, developed by Studier and Moffatt (Studier et Moffatt 1986), is currently used in many laboratories for heterologous protein production. The system is based on the T7 bacteriophage RNA polymerase (T7 RNAPol), which directs selective transcription of genes cloned downstream of the major T7 late promoter. T7 RNAPol is characterized by a very high activity, elongating chains about five times faster than *E. coli* RNA polymerase, and can generate very long mRNAs (Golomb et Chamberlin 1974; Studier et Moffatt 1986). Although *E. coli* has proved to be useful for gene overexpression, different problems can occur and limit the productivity (Terpe 2006). As such, the use of an alternative expression host is often desirable to obtain adequate protein production.

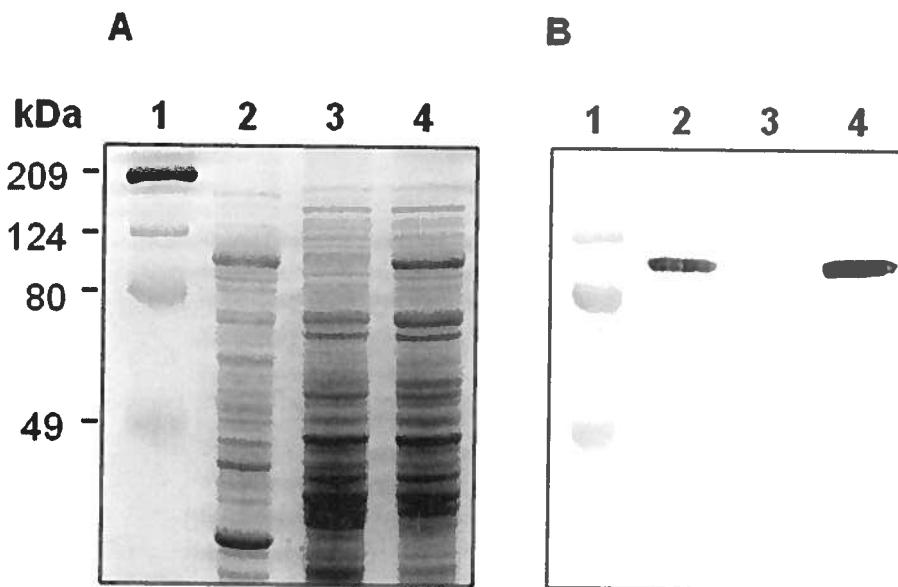
Streptomycetes are Gram-positive G+C-rich bacteria known for their high secretion capacity and have been used extensively in commercial settings for antibiotic production in very large-scale fermentation systems (Chater 2006). Among the streptomycetes, the readily transformable *Streptomyces lividans* has been used for the expression of a wide variety of genes from diverse sources (Binnie *et al.* 1997). This host can secrete directly in the culture medium large quantities of proteins in mature conformation, and given that few endogenous proteins are present in the media, downstream purification processes are simplified. *S. lividans* also displays a very low level of endogenous extracellular protease activity, making it a suitable host for heterologous protein production

(Kieser *et al.* 2000). Most of the *S. lividans* expression systems are based on strong constitutive promoters (Schmitt-John et Engels 1992; Binnie *et al.* 1997). Few inducible promoters are also used, the thiostrepton-inducible *tipA* promoter (*tipAp*) being the most popular (Takano *et al.* 1995).

Since the original publication of the *E. coli* T7 expression system, it has been adapted to mammalian cells and several bacteria (Elroy-Stein et Moss 1990; Conrad *et al.* 1996; Barnard *et al.* 2004; Drepper *et al.* 2005; Kang *et al.* 2007; Gamer *et al.* 2009; Katzke *et al.* 2009). In this report, the T7 expression system was adapted to *S. lividans* to combine the T7 RNAPol efficiency to the great features of this host. During the revision process of this paper, we learned that a similar system had been developed and published in a thesis at the University of Stuttgart (Fischer 1996).

## 2.7 T7 RNA polymerase production in *S. lividans*

To efficiently express the T7 RNAPol gene in *S. lividans*, the four rare TTA<sub>leu</sub> codons were replaced by CTC<sub>leu</sub> codons by overlap extension PCR (Higuchi *et al.* 1988) (see the supplemental material). The production of the T7 RNAPol was evaluated with the codon-modified (CM) and the wild-type (W) genes. They were cloned into the pIJ702-derived (Katz *et al.* 1983) multicopy expression vector pIAFC109 (François Shareck, personal communication), under the control of the constitutive promoter C109, resulting in pIAFC109\_T7CM and pIAFC109\_T7W. Both constructions were introduced in *S. lividans* 10-164 (Hurtubise *et al.* 1995) by protoplasts transformation according to Kieser *et al.* (2000). Protein production and mycelium disruption was conducted as described by Nisole *et al.* (2006). The intracellular protein fractions were analyzed by Coomassie-stained SDS-PAGE and Western Blot (Fig. 5).



**Figure 5.** Analysis of the T7 RNAPol production by (A) Coomassie-stained SDS-PAGE and (B) Western blot. Lane 1, molecular mass standard; lane 2, positive control; lane 3, *S. lividans* 10-164/pIAFC109\_T7W; lane 4, *S. lividans* 10-164/pIAFC109\_T7CM.

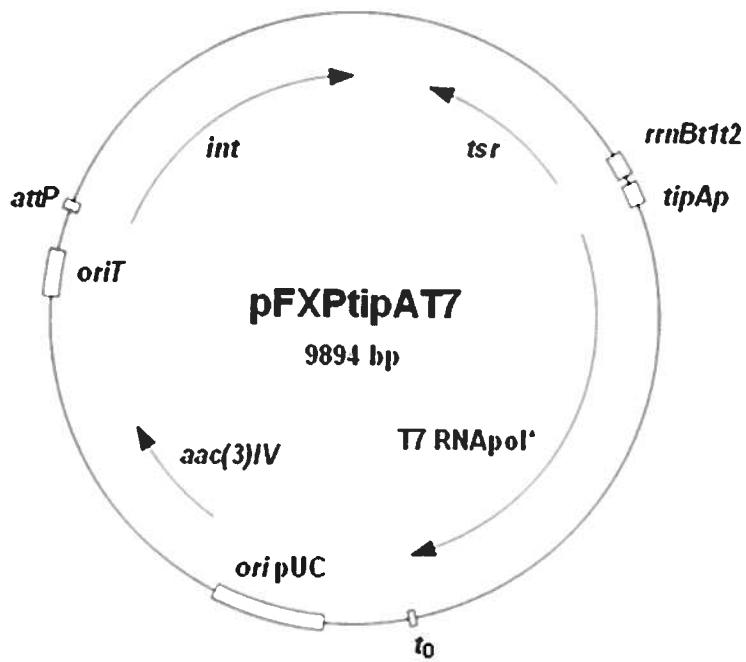
The T7 RNAPol producing strain *E. coli* BL21/pAR1219 (Davanloo *et al.* 1984) was used as positive control. Western blot was performed with anti-T7 RNA polymerase mouse monoclonal antibody (Novagen) and alkaline phosphatase-conjugated goat anti-mouse antibody (GE Healthcare). Colorimetric-detection was performed with the substrates Nitroblue Tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad). On the SDS-PAGE, a band of about 100 kDa appeared in the intracellular fraction of 10-164/pIAFC109\_T7CM, but was undetectable in 10-164/pIAFC109\_T7W. Western blot analysis confirmed that T7 RNAPol was produced with the codon-modified gene, but not detected with the wild-type version.

In *Streptomyces lividans*, TTA<sub>leu</sub> is the rarest codon (Codon Usage Database; [www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/)) and can be efficiently transcribed by only one tRNA encoded by *bldA* (Lawlor *et al.* 1987; Ueda *et al.* 1993). The *bldA* gene is constitutively transcribed, but the mature form of the tRNA seems to be present

only near the end of the logarithmic growth phase (Leskiw *et al.* 1993), when the secondary metabolism is activated (Chater et Chandra 2008). Different papers have shown that the presence of a TTA<sub>leu</sub> codon in a gene causes temporal expression delay or prevents translation in a *bldA* mutant strain (Leskiw *et al.* 1993; Kataoka *et al.* 1999; Rebets *et al.* 2006; O'Rourke *et al.* 2009). Since the T7 RNAPol gene contains four TTA<sub>leu</sub>, it was not surprising that its expression was impaired in *S. lividans*. By mutating these four codons, the T7 RNAPol gene was over-expressed for the first time in *S. lividans*.

## 2.8 *S. lividans* T7 expression strain

The codon-modified T7 RNAPol gene was cloned into a derivative of pSET152 (Bierman *et al.* 1992) under the control of the thiostrepton-inducible promoter *tipAp* (Murakami *et al.* 1989), resulting in pFXPtipAT7 (Fig. 6) (for detailed construction, see the supplemental material). This construction was integrated into the chromosome of *S. lividans* 10-164 via the bacteriophage φC31 att/int system (Bierman *et al.* 1992), giving rise to the apramycin and thiostrepton resistant *S. lividans* 10T7 strain. Integration at the chromosomal *attB* site was confirmed by PCR and DNA sequencing, but T7 RNAPol production by *S. lividans* 10T7 under thiostrepton-induced condition could not be detected by SDS-PAGE or Western blot (results not shown). The strain 10T7 has not shown any growth retardation in presence of thiostrepton compared to that of a non-induced culture (results not shown). It has to be noted that pSET152-derived plasmids can sometimes integrate as tandem repeat and into at least three pseudo-*attB* sites with a 300-fold lower efficiency (Combes *et al.* 2002). The plasmid pFXPtipAT7 can be used to create T7 expression strains in *S. lividans* and *Streptomyces coelicolor* with one simple transformation step (Takano *et al.* 1995; Combes *et al.* 2002).

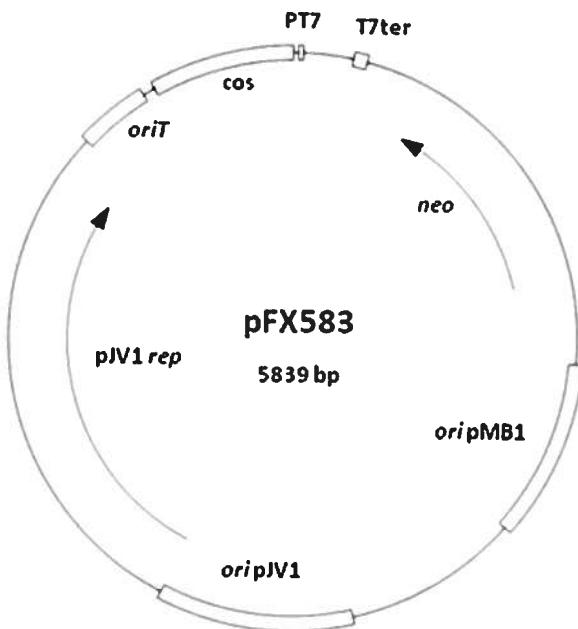


**Figure 6.** Structure of the integrative plasmid pFXPtipAT7. T7 RNAPol\*, codon-modified gene of the T7 RNA polymerase; *tipAp*, thiostrepton-inducible promoter; *aac(3)IV*, apramycin acetyltransferase, apramycin resistance; *tsr*, 23S A<sup>1067</sup> rRNA methylase, thiostrepton resistance; *int*, φC31 integrase; *rrnBt1t2*, transcription terminators T1 and T2 from *E. coli* *rrnB* gene; *t<sub>0</sub>*, λ transcription terminator; *ori pUC*, replication origin of pUC18; *oriT*, RK2 origin of transfer; *attP*, φC31 phage attachment site.

## 2.9 Bifunctional T7 expression vector

Construction of pFX583 was realized by using the *E. coli* T7 expression vector pET-9a (Novagen) (Studier *et al.* 1990) as backbone (Fig. 7). Detailed construction is presented in the supplemental material. The pFX583 vector contains pMB1 (Bolivar 1979) and pJV1 (Bailey *et al.* 1986) replicons, allowing replication in *E. coli* and *S. lividans* with a high-copy number. Kanamycin or neomycin selection can also be used with both bacteria due to the FD Neo-S cassette (Denis et Brzezinski 1991). The vector pFX583 is compatible with the

widely used *Streptomyces* pIJ101 replicon (Bailey *et al.* 1986). Shuttle vectors are very attractive because they allow to do all the DNA manipulation in *E. coli*, but are sometimes structurally unstable in *Streptomyces* for unknown reasons (Kieser *et al.* 2000). Here, pFX583 has been maintained in *E. coli* and *S. lividans* in presence of selection without notable structure instability. Since pFX583 harbours an *oriT* sequence, it can be transferred by conjugation from *E. coli* to *Streptomyces* strains that are difficult to transform. The presence of a  $\lambda$  *cos* sequence allows the use of pFX583 as a cosmid vector for large DNA fragment cloning.



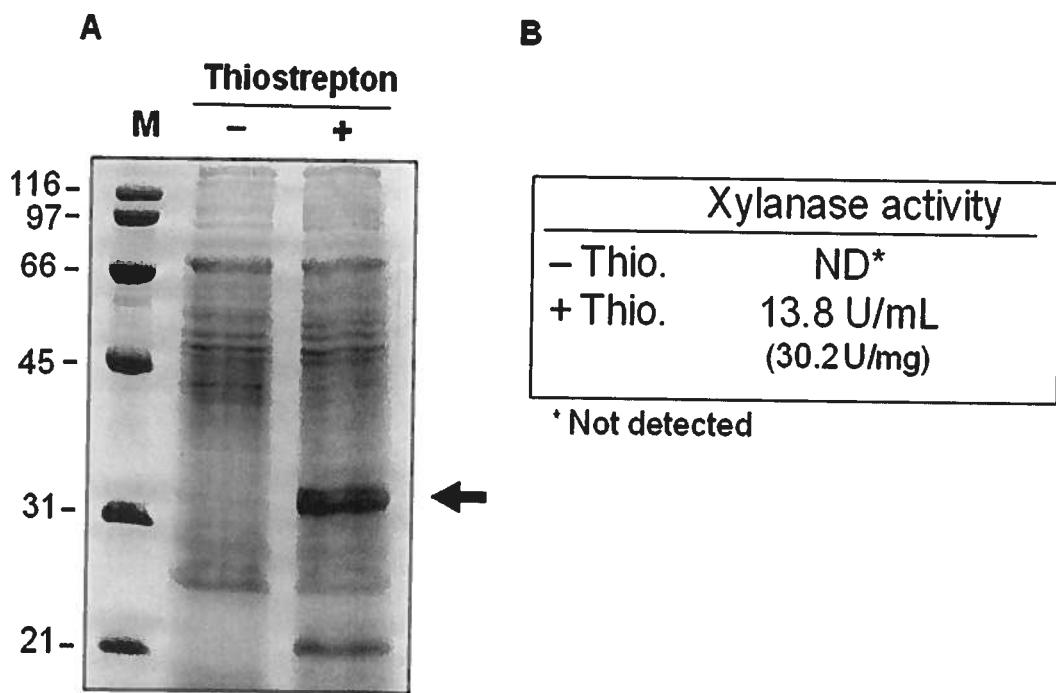
T7 promoter →  
 CCGCGAAATTAAATACGACTCACTATAAGGGAGACCACAACACGGTTCCCTAGAAATAATTTGTTAACCTT  
 SD NdeI XbaI  
 TAAGAAGGAGATATACATATGGCTAGCATGACTGGTGACAGCAAATGGTCGCGGATCCGGCTGCTAACA  
 AAGCCCGAAGGAAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTGGGGCCTCT  
 T7 terminator  
 AACGGGTCTTGAGGGTTTTTGCTGAAAGGAGGAA

**Figure 7.** Structure of the bifunctional T7-expression vector pFX583. PT7, T7 gene  $\phi 10$  promoter; T7ter, T7 transcription terminator  $\phi t$ ; neo, Tn903 aminoglycoside phosphotransferase, kanamycin and neomycin resistance; ori pMB1, replication origin of pMB1; ori pJV1, replication origin of pJV1; pJV1 rep, pJV1 Rep protein; oriT, RK2 origin of transfer; cos,  $\lambda$  cohesive end site; SD, Shine-Dalgarno sequence.

## 2.10 T7 RNA polymerase-directed xylanase production.

The efficacy of the *S. lividans* T7 expression system was demonstrated by over-expressing the gene encoding a truncated version of the *S. lividans* xylanase A (*xlnA2*) (Derewenda *et al.* 1994). The *xlnA2* gene was amplified by PCR and cloned into the *NdeI* and *BamHI* sites of pFX583. The resulting construction, pFX583xlnA2, was introduced into the *S. lividans* T7 expression strain 10T7 by protoplast transformation. Transformants were picked from R5 medium (Kieser *et al.* 2000) and streaked onto Bennett agar (Kieser *et al.* 2000) containing 50 µg/ml apramycin and 50 µg/ml kanamycin. After 3 to 4 days of incubation at 34°C, the sporulated mycelium was used to inoculate Tryptic Soy Broth medium (Difco) and cultured in Erlenmeyer flasks. Incubation was carried out at 34°C on a rotary shaker at 240 rpm for 48 h. Recombinant expression of *xlnA2* was induced by addition of thiostrepton to the culture medium. Different concentrations of thiostrepton were tested and 25 µg/ml allowed the highest XlnA2 production (see Fig. 9 in the supplemental material). Maximal enzyme production was obtained when thiostrepton was added at the beginning of the incubation period and significant activity increase stopped after 48 h (see Fig. 10 in the supplemental material).

Equal volumes of culture supernatants were analyzed by Coomassie-stained SDS-PAGE and xylanase activity was measured as described by Ebanks *et al.* (Ebanks *et al.* 2000) (Fig. 8). Xylanase A2 was absent in non-induced cultures, while readily detected as a 31-kDa band by SDS-PAGE in the presence of thiostrepton. As for all *tipAp*-based expression system, thiostrepton also induced the production of the TipAL protein that can be seen on the SDS-PAGE around 20 kDa (Murakami *et al.* 1989). Xylanase activity assays were consistent with SDS-PAGE analysis. After 48 h, no xylanase activity was measured in non-induced cultures, while 13.8 U/ml (30.2 U/mg) were detected under induced conditions, clearly demonstrating inducible *xlnA2* expression in *S. lividans* 10T7. Based on the specific activity of the purified XlnA2 (286 U/mg) (Ebanks *et al.* 2000), the concentration can be estimated to 48 mg/liter.



**Figure 8.** (A) Coomassie-stained SDS-PAGE showing the extracellular production of XlnA2 in non-induced (-) and induced (+) cultures. M, molecular mass standard. (B) Xylanase activity in non-induced (- Thio.) and induced (+ Thio.) cultures.

To confirm that pFX583 is also functional in *E. coli*, pFX582xlnA2 was introduced into the T7 expression strain BL21(DE3) (Novagen). Protein production was induced with 0.025 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and conducted for 20 h at 37°C. Supernatants of induced and non-induced cultures were analyzed by Coomassie-stained SDS-PAGE (see Fig. 11 in the supplemental material). As for *S. lividans*, the Xylanase A2 was produced under induced condition without obvious expression leaking in absence of IPTG.

## **2.11 Conclusion**

Combined with the bifunctional T7 expression vector pFX583, *S. lividans* 10T7 allowed inducible T7 RNAPol-directed overproduction of the xylanase A2 without detectable expression leaking in absence of inducer. Although the amount of protein produced was relatively low compared to what can be obtain with non-inducible *Streptomyces* expression systems (Lammertyn *et al.* 1997; Nisole *et al.* 2006; Dubé *et al.* 2008), the T7 expression system developed here presents interesting features. It is well regulated, has the potential to transcribe very large DNA fragments and can be used in combination with pIJ101-derived plasmids. The vector pFX583 is functional in *E. coli* and *Streptomyces* strains producing T7 RNAPol. With a single construction it is therefore possible to compare the expression of a gene in two kinds of host and determine which one is the most appropriate based on the productivity and requirements of the study.

## **2.12 Acknowledgments**

This work was supported by a Strategic Grant from the Natural Sciences and Engineering Council of Canada. Plasmid pSET152 was kindly provided by Marie A. Elliot from McMaster University.

## **2.13 Supplemental material**

### **2.13.1 T7 RNA polymerase gene mutation**

The four TTA<sub>leu</sub> codons in the T7 RNA polymerase gene were mutated by overlap extension PCR. Individual mutations were introduced in five independent PCR amplifications using primers pairs T7.1/T7.2, T7.3/T7.4, T7.5/T7.6, T7.7/T7.8 and T7.9/T7.10, using the plasmid pAX1219 (T7 RNAPol gene cloned in pAX01 (Hartl *et al.* 2001) as the template. The complete T7 RNAPol gene was reconstituted in a second PCR reaction in which the five fragments were mixed in

equal ratios, using primers pair T7.1/T7.10. PCR amplifications were performed using proof reading DNA polymerase Pfu Turbo (Stratagene) and the final product was sequenced at the McGill University and Génome Québec Innovation Centre.

### **2.13.2 Construction of pFXPtipAT7**

The *S. lividans* *tipA* promoter (*tipAp*) was amplified from chromosomal DNA using primer pair ptipA5/ptipA3. To place a transcription terminator (lambda *t<sub>O</sub>*) downstream of *tipAp*, the promoter was cloned between the *Bg*/II and *Sma*I restriction sites of a modified version of pAX01 (Hartl *et al.* 2001). The promoter and the *t<sub>O</sub>* terminator were excised with *Bg*/II and *Eco*RV and ligated to the *S. lividans* integrative plasmid pSET152 (Bierman *et al.* 1992) digested by *Bam*HI and *Eco*RV. To block transcriptional read-through, the transcription terminators T1 and T2 from *E. coli rrnB* gene obtained from pAX01 were cloned upstream of *PtipA* into the *Xba*I restriction site. The codon-enhanced T7 RNAPol gene was inserted in the *Bam*HI site located downstream of *tipAp*. The selection marker conferring resistance to thiostrepton (*tsr*) was excised of pIJ702 (Katz *et al.* 1983) by *Bcl*I digestion and introduced into the *Bg*/II site to generate the integrative plasmid pFXPtipAT7.

### **2.13.3 Construction of the pFX583**

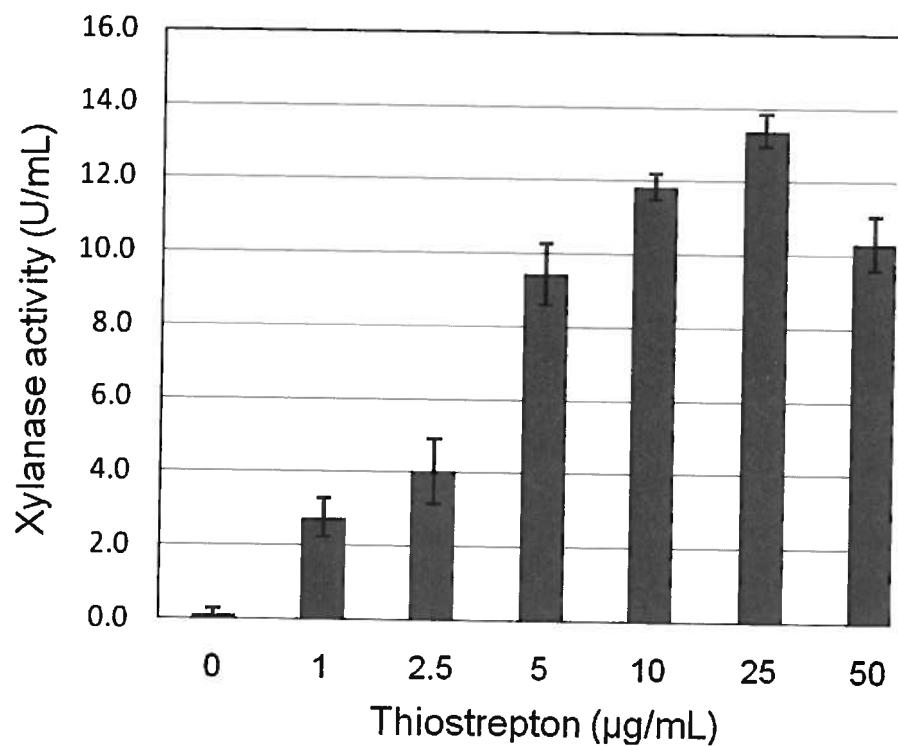
The *Nru*I/*Psil* fragment from pFD666 (Denis et Brzezinski 1992) containing the pJV1 *Streptomyces* replicon was introduced into the pET-9a (Novagen) (Studier *et al.* 1990) *E. coli* T7 expression vector digested with *Nru*I and *Xmn*I. The kanamycin resistance gene of pET-9a was replaced by the FD Neo-S cassette (Denis et Brzezinski 1991). The FD Neo-S cassette was amplified by PCR with the primer pair neoxho5/neoebo3 using pFD666 as a template and introduced between *Xho*I/*Eco*RI sites. A  $\lambda$  *cos* and a RK2 *oriT* sequences were introduced in *Bg*/II/*Sph*I and *Acl*I/*Sph*I sites, respectively, to generate the shuttle T7 expression

vector pFX583. The  $\lambda$  *cos* was amplified from pWEB-TNC (Epicentre Biotechnologies) and *oriT* from pSET152 using the primer pairs cosII5/cosII3 and ori5/ori3, respectively.

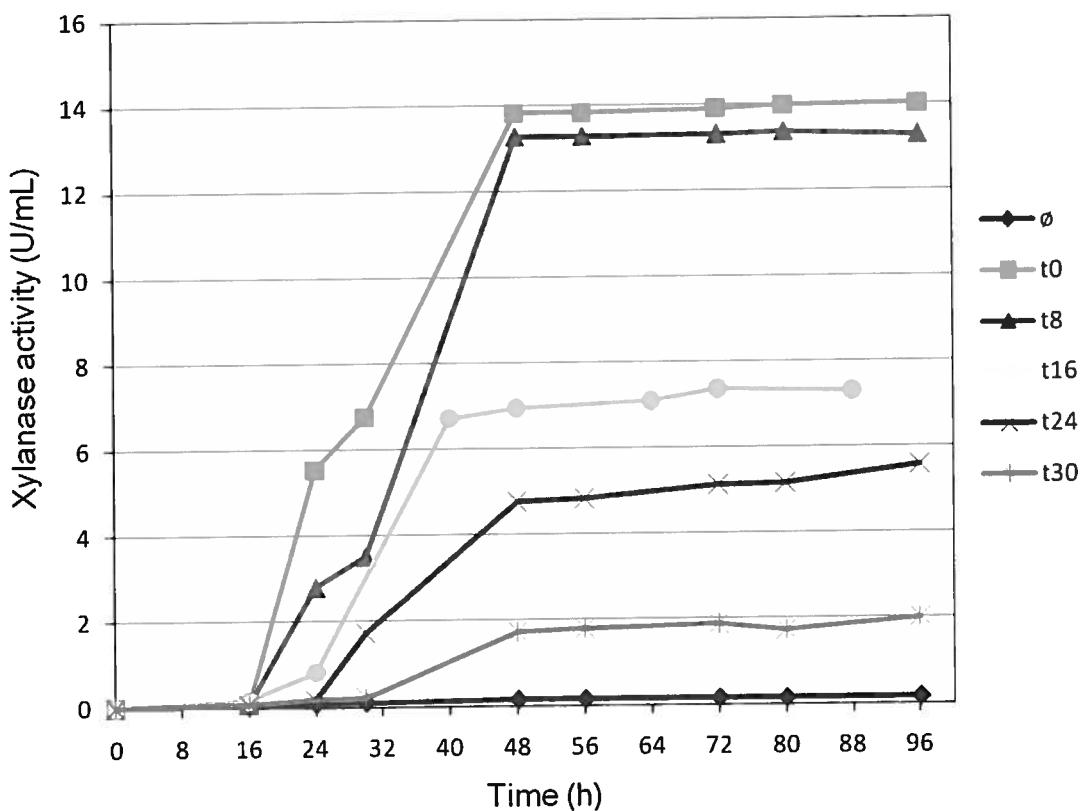
**Table VII.** Oligonucleotides used in this work.

Oligonucleotide	Sequence (5'→3')	Usage
T7.1	GGGAAATGACAAATGGTCCAAACTAGTG	T7 RNAPol mutation
T7.2	CTGTTCGCGAGCGAGACGCTCACCGTAATG	T7 RNAPol mutation
T7.3	CATTACGGTGAGCGTCTCGCTCGAACAG	T7 RNAPol mutation
T7.4	TTTGGCGGTGGAGGCTAACCATTCGGTTGAC	T7 RNAPol mutation
T7.5	ACCGGAATGGTTAGCCTCCACCGCCAAATGCTG	T7 RNAPol mutation
T7.6	GTGTTAATGGTAGGCTGGAGGCGGAACGTGAC	T7 RNAPol mutation
T7.7	CGGTCAGTCCGCCTCCAGCCTACCATAACAC	T7 RNAPol mutation
T7.8	GAAGTCGACTCGAGGATGTCACGGAGGTTG	T7 RNAPol mutation
T7.9	AACCTCCGTGACATCCTCGAGTCGGACTTC	T7 RNAPol mutation
T7.10	GATATGGTGCAAGTCAGCACGAAC	T7 RNAPol mutation
ptipA5	<u>CATAGAT</u> CTCCGGGCTGAGGGAGCCGAC	<i>tipAp</i> amplification
ptipA3	AAGAATT CGGATCCTCCGCTCCCTCTCTG	<i>tipAp</i> amplification
neoxho5	GAACACT <u>CGAG</u> CGCAGCGAGTCAGTGAGC	FD Neo-S cassette amplification
neoeco3	ATCCC <u>GAATT</u> CCCGCTCAGAAGAACTCGTC	FD Neo-S cassette amplification
cosII5	GTCCTGT <u>AGAT</u> CTGCCCTCGCTGGCCTG	$\lambda$ <i>cos</i> amplification
cosII3	GTGTTAG <u>CATG</u> CAGCAGTGCCGTGATAG	$\lambda$ <i>cos</i> amplification
ori5	GGTTT <u>CAAC</u> GGTTCATCCGCTTGCCCTCATCTG	<i>oriT</i> amplification
ori3	TCCGTCA <u>GAGT</u> ACGGGCCATAG	<i>oriT</i> amplification
xlnANdeI5	GGAGAAC <u>ATAT</u> CGCCTCCACGCCCTTCC	<i>xlnA2</i> amplification
xlnABamHI3	CTGATGG <u>ATCC</u> TCAGCCGTTGAGTGCCTC	<i>xlnA2</i> amplification

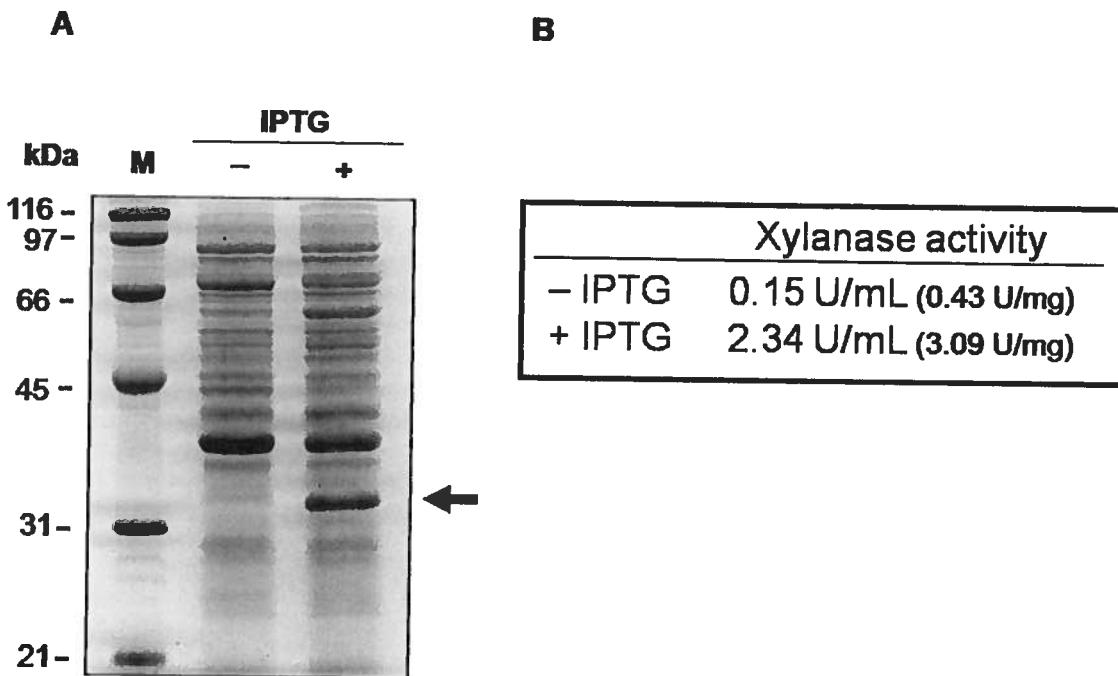
Restriction sites are underlined.



**Figure 9.** Xylanase activity *vs.* concentration of thiostrepton used for induction. Cultures were incubated 48 h at 34 °C and thiostrepton was added at the beginning of the incubation period. Xylanase activity assay was performed on culture supernatants.



**Figure 10.** Xylanase activity from cultured induced at different times. Cultures were incubated at 34 °C up to 96 h. Thiostrepton (25 µg/mL) was added at the beginning of the incubation period (t0), after 8 h (t8), 16 h (t16), 24 h (t24) and 30 h (t30). Xylanase activity was also measured in a non-induced culture (Ø).



**Figure 11.** Xylanase production in *E. coli* BL21(DE3)/pFX583xlnA2. (A) Coomassie-stained SDS-PAGE showing the extracellular production of XlnA2 in non-induced (-) and induced (+) cultures. M, molecular mass standard. (B) Xylanase activity in non-induced and induced cultures. The activity was determined by incubating the culture supernatants with birchwood xylan (5 mg/mL) for 10 min at 60 °C and measuring the reducing sugars. One unit (U) represents the amount of enzyme releasing 1 µmol/min of reducing sugar.

**3. DÉVELOPPEMENT D'UN SYSTÈME  
D'EXPRESSION INDUCTIBLE CHEZ *BACILLUS*  
*SUBTILIS***

### **3.1. Mise en contexte**

Après *Escherichia coli*, *Bacillus subtilis* est l'hôte bactérien le plus couramment utilisé pour la production de protéines recombinantes (Terpe 2006). *B. subtilis* présente plusieurs caractéristiques avantageuses, dont une grande capacité de sécrétion, l'absence d'endotoxines et un usage de codons non-biaisé (Li *et al.* 2004; Terpe 2006). Cette section présente le développement d'un nouveau système d'expression inductible chez *B. subtilis*, basé sur l'ARN polymérase du phage T7. La présente section devait faire l'objet d'un article et a donc été rédigée dans cette optique.

### **3.2 Contribution de l'étudiant**

Les travaux présentés dans cette section ont été entièrement réalisés par l'étudiant, sous la supervision de François Denis (Ph.D.) et François Shareck (Ph.D.).

### **3.3 Résumé**

Le système d'expression T7 développé initialement chez *E. coli* a été adapté à *B. subtilis*. Le gène codant pour l'ARN polymérase T7 a été intégré dans le chromosome de *B. subtilis* WB800, sous le contrôle d'un promoteur inducible par le xylose. La souche WB800 comporte une activité protéolytique très réduite de par l'inactivation des huit protéases extracellulaires connues. Un vecteur d'expression navette a été construit à l'aide du réplicon de pBS72, reconnu pour sa grande stabilité structurelle. Ce vecteur peut être utilisé chez *E. coli* et *B. subtilis*. La fonctionnalité du nouveau système d'expression T7 de *B. subtilis* a été évaluée à l'aide du gène *amyE* codant pour une  $\alpha$ -amylase sécrétée. L'ajout de xylose au milieu de culture induisait la production d'ARN polymérase T7 qui transcrivait à son tour le gène cloné dans le vecteur. Ainsi, l' $\alpha$ -amylase a été produite dans le milieu de culture, démontrant la fonctionnalité du système. Toutefois, la quantité d'enzymes produites était relativement faible. Afin d'augmenter la productivité, les limitations du système d'expression devraient être identifiées et si possible, corrigées.

### **3.4 Abstract**

The T7 expression system originally developed in *E. coli* was adapted to *B. subtilis*. The gene encoding the T7 RNA polymerase was integrated into the chromosome of *B. subtilis* WB800, under the control of an inducible promoter. The strain WB800 has a very low extracellular proteolytic activity and is therefore suited for the production of heterologous proteins. A shuttle expression vector was constructed using the replicon of pBS72, which confers a high structural stability. This vector can be used in both *E. coli* and *B. subtilis*. The functionality of the new T7 expression system of *B. subtilis* was evaluated using a gene encoding a secreted  $\alpha$ -amylase. The addition of xylose to the culture medium induced the production of T7 RNA polymerase, which in turn transcribed the gene cloned into the vector. Thus, the  $\alpha$ -amylase was produced in the culture medium,

demonstrating the functionality of the expression system. However, the amount of enzyme produced was relatively low. To increase the productivity, the limitations of the expression system should be identified and if possible, corrected.

### 3.5 Introduction

The production of recombinant proteins is the cornerstone of numerous research domains of life science. For this purpose, a variety of expression systems are available, using bacteria, yeast, insect or mammalian cells as hosts. Among those, the Gram-negative bacterium *Escherichia coli* is the most widely used, being well established in many laboratories (Terpe 2006). *E. coli* is also commonly used for the production of industrial and pharmaceutical proteins (Schmidt 2004). The popularity of this bacterial host can be explained by its ease of handling, its well-characterized genetics, its short generation time in inexpensive medium and the availability of a large number of vectors and specialized strains. Many expression systems have been developed in *E. coli* based on different inducible promoters such as *P<sub>lac</sub>*, *ParABAD* and *P<sub>trc/tac</sub>* (Gronenborn 1976; Brosius *et al.* 1985; Guzman *et al.* 1995). Although it is difficult to predict which expression system is best suited for the heterologous production of a specific protein, the *E. coli* T7-expression system developed by Studier and Moffatt (1986) has shown a high success rate (Graslund *et al.* 2008). In this expression system, the host harbours a chromosomal copy of the gene encoding for T7 RNA polymerase (T7 RNAPol) under the control of the *lacUV5* promoter (Wanner *et al.* 1977). Upon induction, the T7 RNAPol is produced and specifically transcribes the target gene cloned downstream of a T7 late promoter in the plasmid vector. The very high processivity of the T7 RNAPol is mainly responsible for the great efficiency of the *E. coli* T7-expression system. The T7 RNAPol can elongate mRNA chains about five times faster than *E. coli* RNA polymerase (Golomb et Chamberlin 1974), producing enough messenger RNA (mRNA) to saturate the protein-synthesising machinery of the host. It also possesses a stringent selectivity toward

T7 late promoters, so it only transcribes target genes cloned into the expression vector.

When using *E. coli* as a host, heterologous proteins can be produced either in the cytoplasm or in the periplasm. Secreted proteins can also leak from the periplasm and be recovered in the culture medium (Shokri *et al.* 2003). Because of higher yield, cytoplasmic production is more often used, but over-produced proteins tend to accumulate in inclusion bodies (Sorensen et Mortensen 2005). The process for recovering properly folded and active proteins involves cell disruption, inclusion body isolation, denaturation and renaturation steps, followed by purification of the target from contaminating host proteins. Although *E. coli* has proven to be a very useful expression host, different problems can hamper the production of recombinant proteins, such as cell toxicity, misfolding, inappropriate post-translational modification and codon bias. When those problems cannot be circumvented, one solution is the use of an alternative host.

After *E. coli*, the Gram-positive bacterium *Bacillus subtilis* is perhaps the most popular expression host for the production of heterologous proteins (Terpe 2006). Like *E. coli*, *B. subtilis* is very well characterized and has become the paradigm of Gram-positive genetics and physiology. Over the years, several expression systems have been developed, based on replicating and integrative vectors (Schumann 2007). One of the great features of *B. subtilis* is its huge capacity to secrete proteins directly into the culture medium, which greatly facilitates downstream processing (Harwood 1992). In contrast to *E. coli*, *B. subtilis* does not produce pyrogenic lipopolysaccharides and bears “generally recognized as Safe” (GRAS) status from the U.S. Food and Drug Administration (FDA). Also, it has no significant codon usage bias (Codon Usage Database, <http://www.kazusa.or.jp/codon/>). Nevertheless, one of the major limitations of *B. subtilis* as an expression host has been its high extracellular proteolytic activity (Wong 1995). This problem was largely solved by the development of protease-deficient strains, notably the strains WB800 (Wu *et al.* 2002) and Dpr8 (Kodama *et al.* 2007b), in which the eight known extracellular proteases were inactivated. One other problem that has limited the use of *B. subtilis* is the structural

instability of the expression vectors that replicate by rolling-circle mechanisms (Bron *et al.* 1991b). Recently, a series of expression vectors was developed using the replicon of the cryptic *B. subtilis* plasmid pBS72 (Nguyen *et al.* 2005). These low copy expression vectors replicate as theta circles and exhibited full structural stability.

Since the original publication of the *E. coli* T7-expression system, it has been adapted to mammalian cells, *Pseudomonas aeruginosa*, *Erwinia carotovora*, *Salmonella choleraesuis*, *Agrobacterium tumefaciens*, *Chromobacterium violaceum*, *Rhodobacter capsulatus*, *Ralstonia eutropha*, *Bacillus megaterium* and *Streptomyces lividans* (Elroy-Stein et Moss 1990; Barnard *et al.* 2004; Drepper *et al.* 2005; Kang *et al.* 2007; Gamer *et al.* 2009; Lussier *et al.* 2010b). A T7-expression system was also developed in *B. subtilis* (Conrad *et al.* 1996), but its use has not been reported in any publication since it was developed more than ten years ago. The expression strain have been constructed in the *B. subtilis* strain BD170 (*trp-2 thr-5*) (Dubnau *et al.* 1973) and the vector was based on the rolling circle-type plasmid pUB110 (Keggins *et al.* 1978). During the redaction of this paper, another T7-expression system was developed in *B. subtilis*, using an integrative expression vector (Chen *et al.* 2010).

In this study, a T7-expression system was developed with the protease-deficient strain *B. subtilis* WB800 with the aim to combine the T7 RNAPol efficiency to the great secretion capacity of this host. A shuttle vector, allowing T7 RNAPol-directed transcription in *B. subtilis* and *E. coli*, was constructed with the pBS72 replicon and the commercial plasmid pET-9a (Novagen). The new expression system efficacy was assessed using a *B. subtilis*  $\alpha$ -amylase.

### 3.6 Materials and methods

#### 3.6.1 Bacterial strains, plasmids and oligonucleotides

Bacterial strains and plasmids used in this study are listed in Table VIII, and the oligonucleotides used for PCR amplification are presented in Table IX.

**Table VIII.** Strains and plasmids used in this work.

Strain and plasmid	Genotype and description	Source or reference
<b>Strain</b>		
<i>E. coli</i> DH5α	<i>F</i> $\phi$ 80lacZΔM15 Δ( <i>lacZYA-argF</i> )U169 <i>deoR</i> <i>recA1 endA1 hsdRI7(r<sub>k</sub><sup>-</sup>, m<sub>k</sub><sup>+</sup>)phoA supE44</i> <i>thi-1 gyrA96 relA1 λ-</i>	Invitrogen
<i>B. subtilis</i> WB800	<i>nprE aprE epr bpr mpr::ble nprB::bsr Δvpr wprA::hyg</i>	(Wu <i>et al.</i> 2002)
<i>B. subtilis</i> WBT7	Derivative of WB800, <i>lacA::[xylR-xylAp-T7 RNApol-ery]</i>	This study
<b>Plasmid</b>		
pAR1219	<i>bla, ori pMB1, T7 RNApol</i>	(Davanloo <i>et al.</i> 1984)
pAX01	<i>bla, ori pMB1, lacA::[xylR-xylAp-ery]</i>	(Hartl <i>et al.</i> 2001)
pAXT7	Derivative of pAX01 containing T7 RNApol	This study
pET-9a	<i>km, ori pMB1, T7 pro moter and terminator</i>	Novagen, (Studier <i>et al.</i> 1990)
pHMC04	<i>bla, cat, ori/rep pBS72, ori pBM1, xylR, xylAp</i>	(Nguyen <i>et al.</i> 2005)
pFX583	<i>neo, ori/rep pJV1, ori pMB1, RK2 oriT, λ cos, T7 pro moter and terminator</i>	(Lussier <i>et al.</i> 2010b)
pFX64	<i>spc, km, ori/rep pBS72, ori pBM1, , RK2 oriT, λ cos, T7 pro mter and terminator</i>	This study

*ble, bsr, hyg, ery, bla, km, cat, neo* and *spc* are antibiotic resistance markers for bleomycin, blasticidin S, hygromycin, erythromycin, ampicillin, kanamycin, chloramphenicol, neomycin and spectinomycin, respectively.

**Table IX.** Oligonucleotides used in this work.

Oligonucleotide	Sequence (5'→3')
yvfM5	AGCCCTGCTTATCAGCATAC
erm3	GAACACTAGGGTTGCTCTTG
xylR5	AGCGTTACTGGCTTCGCTCTC
yvfO3	ATGCGGACGCGAACATAGTTG
spec5	GGCCAT <u>GAATTCT</u> TGACATTTTCTTGTGGATCTGTAT
spec3	AATAAAGAAATAATTA
repori5	GGCCAT <u>GATATCCA</u> ATAGTTACAAATTGTTCACTA
repori3	TCTGGCAGCATTCCGAAGG
amyE5rbs	GCTAC <u>GAATTCT</u> TGACTTAGCGACCCTGAG
amyE5	GAGGGACAT <u>ATGACTCCGAA</u> GTAAAGTCTTC
amyE3	GTGTCAC <u>ATATGTTGCAA</u> ACGATTCAAAAC
	TAGCCGGAT <u>CCTCA</u> ATGGGGAAAGAGAAC

Restriction sites are underlined.

### 3.6.2 Culture media, growth conditions and recombinant gene expression

*Escherichia coli* and *Bacillus subtilis* strains were grown in 2xTY medium (16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl) with or without agar, at 37 °C. Antibiotics were added to growth media at the following concentrations: kanamycin at 30 µg/ml, erythromycin at 1 µg/ml and spectinomycin at 170 µg/ml. In *B. subtilis*, recombinant gene expression was induced by the addition of D-xylose (2 %) at an OD<sub>578</sub> of 0.4.

### 3.6.3 DNA preparation and manipulation

Standard procedures were used for the preparation and manipulation of DNA (Sambrook *et al.* 1989). All plasmid constructions and gene cloning were realized in *E. coli* DH5α (Invitrogen). *B. subtilis* was transformed by electroporation as

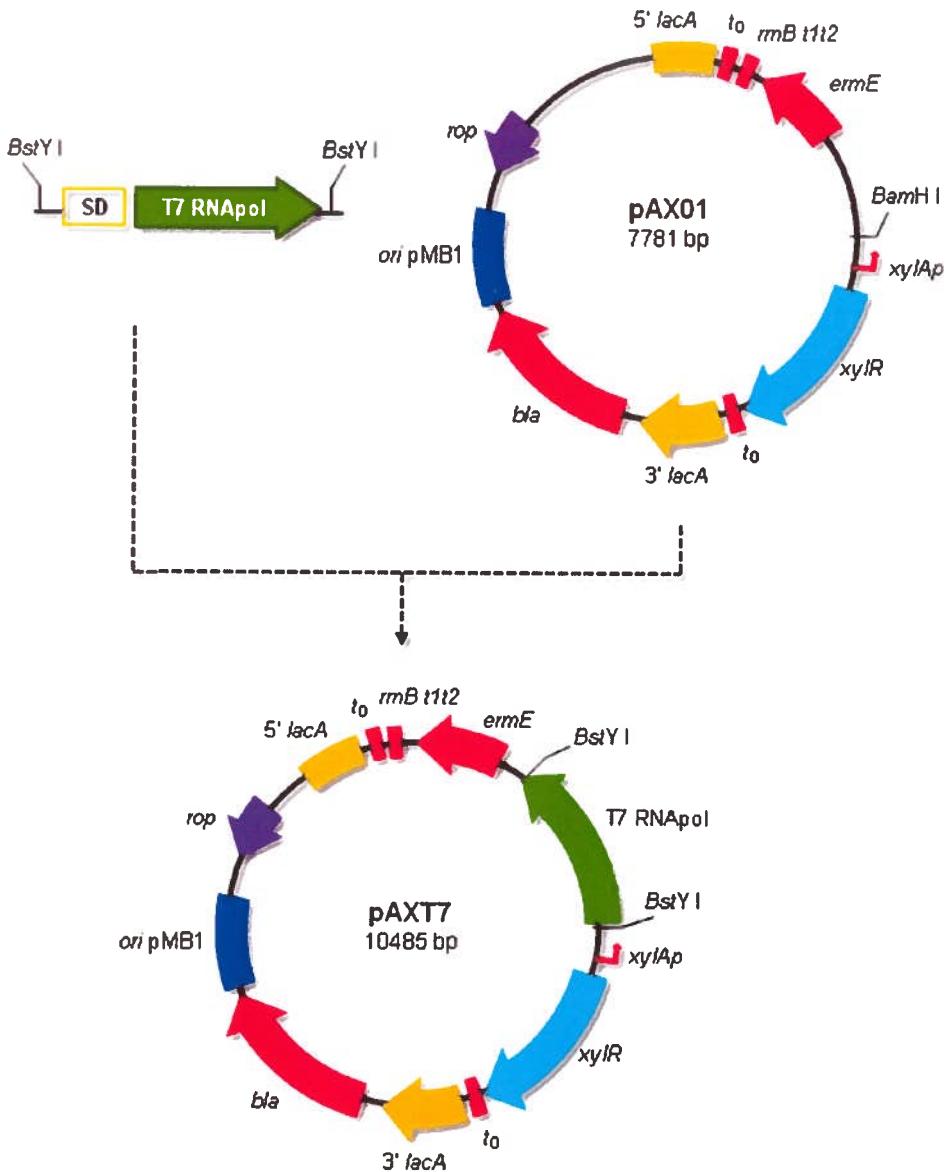
described by Xue *et al.* (1999). PCR amplifications were performed using proof reading Pwo DNA polymerase (Roche) and all products were sequenced at the McGill University and Génome Québec Innovation Centre.

### 3.6.4 Construction of the *B. subtilis* T7-expression host

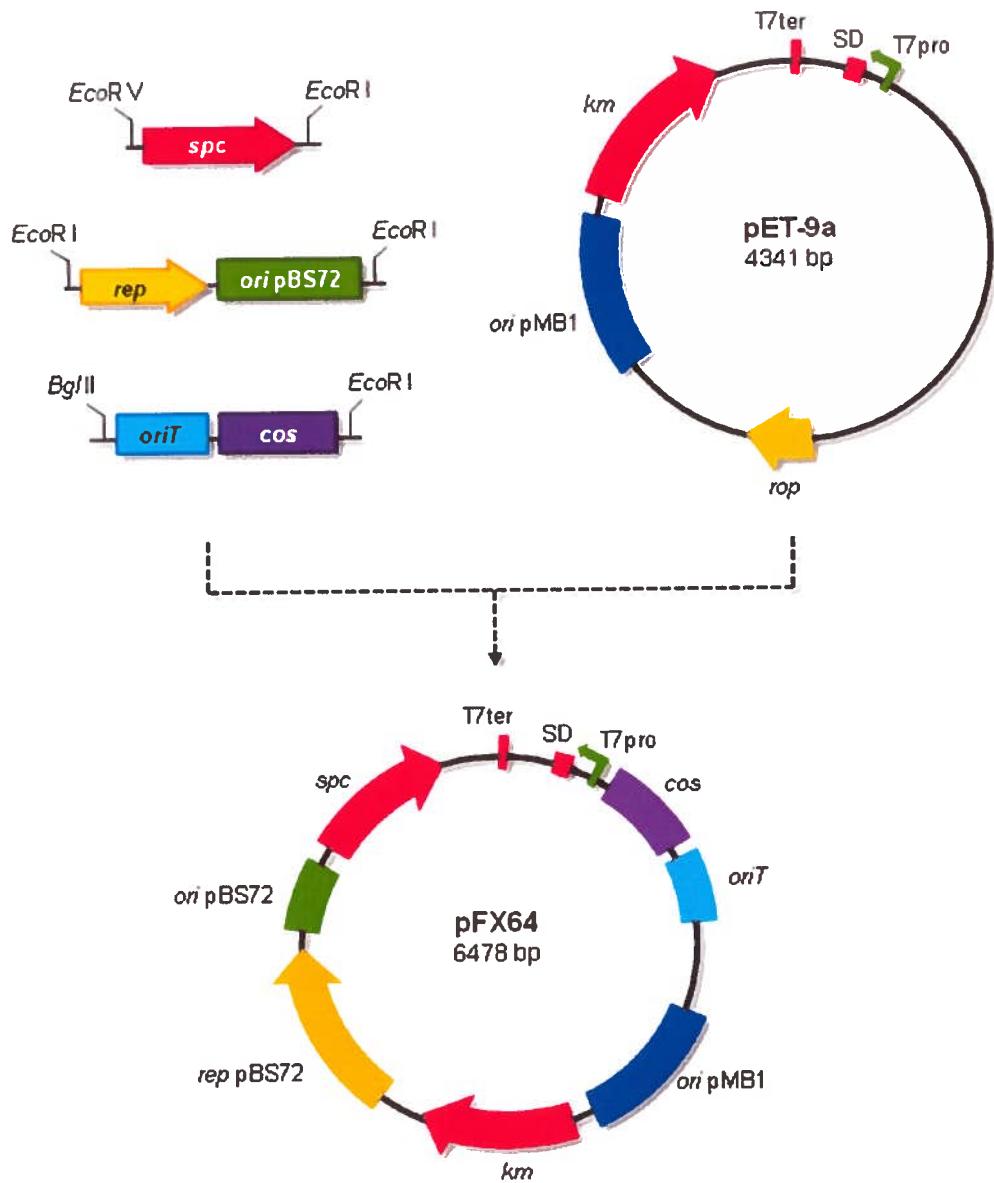
The T7 gene *l*, encoding for the phage RNA polymerase (T7 RNAPol), was obtained from pAR1219 (Davanloo *et al.* 1984) by a *Bst*YI digest and ligated into *Bam*HI-linearized vector pAX01(Hartl *et al.* 2001) (Figure 12). The resulting construction pAXT7 was introduced into the *B. subtilis* protease-deficient strain WB800 (Wu *et al.* 2002) via natural competence (Yasbin *et al.* 1975). Transformants were selected with erythromycin and chromosomal integration in the *lacA* locus was confirmed by PCR with pairs of primers yvfM5/erm3 and xylR5/yvfO3, and by sequencing.

### 3.6.5 Construction of the T7-expression vector

The pFX64 shuttle T7-expression vector was constructed by using the *E. coli* pET-9a vector (Novagen) as backbone. The *spc* gene conferring resistance to spectinomycin was amplified from *B. subtilis* IHA01 (Hartl *et al.* 2001) chromosomal DNA using primers spec5 and spec3. The resulting 1.1 kb fragment was digested with *Eco*RV and *Eco*RI and ligated into pET-9a linearized with the same restriction enzymes. The *B. subtilis* pBS72 (Lagodich *et al.* 2004) replicon, amplified from vector pHCMC04 (Nguyen *et al.* 2005) with primers repori5 and repori3, was cloned into the *Eco*RI site. A *Bgl*II/*Acl*I fragment from pFX583 (Lussier *et al.* 2010a) containing a RK2 transfer origin (*oriT*) and a  $\lambda$  *cos* sequence was cloned into the same restriction site. Finally, a 198 pb DNA fragment containing *rop* was deleted by a *Acl*I/*Bpm*I digestion and the blunted plasmid was recircularized, giving rise to pFX64 (Figure 13).



**Figure 12.** Structure of the integrative plasmid pAXT7. T7 RNAPol, T7 RNA polymerase-encoding gene; SD, Shine-Dalgarno sequence; *rop*, repressor of primer (copy number control); *ori pMB1*, replicon for replication in *E. coli*; *bla*, β-lactamase, ampicillin resistance; 5'*lacA* and 3'*lacA*, disrupted *lacA* for chromosomal integration in *B. subtilis*; *rrnBt1t2*, transcription terminators T1 and T2 from *E. coli rrnB* gene; *t<sub>0</sub>*, λ transcription terminator; *xyIR*, xylose repressor protein; *xylAp*, xylose-inducible promoter, *ermE*, N-6-aminoadenine-N-methyltransferase, erythromycin resistance.



**Figure 13.** Structure of the expression vector pFX64. T7pro, T7 gene  $\varphi 10$  promoter; T7ter, T7 transcription terminator  $\varphi t$ ; SD, Shine-Dalgarno sequence; *spc*, spectinomycin adenyltransferase, spectinomycin resistance; *ori/rep* pBS72, replicon for replication in *B. subtilis*; *oriT*, RK2 origin of transfer; *cos*,  $\lambda$  cohesive end site; *km*, aminoglycoside phosphotransferase, kanamycin resistance; *ori pMB1*, replicon for replication in *E. coli*; *rop*, repressor of primer (copy number control).

### **3.6.6 T7 RNA polymerase-directed gene expression**

The  $\alpha$ -amylase-encoding gene (*amyE*) was amplified with and without its ribosome binding site (RBS) sequence using, respectively, the primers pairs amyE5rbs/amyE3 and amyE5\amyE3. The resulting DNA fragments were cloned into pFX64 using the *Nde*I and *Bam*HI restriction sites. The ligated DNA was then transformed in *E. coli*. After obtaining the right constructions, the *amyE*-containing vectors were transferred into *B. subtilis* WBT7. Expression was induced by the addition of 2 % D-xylose and incubated at 37 °C for 8 h.

### **3.6.7 Analysis of recombinant protein production**

Culture supernatant and cells were recovered by centrifugation. Cells were resuspended in 100 mM sodium phosphate buffer (pH 7.0) containing 0.1 mM phenylmethylsulfonyl fluoride and disrupted by sonication (XL2020TM, Heat Systems). Supernatants and lysates were kept at -20 °C. The Bradford protein assay method (Bradford reagent, Sigma-Aldrich) was used for protein concentration determination. Acetone precipitation was used to concentrate proteins from culture supernatants (2 volumes acetone : 1 volume supernatant). Separation of proteins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (Laemmli 1970) and proteins were visualized by Coomassie blue staining. Detection of T7 RNAPol by Western blotting was performed as described by Lussier *et al.* (2010b). Amylase assays were realized on agar medium supplemented with 1 % (m/v) starch. Culture supernatants were loaded into wells in the agar medium and the plates were incubated 16 h at 37° C. Hydrolyse zones were revealed by iodine staining.

## 3.7 Results

### 3.7.1 *B. subtilis* T7-expression host

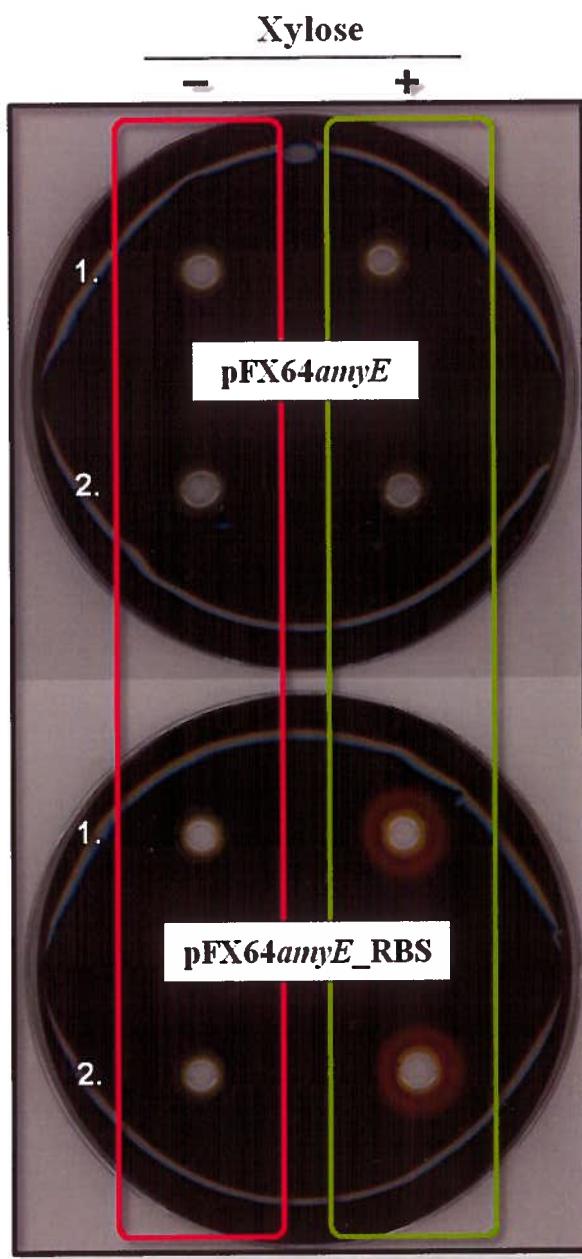
The *B. subtilis* WBT7 strain was obtained by using the integrative vector pAXT7, in which the T7 RNAPol gene was placed under the control of the xylose-inducible *xylA* promoter (*xylAp*) of *Bacillus megaterium* (Rygus et Hillen 1991). Ectopic integration into the *lacA* locus of *B. subtilis* WB800 resulted in an erythromycin-resistant strain that can produce T7 RNAPol under xylose-induced condition. PCR amplification and DNA sequencing confirmed the chromosomal integration by double crossover. Production of T7 RNAPol by *B. subtilis* WBT7 was evaluated by Western blotting. Cells were grown under induced condition for a period of 20 h and processed as mentioned above. No T7 RNAPol could be detected (result not shown).

### 3.7.2 Shuttle T7-expression vector

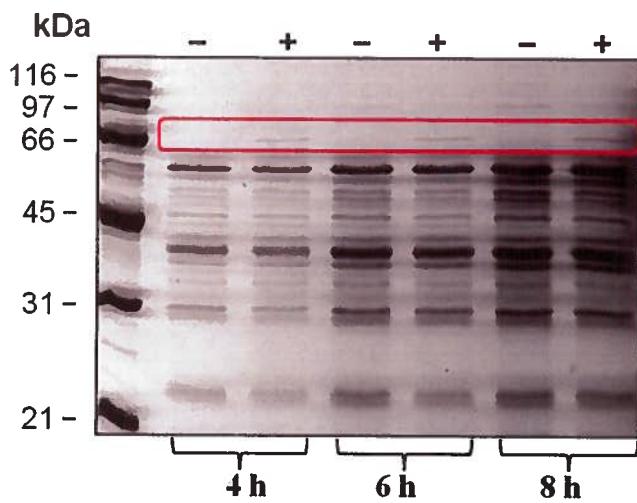
To be used with *B. subtilis* WBT7, a shuttle vector containing the T7 promoter  $\phi 10$  and the T7 transcription terminator  $t\phi$  was constructed (Figure 13). The vector pFX64 features a pMB1 and a pBS72 replicon, allowing replication in *E. coli* and *B. subtilis*. The  $\lambda$  *cos* sequence allows the use of pFX64 as a cosmid vector for large DNA fragment cloning and can be mobilized via the RK2 *oriT* sequence. Because the kanamycin selection marker (*km*) originally present in pET-9a was not functional in *B. subtilis*, *spc* was added to the construction. For unknown reasons, the copy number of the vector seemed to have been reduced in *E. coli* after the addition of the *oriT* and *cos* sequences. Deletion of *rop* helped to restore the high copy number.

### **3.7.3 Extracellular protein production**

The efficiency of the *B. subtilis* T7 RNAPol-dependent expression system to produce secreted proteins was evaluated by using the homologous  $\alpha$ -amylase (AmyE) as model. The gene *amyE* was cloned in pFX64 with and without its RBS sequence, resulting in pFX64*amyE*\_RBS and pFX64*amyE*, respectively. After 8 h of incubation under induced condition, the culture supernatants were recovered and the amylolytic activity was assayed on agar plates containing starch (Figure 14). After iodine staining, hydrolyse zones could be seen for WBT7/ pFX64*amyE*\_RBS but not for WBT7/ pFX64*amyE*. No amylolytic activity was detected for the non-induced controls. Culture supernatants were also analyzed by Coomassie-stained SDS-PAGE (Figure 15). In accordance with the amylolytic assay, AmyE was detected only in the supernatant of WBT7/ pFX64*amyE*\_SD as a very faint band. Numerous unknown proteins were also detected in the supernatant, independently of the presence of the inducer.



**Figure 14.** Agar plate assay for the evaluation of the  $\alpha$ -amylase production in the culture supernatant of WBT7/pFX64*amyE* and WBT7/pFX64*amyE\_RBS*, under induced and non-induced conditions. Two clones were analyzed for each strain.



**Figure 15.** Coomassie-stained SDS-PAGE showing the extracellular production of  $\alpha$ -amylase in non-induced (-) and induced (+) cultures of WBT7/pFX64amyE\_RBS, 4h, 6h and 8h post-induction.

### 3.8 Discussion

Conrad *et al.* (1996) have already demonstrated that T7 RNAPol-dependent expression was feasible in *B. subtilis*. In their expression system, the T7 RNAPol-encoding gene was chromosomally integrated into the *amyE* locus of the *B. subtilis* strain BD170, under the control of the xylose-inducible promoter *xylAp*. They showed that the T7 RNAPol was produced as soon as 30 min after induction. By using their multicopy pUB110-derivative T7-expression vector, they were able to produce the *E. coli*  $\beta$ -galactosidase and the *Thermoactinomyces vulgaris*  $\alpha$ -amylase. However, their expression system presented two main limitations. First, *B. subtilis* BD170 is very similar to the wild type strain 168 (Dubnau *et al.* 1967) and can therefore produce numerous extracellular proteases. This last characteristic has been one of the main bottlenecks in expression and secretion of heterologous proteins in *B. subtilis* (Li *et al.* 2004). Since the publication of Conrad *et al.*, strains with reduced extracellular proteolytic activity

have been developed, such as WB800 (Wu *et al.* 2002) and Dpr8 (Kodama *et al.* 2007b). The second limitation of the system comes from the pUB110 replicon used in their expression vector. Plasmids derived from pUB110 replicate stably in *B. subtilis*, but cloned recombinant DNA can cause structural and sometimes segregational instability (Bron *et al.* 1991b). The basis of the structural instability is linked to their replication mechanism. These plasmids replicate as rolling circles by generating single-stranded DNA intermediates in which short direct repeats, introduced by the cloned DNA, can lead to deletions (Bron *et al.* 1991a). This can be particularly problematic when large DNA fragments are cloned.

The T7-expression system developed by our group present all the interesting features of the one developed by Conrad *et al.*, without the limitations. The *B. subtilis* strain WB800 used for the construction of our expression host has no detectable extracellular proteolytic activity. *B. subtilis* WB800 was developed by Wu *et al.* (2002) by inactivating the eight known extracellular proteases encoded by *nprE*, *aprE*, *epr*, *bpr*, *mpr*, *nprB*, *vpr* and *wprA*. Improperly folded secreted proteins are known to be targeted by proteases. In *B. subtilis*, even correctly folded heterologous proteins can be degraded by extracellular proteases. The strain WBT7 is therefore a more appropriate expression host than the BD170-derivative. The structural instability of the rolling circle-type plasmids has lead to the development of expression vectors derived from theta-type replicating plasmids. A series of low-copy expression vectors, with full structural stability, have already been constructed using the pBS72 replicon (Nguyen *et al.* 2005). In this study, the pBS72 replicon was used to develop the T7-expression vector pFX64. Although our vector is similar to the one developed by Conrad *et al.*, no structural instability is expected. Also, pFX64 features a  $\lambda$  *cos* sequence that allows cloning of large DNA fragments.

Compared to *E. coli*, *B. subtilis* requires a stronger complementarity between ribosomes and Shine-Dalgarno sequences for an efficient initiation of the translation (Vellanoweth et Rabinowitz 1992). Also, it has been reported that *B. subtilis* inefficiently translated T7 late mRNA, possibly because of the RBS (Hager et Rabinowitz 1985a). To assess the new *B. subtilis* T7-expression system,

*amyE* was used as a model gene. Expression was evaluated with the native RBS of *amyE* and that of the T7 *gene 10* originally present in pET-9a. Agar plate assay done with culture supernatants showed that AmyE was produced under induced-condition, demonstrating the functionality of *B. subtilis* WBT7 and pFX64. The absence of activity when using the T7 *gene 10* RBS is in accordance with the literature. Although this RBS is really effective in *E. coli*, it seems to be inefficient in initiating translation in *B. subtilis*. For future use, the RBS present in pFX64 should be replaced by one that is functional in *B. subtilis* and *E. coli*, to have a true bifunctional T7-expression vector.

The agar plate assay showed that the *B. subtilis* T7-expression system was functional. The strain WBT7 allowed inducible production of proteins without detectable expression leaking. Conrad *et al.* (1996) also reported tightly regulated expression using the same promoter. In the absence of inducer, *xylAp* is known to be strongly repressed by XylR (Kim *et al.* 1996). Also, glucose exerts catabolite repression and acts as an anti-inducer for XylR by competing with xylose (Rygu et Hillen 1992; Dahl *et al.* 1995). Since the *cis*-acting element (CRE) is not present on pFXT7, expression should not be subjected to catabolite repression. Tightly regulated systems, like ours, are of particular interest for the production of toxic proteins and allow turning expression “on” when cells reach high density or at a particular time point.

Although pFX64 and *B. subtilis* WBT7 allowed inducible T7 RNAPol-directed expression, the amount of AmyE secreted was relatively low compared to what was reported in another study (Leloup *et al.* 1997)Q32. A first bottleneck could result from the quantity of T7 RNAPol produced by the host. Although small amounts of T7 RNAPol are required to generate very large quantity of mRNA (Studier et Moffatt 1986), insufficient production would directly impair expression of the cloned gene. Western blot failed to detect the T7 RNAPol in the cytosolic protein fraction of *B. subtilis* WBT7 under induced condition. The low production of T7 RNAPol could be explained by the native phage RBS used in pAXT7. Replacing the RBS upstream of the T7 RNAPol gene by a more efficient one would possibly increase the productivity of the system. The low yield of

secreted AmyE could also result of misfolding or burden of the secretion machinery. Under overproduction conditions, proteins tend to aggregate as inclusion bodies (Mogk *et al.* 2002). One approach to enhance proper folding is to increase the production of molecular chaperones (Wu *et al.* 1998). Constitutive expression of the extracellular membrane-associated chaperone PrsA has increased the production of secreted  $\alpha$ -amylases and of a protease (Kontinen et Sarvas 1993). Other alternative methods consist of lowering the cells incubation temperature (Schein et Noteborn 1988; Thuy Le et Schumann 2007) and tuning inducer concentration (Turner *et al.* 2005). Different concentrations of D-xylose were tested, but no notable effect on the amount of secreted AmyE was observed. In their *B. subtilis* T7-expression system, Chen *et al.* (2010) showed that the use of a multicopy expression vector was unnecessary and even adverse. When using a pUB110-derived vector, the secretion pathway of *B. subtilis* seemed to be saturated. With an integrative vector, 50 % more target protein could be secreted. A similar approach could be used in an attempt to increase the productivity of our expression system. Still, further investigation is needed to determine what limits the productivity of the expression system.

Some improvement is still needed to obtain a highly efficient *B. subtilis* T7-expression system, but the strain and the vector presented here are a solid base for future work. This expression system could be a very appealing alternative to *E. coli*. The *B. subtilis* T7-expression system combines the great features of this host to the high activity of the T7 RNAPol. The  $\lambda$  *cos* sequence of pFX64 allows cloning of large DNA inserts for genomic and metagenomic libraries construction. Libraries constructed with pFX64 could be screened in *E. coli* and *B. subtilis*. Use of multiple hosts in function-based screening is known to enhance gene detection frequencies (Daniel 2005). The *B. subtilis* strain WBT7 and the shuttle vector pFX64 offer great potential for heterologous gene expression and metagenomic library construction and screening.

### **3.9 Acknowledgements**

This work was supported by a Strategic Grant from the Natural Sciences and Engineering Council of Canada. We are grateful to Raymonde Jette, Amélie Côté and Guillaume Brault for helpful discussions. We also thank Alexia Monges for critical review.

**PARTIE II:**  
**CONSTRUCTION ET CRIBLAGE D'UNE BANQUE**  
**MÉTAGÉNOMIQUE**

#### **4. REVUE DE LITTÉRATURE**

## 4.1 Métagénomique

Pour l'industrie des biotechnologies, la vaste diversité microbienne représente une source incroyable de biocatalyseurs et de molécules aux propriétés diverses. Le nombre total de cellules procaryotes sur terre est en effet estimé à  $4-6 \times 10^{30}$  (Whitman *et al.* 1998). Les bactéries sont présentes dans pratiquement tous les environnements et seule une infime proportion d'entre elles ont été découvertes ou étudiées. On estime à environ 0.1% la portion des bactéries pouvant être cultivées en laboratoire avec les techniques actuelles (Torsvik *et al.* 1990; Amann *et al.* 1995a; Lorenz *et al.* 2002; Torsvik et Ovreas 2002). La grande majorité de la diversité génétique microbienne demeure donc inaccessible par les approches classiques basées sur l'isolement de bactéries en cultures pures. Pour accéder à ce matériel génétique, des approches alternatives ont été développées. La métagénomique a émergé à la fin des années 1990 comme une approche permettant d'accéder et d'étudier le matériel génétique d'une communauté microbienne présente dans un environnement donné (Handelsman *et al.* 1998). Dans cette approche, l'ADN génomique est isolé directement d'un consortium bactérien, cloné dans un vecteur et introduit dans un hôte hétérologue pour ainsi générer une banque métagénomique. La banque peut ensuite être analysée par séquençage ou criblée selon des approches basées sur la séquence ou l'expression de traits particuliers.

L'approche métagénomique a été utilisée avec succès pour la découverte d'une grande diversité d'enzymes (Tableau X) (Streit *et al.* 2004; Daniel 2005; Steele *et al.* 2009). Parmi les groupes d'enzymes les plus souvent représentées, on retrouve les estérases, les lipases, les protéases et diverses enzymes impliquées dans l'hydrolyse de polysaccharides et la synthèse de métabolites secondaires. D'une étude à l'autre, le degré de caractérisation des enzymes découvertes varie grandement, allant de la simple séquence peptidique à la complète caractérisation biochimique.

**Tableau X.** Enzymes découvertes grâce à l'approche métagénomique.

Enzyme	Source d'ADN	Type de vecteur	Nombre de clones	Référence(s)
Lipases	Sources variées; enrichissement en fermenteur avec de la gélatine	Cosmide	10000	(Meilleur <i>et al.</i> 2009; Côté et Shareck 2010)
	Sédiments aquatiques	Fosmide	2400	(Couto <i>et al.</i> 2010)
	Sol	Plasmide	Non spécifié	(Wei <i>et al.</i> 2009)
	Rumen bovin	BAC	15360	(Liu <i>et al.</i> 2009)
	Sédiments marins	Fosmide	> 7000	(Hardeman et Sjoling 2007)
Estérases	Sol	Fosmide	15000	(Buntermgsook <i>et al.</i> 2010)
	Sol	Fosmide	142900	(Lee <i>et al.</i> 2010)
	Sédiments marins	Fosmide	Non spécifié	(Hu <i>et al.</i> 2010)
	Compost	Fosmide	23400	(Kim <i>et al.</i> 2010)
Protéases	Boues activées	Fosmide	Non spécifié	(Zhang et Han 2009)
	Sol	Plasmide	80000	(Waschkowitz <i>et al.</i> 2009)
	Sédiments marins	Fosmide	30000	(Lee <i>et al.</i> 2007)
β-galactosidase	Sol	Plasmide	12000	(Wang <i>et al.</i> 2010a)
Xylanase	Sol; enrichissement avec de la paille de riz	Cosmide	12000	(Mo <i>et al.</i> 2010)
	Sol	Plasmide	24000	(Hu <i>et al.</i> 2008)
Amylase	Eau usée agricole	Phagemide	5000000	(Lee <i>et al.</i> 2006b)
	Sol	Cosmide	35000	(Sharma <i>et al.</i> 2010)
	Rumen bovin	Cosmide	15000	(Duan <i>et al.</i> 2009)
Cellulase	Compost	Cosmide	100000	(Pang <i>et al.</i> 2009)
	Sol	Fosmide	70000	(Kim <i>et al.</i> 2008)
	Sol	Cosmide	1532	(Voget <i>et al.</i> 2003)

## 4.2 Construction de banques métagénomiques

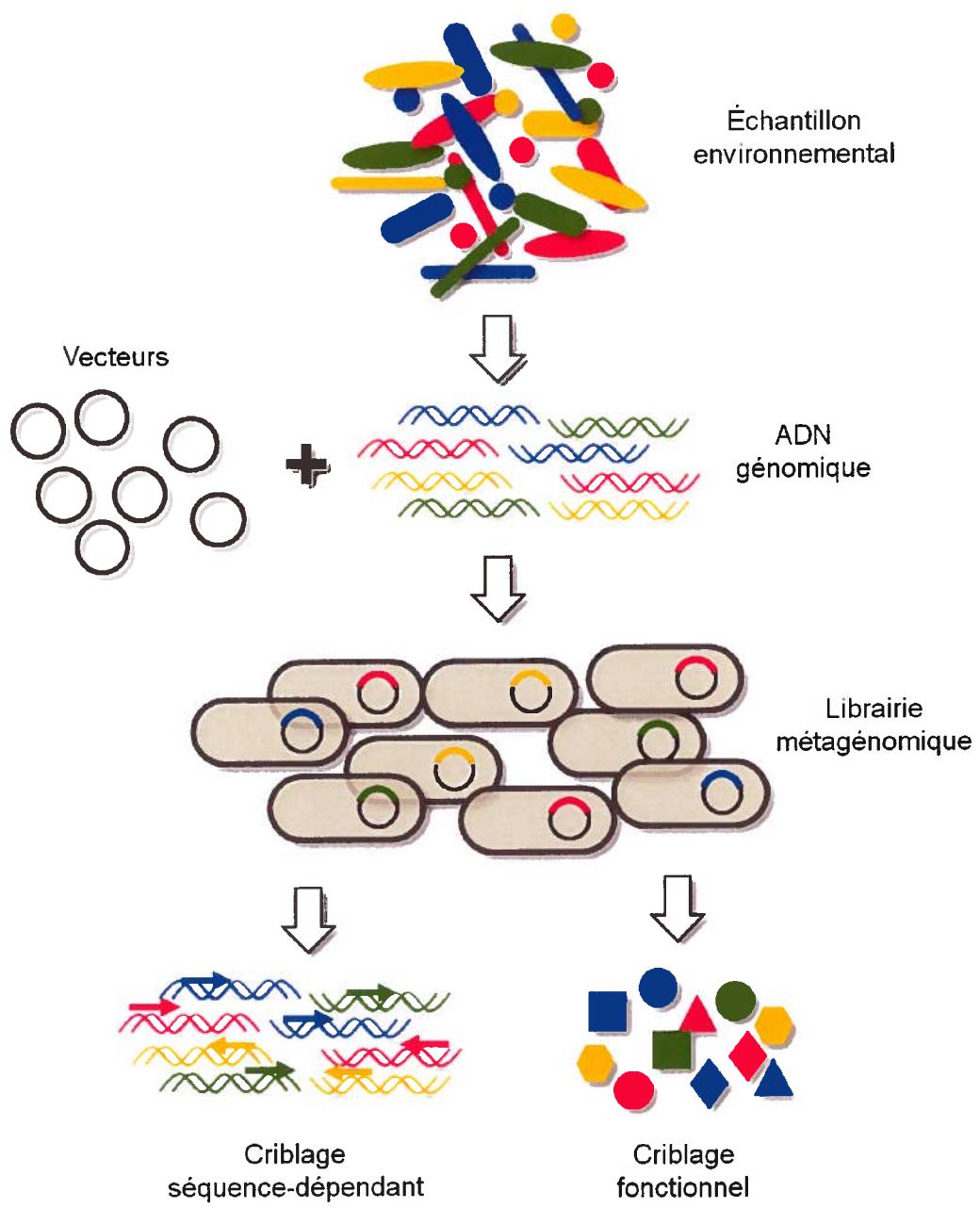
La construction d'une banque métagénomique débute par l'extraction d'ADN génomique à partir d'un échantillon complexe contenant une biomasse bactérienne. L'ADN est ensuite fragmenté à l'aide d'enzymes de restriction ou de moyens mécaniques pour générer des fragments pouvant varier entre 5 et 40 kb. Selon la longueur des fragments d'ADN, différents types de vecteurs peuvent être utilisés. L'ADN génomique lié au vecteur est finalement introduit dans un hôte bactérien hétérologue pour y être conservé et/ou criblé (Figure 16).

### 4.2.1 Source d'ADN

Le choix de la source d'ADN influence l'issue d'une étude métagénomique en déterminant la diversité génétique et la fréquence des gènes codant pour l'activité recherchée. Trois grandes approches peuvent être dégagées des études métagénomiques publiées. La première approche consiste à extraire l'ADN d'un environnement naturellement riche en organismes produisant le type d'enzyme recherchée, tel que l'intestin d'insectes phytophages pour la découverte de xylanases (Brennan *et al.* 2004). La seconde approche utilise des environnements comportant une grande diversité microbienne, tel que le sol. Il est estimé qu'un gramme de sol provenant d'un simple champ cultivé contient  $2 \times 10^9$  cellules procaryotes (Paul et Clark 1989). Selon différentes études, un gramme de sol contiendrait entre 2000 et 18000 génomes distincts (Torsvik *et al.* 1996; Torsvik *et al.* 1998; Doolittle 1999; Torsvik et Ovreas 2002). De ce fait, un grand nombre de banques métagénomiques sont construites à partir d'ADN génomique extrait du sol (Tableau X). Une troisième approche cible des environnements extrêmes, tel qu'hypersalin, alcalin, chaud ou froid, avec pour but d'isoler des enzymes actives et stables dans ces conditions particulières (Ferrer *et al.* 2005; Cieslinski *et al.* 2009; Heath *et al.* 2009; Jeon *et al.* 2009; Grant et Heaphy 2010).

Dans l'échantillon utilisé pour extraire l'ADN génomique, la fréquence des organismes possédant l'activité recherchée détermine le nombre de clones

devant être criblés pour isoler un gène d'intérêt (Daniel 2005). Une stratégie pour augmenter cette fréquence consiste à soumettre la biomasse à des conditions de culture favorisant les organismes comportant les caractéristiques souhaitées, tel que le pH, la température, la source de carbone et d'azote (Knietsch *et al.* 2003; Rees *et al.* 2003; Daniel 2004; Gabor *et al.* 2004b; Meilleur *et al.* 2009). L'étape d'enrichissement doit toutefois être conduite avec précaution car elle peut mener à une perte de diversité et réduire la probabilité de découverte de nouvelles enzymes (Elend *et al.* 2006).



**Figure 16.** Étapes générales d'une étude métagénomique.

#### **4.2.2 Isolement d'ADN**

L'isolement de l'ADN génomique est une étape critique dans la construction d'une banque métagénomique. Elle détermine en effet la qualité de la banque en définissant la diversité des organismes qui y sont représentés et le poids moléculaire de l'ADN. L'extraction d'ADN génomique, que ce soit à partir d'une culture pure ou d'un échantillon environnemental, requiert que les cellules soient efficacement lysées. À la différence des cultures pures, les échantillons environnementaux nécessitent un protocole de lyse adapté à plusieurs types de bactéries. Aux fins de construction de banques métagénomiques, plusieurs méthodes ont été développées pour l'extraction d'ADN à partir de divers échantillons environnementaux (Zhou *et al.* 1996; Krsek et Wellington 1999; Maarit Niemi *et al.* 2001; Roose-Amsaleg *et al.* 2001; Bey *et al.* 2010). L'extraction d'ADN métagénomique doit être un compromis entre une lyse vigoureuse, nécessaire pour que la majorité des génomes soient représentés, et la minimisation de la fragmentation de l'ADN. Les méthodes mécaniques sont connues pour leur grande efficacité de lyse et permettent généralement d'obtenir la plus grande diversité génétique (Maarit Niemi *et al.* 2001). Les méthodes de lyse chimique sont elles plus douces et sont mieux adaptées à l'isolement d'ADN génomique de hauts poids moléculaires (Bertrand *et al.* 2005; Cowan *et al.* 2005).

#### **4.2.3 Types de banques**

Les banques métagénomiques peuvent être classées en deux grands groupes selon la taille des fragments d'ADN clonés, les banques à petits et à grands inserts. Les banques à petits inserts (< 15 kb) sont basées sur des vecteurs plasmidiques, tels que pBluescript SK+ et pUC19 (Henne *et al.* 1999; Ranjan *et al.* 2005). Dans le cas des banques à grands inserts, des cosmides ou des fosmides peuvent être utilisés pour des fragments d'ADN allant jusqu'à 40 kb et des vecteurs BAC (chromosome artificiel bactérien) pour les plus grands inserts (Rondon *et al.* 2000; Courtois *et al.* 2003; Kim *et al.* 2006). Le choix du vecteur utilisé pour la

construction d'une banque métagénomique est donc influencé par le poids moléculaire de l'ADN extrait. De par la taille de leurs inserts, les banques plasmidiques sont adéquates pour l'isolement de gènes uniques ou d'opérons. Les banques à grands inserts permettent d'isoler de plus importants groupes de gènes codant pour des voies métaboliques et facilitent la caractérisation des génomes présents dans l'échantillon environnemental (Rondon *et al.* 2000; Cowan *et al.* 2005). Un des avantages des banques plasmidiques est leur nombre élevé de copies par cellule qui permet la détection de gènes faiblement exprimés. Aussi, la petite taille des fragments rend possible la transcription par le promoteur du vecteur. Par contre, de par la taille des inserts, les banques plasmidiques requiert le criblage d'environ 10 fois plus de clones que les banques à grands inserts pour couvrir la même longueur d'ADN (Steele *et al.* 2009).

## 4.3 Criblage

Différentes approches ont été développées pour accéder à l'information génétique contenue dans une banque métagénomique, soit le criblage basé sur la séquence et celui dépendant de l'expression génétique. Pour être efficace, la méthode de criblage utilisée doit permettre un haut débit d'analyse et être suffisamment sensible.

### 4.3.1 Criblage dépendant de la séquence

Dans la recherche de nouvelles enzymes, l'approche de criblage séquence-dépendant a longtemps reposé sur la séquence de gènes déjà connus pour le design d'oligonucléotides servant de sondes d'hybridation ou utilisés pour l'amplification par PCR (Schloss et Handelsman 2003). Dans les deux cas, cette approche de criblage se limite à la détection de gènes similaires à ceux déjà connus (Cottrell *et al.* 2000). Néanmoins, l'approche de criblage basée sur la séquence présente l'avantage de ne pas dépendre de l'expression des gènes chez l'hôte hétérologue. Moins utilisée que le criblage fonctionnel, cette méthode a tout

de même permis la découverte de nouvelles enzymes, telles que des lipases, des chitinases et des amylases (Cottrell *et al.* 2000; Bell *et al.* 2002; Jiang *et al.* 2006; Tang *et al.* 2006).

Au cours des dix dernières années, les capacités de séquençage se sont rapidement accrues et il est maintenant possible de séquencer un génome complet à un coût raisonnable. Les nouvelles plateformes de séquençage, telles que le Genome Sequencer (Roche 454 Life Sciences), le HiSeq 2000 (Illumina), le SOLiD System (Applied Biosystems) et le HeliScope (Helicos), permettent de séquencer des gigabases d'ADN en quelques jours (Pettersson *et al.* 2009; Voelkerding *et al.* 2009). Ces technologies sont principalement utilisées dans des études métagénomiques visant à caractériser des communautés microbiennes (Park *et al.* 2010; Weber *et al.* 2010; Xie *et al.* 2010; Youssef *et al.* 2010) et plus rarement pour la découverte d'enzymes (Hess *et al.* 2011).

#### **4.3.2 Criblage fonctionnel**

L'approche de criblage la plus utilisée pour la découverte d'enzymes repose sur la détection de l'activité catalytique (Meilleur *et al.* 2009; Waschkowitz *et al.* 2009; Buntergsook *et al.* 2010; Côté et Shareck 2010; Couto *et al.* 2010; Mo *et al.* 2010; Sharma *et al.* 2010). Une fois la banque métagénomique construite et introduite dans l'hôte hétérologue, le criblage fonctionnel s'effectue le plus souvent sur un milieu agar contenant un substrat permettant de visualiser l'activité recherchée. L'activité enzymatique peut se traduire par un changement d'apparence du substrat, tel la formation d'une zone claire ou turbide. La détection de lipases et d'estérases, par exemple, s'effectue souvent à l'aide de tributyrine, dont l'hydrolyse se traduit par la formation d'une zone claire autour de la colonie (Thomson *et al.* 1999). D'autres approches utilisent des substrats fluorogéniques ou emploient des colorants (Kouker et Jaeger 1987; Kim *et al.* 2008). Le criblage fonctionnel ne dépend pas de la séquence de gènes connus et peut donc mener à la découverte de classes d'enzymes complètement nouvelles. Cette approche présente aussi l'avantage de sélectionner des gènes fonctionnels et

complets. Par contre, le succès du criblage fonctionnel repose entièrement sur l'expression adéquate des gènes chez l'hôte hétérologue, ce qui est en fait la principale limitation de cette approche. À partir de l'analyse de 32 génomes procaryotes, il est estimé que 40 % des gènes peuvent être exprimés chez *E. coli* (Gabor *et al.* 2004a). Même quand les gènes sont exprimés, la quantité d'enzyme produite peut être trop faible pour que l'activité soit détectée. Le choix de l'hôte utilisé peut donc influencer grandement les résultats du criblage.

#### 4.4 Hôtes hétérologues

Comme pour l'expression de gènes hétérologues, *E. coli* est l'hôte le plus couramment utilisé dans les études métagénomiques. Dans la plupart des publications, *E. coli* est utilisée pour la construction et le maintien de la banque métagénomique, ainsi que pour le criblage (Daniel 2005). Ceci peut entre autres s'expliquer par la disponibilité d'un grand nombre de vecteurs, plasmides, cosmides, fosmides et BAC. Bien qu'*E. coli* soit utilisée avec succès dans bon nombre d'études, il a été démontré que l'utilisation d'hôtes alternatifs permettait d'élargir la diversité des enzymes détectées par criblage fonctionnel (Li *et al.* 2005; Wang *et al.* 2006). À cette fin, divers vecteurs navettes ont été développés, permettant le transfert de banques produites chez *E. coli* vers d'autres hôtes bactériens (Martinez *et al.* 2004; Aakvik *et al.* 2009; Lussier *et al.* 2010b). Les hôtes de criblage alternatifs décrits dans la littérature comprennent *Pseudomonas putida*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Rhizobium leguminosarum*, *Pseudomonas fluorescens* et *Xanthomonas campestris* (Wang *et al.* 2000; Martinez *et al.* 2004; Li *et al.* 2005; Wexler *et al.* 2005; Aakvik *et al.* 2009). Les bactéries possédant une grande capacité de sécrétion sont d'un intérêt particulier, les approches de criblage fonctionnel requérant généralement que les protéines soient produites dans le milieu de culture. Des hôtes eucaryotes pourraient aussi être utiles pour le criblage fonctionnel, les banques métagénomiques pouvant contenir une grande proportion d'ADN eucaryote (Daniel 2005). Par contre,

l'inefficacité de transfert des banques représente un frein réel à l'utilisation d'hôtes alternatifs (Martinez *et al.* 2004).

## **5. PRÉSENTATION DE L'ARTICLE 3**

## **5.1 Mise en contexte de l'article 3**

La métagénomique est une approche permettant d'accéder à la diversité génétique d'une population de microorganismes présents dans un environnement donné. Dans un grand nombre de publications, cette approche a démontré son efficacité pour la découverte de nouvelles enzymes. Dans le cadre d'une étude visant à identifier de nouvelles enzymes d'intérêt industriel, le vecteur pFX583 et la souche *S. lividans* 10T7 ont été utilisés pour la construction et le criblage d'une banque métagénomique. Cet article présente une des applications du système d'expression T7 développé chez *S. lividans* et a été accepté pour publication dans la revue *Journal of Industrial Microbiology and Biotechnology* le 11 novembre 2010.

## **5.2 Contribution de l'étudiant à l'article 3**

Les travaux présentés dans cet article ont principalement été réalisés par l'étudiant sous la supervision de François Shareck (Ph.D.). L'enrichissement de la biomasse en fermenteur a été réalisé par Jean-François Hupé, sous la supervision de Réjean Beaudet (Ph.D.) et de Pierre Juteau (Ph.D.). Les essais enzymatiques ont été réalisés par Olivier Chambenoit (Ph.D.) avec l'assistance technique de Julie Dorion-Thibaudeau. L'article a entièrement été écrit par l'étudiant et révisé par les co-auteurs.

### **5.3 Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector**

François-Xavier Lussier\*, Olivier Chambenoit, Amélie Côté, Jean-François Hupé,  
François Denis, Pierre Juteau, Réjean Beaudet and François Shareck

INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval,  
Québec, Canada H7V 1B7

\* Corresponding author. Mailing address: INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval, Québec, Canada H7V 1B7. Phone: (450) 687-5010, ext. 4274. Fax: (450) 686-5501. E-mail: francois-xavier.lussier@iaf.inrs.ca

Keywords: Metagenomic, functional screening, pFX583, biomass enrichment.

**Référence:** Lussier, F.X., Chambenoit, O., Côté, A., Hupé, J.F., Denis, F., Juteau, P., Beaudet, R. et Shareck, F. (2010) "Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector." J Ind Microbiol Biotechnol DOI: 10.1007/s10295-010-0915-2.

## 5.4 Résumé

L'approche métagénomique a grandement accéléré la découverte de nouvelles enzymes en donnant accès au potentiel génétique de microorganismes provenant de divers environnements. Le criblage fonctionnel de banques métagénomiques dépend de l'expression adéquate des gènes dans l'hôte hétérologue. Ceci peut être particulièrement problématique dans le cas de banques comportant de grands fragments d'ADN. Dans cette étude, le vecteur cosmidique navette pFX583 a été utilisé pour la construction et le criblage d'une banque métagénomique. Ce vecteur permet d'utiliser l'ARN polymérase T7 pour transcrire spécifiquement l'insert d'ADN, chez *Escherichia coli* et *Streptomyces lividans*. L'ADN utilisé pour la construction de la banque a été obtenu à partir d'une biomasse enrichie. La banque a été criblée pour des activités lipolytiques et protéolytiques en utilisant *E. coli* et *S. lividans* comme hôtes. De nombreux clones possédant une activité lipolytique ont été isolés chez *E. coli*. Par contre, aucune protéase n'a pu être détectée chez les deux hôtes. Le criblage fonctionnel a néanmoins permis l'isolement d'un gène codant pour une nouvelle lipase qui a pu être partiellement caractérisée.

## 5.5 Abstract

The metagenomic approach has greatly accelerated the discovery of new enzymes by giving access to the genetic potential of microorganisms from various environments. Function-based screening depends on adequate expression of the foreign genes in the heterologous host, which can be challenging in large-insert libraries. In this study, the shuttle cosmid vector pFX583 was used for the construction and screening of a metagenomic library. This vector allows T7 RNA polymerase-directed transcription of the cloned DNA and can be used in *Escherichia coli* and *Streptomyces lividans*. The DNA used for the library construction was obtained from an enriched biomass. The library was screened for

lipolytic and proteolytic activities using *E. coli* and *S. lividans* as hosts. Numerous *E. coli* clones with lipolytic activity were detected. Unfortunately, proteases could not be detected in both hosts. From the lipolytic activity screen, a gene coding for a new lipase was isolated, and partial characterization was conducted.

## 5.6 Introduction

Considering that nearly 99 % of the microorganisms in nature are not readily culturable (Amann *et al.* 1995b), a large genetic reservoir remains untapped despite years of culture-dependent screening studies. To gain access to this genetic information, methods were developed based on the analysis and direct isolation of nucleic acids from uncultured microorganisms. Among those methods, metagenomics, the analysis of the collective microbial genomes present in a given habitat, has emerged as a powerful approach (Steele *et al.* 2009). Metagenomics usually involves direct isolation of genomic DNA from an environment, construction of a library by cloning the DNA into a suitable vector and subsequent high-throughput sequencing or screening. Screens of metagenomic libraries can be based either on sequence- or function-driven approaches. Sequence-based screening relies on the use of a conserved DNA sequence for designing hybridization probes or PCR primers to detect specific sequences (Schloss et Handelsman 2003). The main advantage of sequence-driven screening methods is their independence toward expression of the cloned gene by the heterologous host. However it tends to recover sequences related to known genes and does not select for complete gene sequences and functional products (Daniel 2005). Function-driven analysis depends on the detection of a specific phenotype expressed by a clone. In order to be successful, it requires the transcription of the cloned genes and an efficient translation by the heterologous screening host, which are the main limitation of the function-driven approach (Handelsman 2004). Expression can be driven by the vector's host-specific promoter in plasmid libraries, but limited in cosmid, fosmid or BAC libraries due to the size of the DNA insert. Nevertheless, those large-insert libraries cover

longer DNA sequences with fewer clones and are more appropriate to recover complex pathways encoded by large gene clusters.

The detection of a target gene in a metagenomic library often involves screening many thousands of clones. To increase the chances of finding positive clones, it is possible to generate a laboratory enrichment culture from an environmental sample, prior to DNA isolation. By using particular growth conditions, it is possible to increase the proportion of organisms harboring the target trait (Entcheva *et al.* 2001; Gabor *et al.* 2004b; Ferrer *et al.* 2005). However, it must be balanced against the overall loss of population diversity (Elend *et al.* 2006).

The *Escherichia coli* T7-expression system developed by Studier and Moffatt (1986) is commonly used to achieve high-level protein production. The system is based on the T7 bacteriophage RNA polymerase (T7 RNAPol), which directs the selective transcription of genes cloned downstream of the major T7 late promoter. The T7RNAPol is characterized by a very high activity, elongating mRNA chains about five times faster than the *E. coli* RNA polymerase (Golomb et Chamberlin 1974). The phage RNA polymerase can also generate very long mRNAs and is poorly terminated by unrelated transcription terminators (McAllister *et al.* 1981). Since the original publication of Studier and Moffatt, the T7-expression system has been adapted to mammalian cells and several bacteria (Elroy-Stein et Moss 1990; Conrad *et al.* 1996; Barnard *et al.* 2004; Drepper *et al.* 2005; Kang *et al.* 2007; Gamer *et al.* 2009; Katzke *et al.* 2010). Lussier *et al.* (2010b) have developed a bifunctional vector, pFX583, that allows T7 RNAPol-directed transcription in *E. coli* and in the filamentous Gram-positive bacterium *Streptomyces lividans*. As pFX583 features a  $\lambda$  cos sequence, it can be used as a cosmid vector for the cloning of large DNA fragments. The different characteristics of pFX583 make it very attractive for the construction of metagenomic libraries and for function-driven screening. To use with this vector, a *S. lividans* strain that inducibly produces T7 RNAPol was also developed by the same group. *S. lividans* is known as a very useful alternative host to expand the

number of genes detected in metagenomic function-based screening (Martinez *et al.* 2004).

Here, in order to isolate new lipolytic and proteolytic enzymes, a metagenomic cosmid library was constructed with pFX583 using DNA extracted from the biomass of an enriched fed-batch reactor. The resulting library was screened for enzymatic activity in *E. coli* and *S. lividans*. The approach described in this paper combined biomass enrichment and multi-host screening, strategies known to increase gene detection frequency (Daniel 2005). The use of pFX583 has allowed T7 RNAPol-directed transcription of the cloned metagenomic DNA fragments, which has potentially enhanced the expression of the foreign genes. Many clones with lipolytic activity were detected, from which a new lipase was isolated and partially characterized. For unknown reason, no protease was detected.

## 5.7 Materials and Methods

### 5.7.1 Bacterial strains, culture media and growth conditions

Bacterial strains and vectors used in this study are listed in Table 1. *E. coli* strains were grown in 2xTY medium (16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl) with or without agar at 37 °C. Screening was performed on 2xTY agar supplemented with 1% tributyrin (Sigma) or 2.5 % skim milk (Difco). *S. lividans* was grown at 34 °C and R5 medium (Kieser *et al.* 2000) supplemented with 2.5 % skim milk was used for protoplast regeneration and for direct screening of proteolytic activity. Antibiotics were added to the growth media at the following concentrations: kanamycin at 50 µg/ml or 200 µg/ml, thiostrepton at 50 µg/ml (solid medium) or 5 µg/ml (liquid medium) and apramycin at 50 µg/ml. The culture medium used for the biomass enrichment was an adaptation of the Basal Salts Medium (Reddy *et al.* 2007) supplemented with ground meat extract.

**Table XI.** Strains and plasmids used in this work.

Strain or plasmid	Genotype and description	Source or reference
<b>Strain</b>		
<i>E. coli</i> RosettaBlue(DE3)/pLysS	<i>endA1 hsdR17(r<sub>K12</sub><sup>-</sup> m<sub>K12</sub><sup>+</sup>) supE44 thi-1 recA1 gyrA96 relA1 lac</i> (DE3) [F' <i>proA</i> <sup>+</sup> <i>B</i> <sup>+</sup> <i>lacI</i> <sup>q</sup> <i>ZΔM15</i> ::Tn10] pRA RE (Cam <sup>R</sup> , Tet <sup>R</sup> )	Novagen
<i>E. coli</i> TOP10	<i>F</i> ' <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) φ80 <i>lacZΔM15</i> Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara-leu</i> ) 7697 <i>galU</i> <i>galK</i> <i>rpsL</i> ( <i>Str</i> <sup>R</sup> ) <i>endA1</i> <i>nupG</i> λ-	Invitrogen
<i>S. lividans</i> 10T7	<i>msiK</i> <sup>-</sup> , T7 RNApol codon-optimized gene under the control of <i>tipAp</i> (Ap <sup>R</sup> , Th <sup>R</sup> )	(Lussier <i>et al.</i> 2010b)
<b>Plasmid</b>		
pFX583	Km <sup>R</sup> ; <i>ori/rep</i> pJV1, <i>ori</i> pMB1, T7 promoter and terminator, λ <i>cos</i> , RK2 <i>oriT</i> .	(Lussier <i>et al.</i> 2010b)
pCR4-TOPO	Am <sup>R</sup> / Km <sup>R</sup> ; <i>ori</i> pUC, <i>lac</i> promoter	Invitrogen

Cam<sup>R</sup>, Tet<sup>R</sup>, Ap<sup>R</sup>, Th<sup>R</sup>, Km<sup>R</sup>, Am<sup>R</sup>: resistance genes for chloramphenicol, tetracycline, apramycin, thiostrepton, kanamycin and ampicillin, respectively.

### 5.7.2 Sequencing Fed-Batch Reactor enrichment

A 10 litre Sequencing Fed-Batch Reactor (SFBR) was inoculated with 10 g of black soil, 10 g of shrimp compost, 50 ml of biomass from a SFBR for the phosphor elimination in swine waste, 50 ml of biomass from an Aerobic Thermophilic Sequencing Batch Reactor for swine waste treatment (Juteau *et al.* 2004) and two swabs from the cafeteria of the INRS-Institut Armand-Frappier. The biomass of the SFBR was enriched with ground meat extract solution as a source of carbon. The solution was prepared with extra lean ground beef composed of 10 % of lipids, 20 % of proteins and 0 % of carbohydrates, based on the product label. In 1 l of ion-free water, 100 g of meat were homogenized and the pH was adjusted to 7.0 with NaOH. The resulting solution was filtered on 850

$\mu\text{m}$  sieve (20 mesh). The SFBR biomass was submitted to 30 cycles of 72 h in which the pH gradually shifted from 7 to 8.5 and back to 7. Simultaneously, the temperature increased from 50°C to 70°C and returned to 50°C.

### **5.7.3 Library construction**

The DNA used for the construction of the metagenomic library was obtained from the biomass of the SFBR. The method used to extract high-molecular-weight DNA included lysozyme digestion, freeze/thaw cycles and phenol/chloroform extractions. Total DNA was partially digested with *Bam*HI in order to generate fragments of 35-45 kb, and ligated with pFX583 linearized with the same restriction enzyme. The concatemeric DNA was packaged using the MaxPlax<sup>TM</sup> Lambda Packaging Extracts (Epicentre Biotechnologies) and transduced into *E. coli* RosettaBlue (DE3)/pLysS (Novagen).

### **5.7.4 Functional-based screening and gene isolation**

Transductants were replica-plated on 2xTY agar supplemented with tributyrin or skim milk and 0.4 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). After 48 h of incubation at 37 °C, the colonies showing a zone of clearance indicated lipolytic or esterolytic activities. The cosmids from the positive colonies were extracted for analysis. To isolate the putative lipase or esterase encoding genes, a subcloned library was generated in the plasmid pCR4-TOPO using the TOPO Shotgun Subcloning Kit (Invitrogen). The subcloned library was rescreened in *E. coli* TOP10 (Invitrogen) and the plasmids from positive colonies were sequenced at the McGill University and Génome Québec Innovation Centre.

### **5.7.5 Partial enzymatic characterization**

Enzymatic assays were done with supernatant from an overnight-grown culture of the *E. coli* TOP10 clone containing *lipF5-11*, which was obtained from the

subcloned library. A clone with an empty plasmid was used as a negative control. Assays were performed at 50 °C in a 96-well microplate with 5 µl of supernatant in a final volume of 100 µl. Reaction mix contained 1mM of a particular p-nitrophenyl (pNP) acyl ester (Sigma) in 50 mM Tris-HCl pH 8, 0.05 % (m/v) CaCl<sub>2</sub>, 0.5% (v/v) Triton X-100. Absorbance was read at 405 nm after 48 min of incubation with a Fusion Microplate Reader (Packard). Substrates with different acyl chain-length were tested: pNP-acetate (C2), pNP-butyrate (C4), pNP-caproate (C6), pNP-caprylate (C8), pNP-caprate (C10), pNP-laurate (C12), pNP-myristate (C14), pNP-palmitate (C16) and p-nitrophenyl stearate (C18). The influence of pH was tested with pNP-palmitate using 50 mM Tris-HCl buffer for pH 6, 7 and 8 and with 50 mM glycine-NaOH buffer for pH 9, 10 and 11. The thermostability of LipF5-11 was evaluated by pre-incubating the culture supernatant for 2 hours at temperatures ranging between 50 °C and 70 °C. The enzymatic assay was then conducted as described above.

## 5.8 Results

### 5.8.1 Metagenomic library construction

The library was constructed using genomic DNA extracted from the biomass of a Sequencing Fed-Batch Reactor enriched for bacteria able to grow at alkaline pH and temperatures ranging between 50 and 70 °C. The genomic DNA was cloned in the vector pFX583, which featured a λ cos sequence, a T7 promoter and a kanamycin/neomycin selection marker. Because of the pMB1 and pJV1 replicons of pFX583, the library can be screened in *E. coli* and *Streptomyces* species.

### 5.8.2 Function-based screening

The packaged library was transduced into *E. coli* RosettaBlue (DE3)/pLysS, a *recA* strain harbouring the T7RNAPol encoding gene under the control of the *lacUV5* promoter. Transductants were replica-plated on 2xTY agar supplemented

with tributyrin or skim milk. From approximately 2000 screened colonies, 17 showed esterolytic/lipolytic activity, but no proteolytic activity was detected. The same number of cosmids was transferred to *Streptomyces lividans* 10T7 by protoplast transformation. Approximately 10 000 colonies were screened on R5 agar supplemented with skim milk. As for *E. coli*, no colonies showed proteolytic activity.

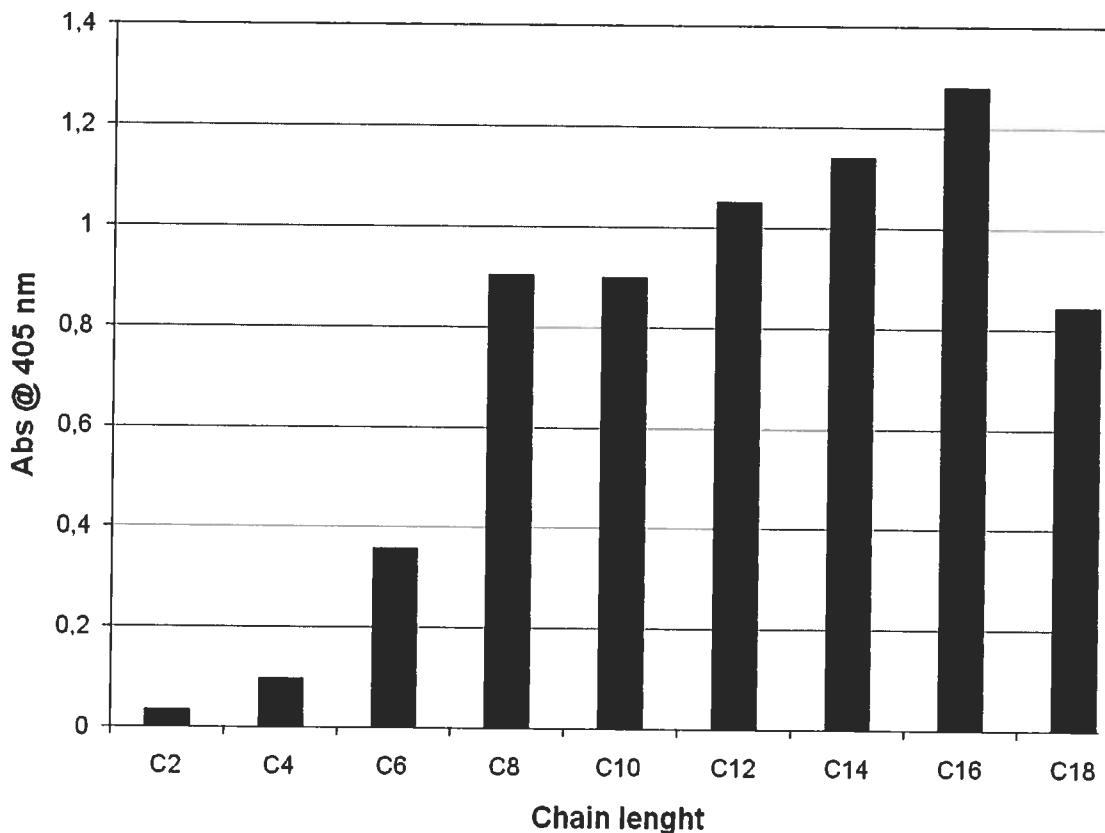
### 5.8.3 Subcloning and sequence analysis

Cosmid from positive clones were sheared by nebulization (Surzycki 2000) into 1 to 6 kb fragments and subcloned into pCR4-TOPO via topoisomerase-directed ligation. The resulting libraries were introduced in *E. coli* TOP10 and plasmids from clones showing activity were sequenced. Nucleotide sequences were analyzed by BLASTX (NCBI) using the non-redundant protein sequences database. Some of the clones contained lipase genes previously isolated from another metagenomic library constructed by our group (Meilleur *et al.* 2009). One of the plasmids contained a new secreted lipase-encoding gene (*lipF5-11*) (GenBank accession no. HQ009871) showing the highest similarity with predicted lipases from *Streptomyces pristinaespiralis* (EDY64232), *Thermomonospora curvata* (ACY99099) and *Conexibacter woeseli* (ADB52323), with identities of 42%, 39% and 39%, respectively. Downstream of *lipF5-11*, the plasmid also contained two ORFs that showed homology with PpiC-type peptidyl-prolyl cis-trans isomerase and lipase chaperone genes. The amino acid sequence of LipF5-11 featured a predicted signal peptide (residues 1-26) and the conserved pentapeptide Ala-Xaa-Ser-Xaa-Gly (residues 99-103) typical of lipases from *Bacillus* (Arpigny et Jaeger 1999) (FIG. 1). The multiple sequence alignment performed with ClustalW2 (Larkin *et al.* 2007) also revealed a putative oxyanion hole sequence (H34-G35) and the catalytic residues Asp171 and His193.

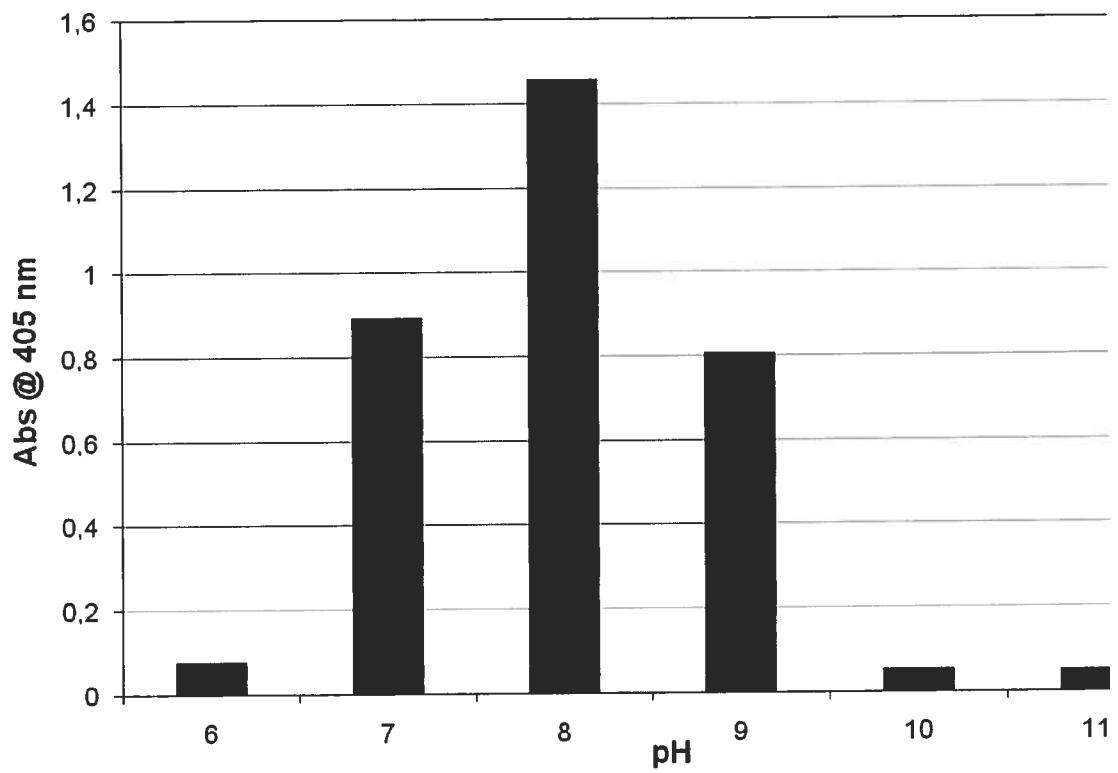
**Figure 17.** Multiple amino acid sequence alignment performed with ClustalW2. LipF5-11 was aligned with the three putative lipases that showed the highest similarity and with a lipase from *B. subtilis*. *Streptomyces pristinaespiralis* (EDY64232), *Thermomonospora curvata* (ACY99099), *Conexibacter woeseli* (ADB52323) and *Bacillus subtilis* (AAA22574.1). Lower cases indicate the predicted signal peptide; the putative oxyanion hole (HG), the penta-peptide (AHSNG) as well as the aspartic acid and the histidine as putative residues of the active site are underlined. "\*" means that the residues in that column are identical in all sequences in the alignment. ":" means that conserved substitutions have been observed. "." means that semi-conserved substitutions are observed.

#### **5.8.4 Partial biochemical characterization**

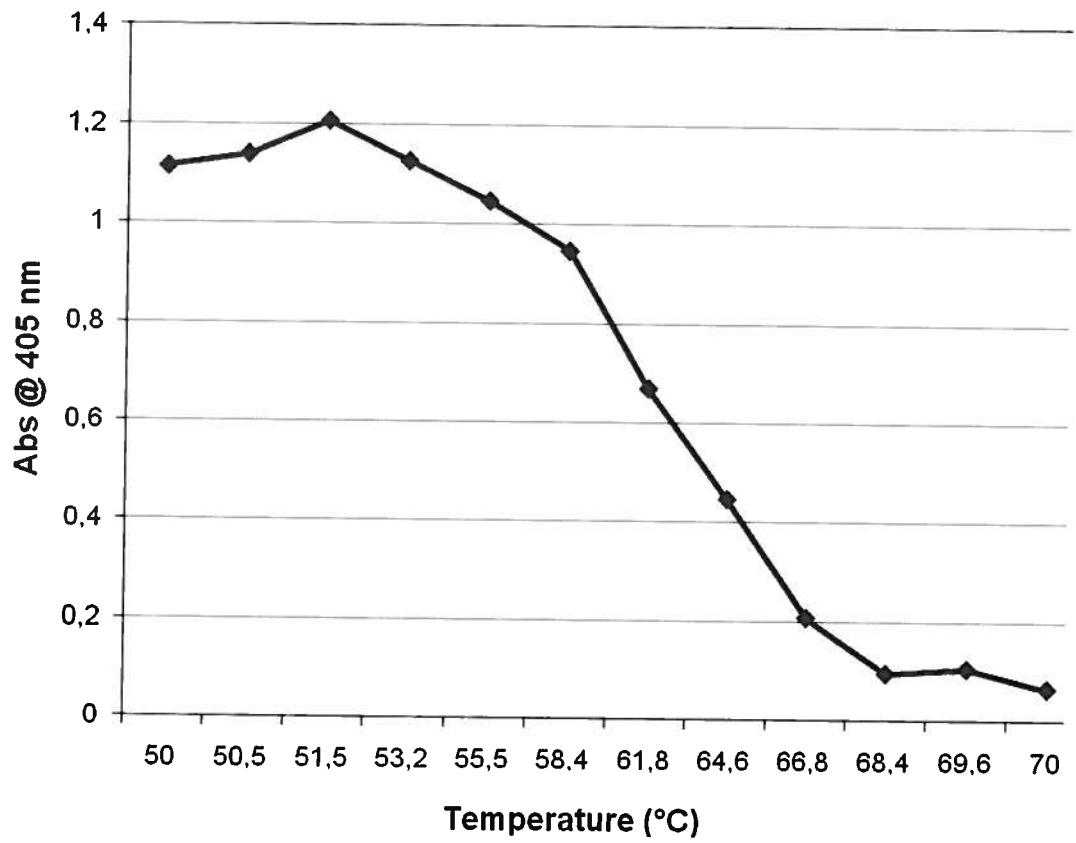
A high-throughput assay was used for the fast partial characterization of the unpurified lipolytic enzyme LipF5-11. The results presented here are representative of what was obtained in independent assays, although the global lipolytic activity of the supernatant could vary. To evaluate the substrate specificity of the enzyme, pNP-acyl esters with carbon chain length varying between 2 and 18 were tested (FIG. 2). Hydrolysis was detected for all substrates tested, but was clearly more important for the long-chain pNP-acyl esters. Optimal pH was determined by incubating the crude protein extract with pNP-palmitate (C16) at pH ranging from 6 to 11. LipF5-11 was most active between pH 6 and 9, with the highest activity at pH 8 (FIG. 3). Thermostability was evaluated by looking at the residual lipolytic activity in the supernatant after a pre-incubation of 2 hours at temperatures ranging from 50 °C to 70 °C (FIG. 4). Pre-incubation below 55 °C did not have an important effect on the activity of LipF5-11. More than 50 % of the activity remained after 2 hours at 60 °C, but dropped to 5 % at 70 °C.



**Figure 18.** Effect of the chain length on the activity of LipF5-11. Assays were conducted at 50 °C in 50 mM Tris-HCl pH 8.0. Absorbance measured at 60 min was used for comparison. pNP-acetate (C2), pNP-butyrate (C4), pNP-caproate (C6), pNP-caprylate (C8), pNP-caprate (C10), pNP-laurate (C12), pNP-myristate (C14), pNP-palmitate (C16) and p-nitrophenyl stearate (C18).



**Figure 19.** Effect of pH on the activity of LipF5-11. Lipase activity was assayed toward pNP-palmitate (C16) at 50 °C in 50 mM Tris-HCl buffer for pH 6, 7 and 8 and in 50 mM glycine-NaOH buffer for pH 9, 10 and 11.



**Figure 20.** Thermostability assay. Activity of LipF5-11 was assayed after a pre-incubation of two hours at temperatures ranging from 50 to 70 °C.

## 5.9 Discussion

In the search for novel biocatalysts or molecules, function-driven screening is the only strategy that has the potential to identify completely new genes (Daniel 2005). However, the functional approach is limited by its reliance on adequate expression of the cloned genes in a surrogate host. Aside from the codon usage of the foreign genes, the regulation of the promoters and their recognition by the host RNA polymerase can also impair the detection of positive clones. In the present study, a metagenomic library was constructed with the bifunctional cosmid vector pFX583. The T7 promoter/terminator in pFX583 allows the use of the highly effective T7 RNAPol to transcribe the cloned DNA. Combined with a T7 RNAPol producing host, pFX583 has the potential to increase the number of colonies with positive activity. A similar approach was successfully used by Leggewie *et al.* (Leggewie *et al.* 2006), using a transposon harboring bidirectional T7 promoters. Introducing this transposon in a cosmid library allowed inducible expression of its flanking regions in both directions, thus enhancing detection of clones with lipolytic activity. However, this transposon approach normally requires extraction of the cosmid library, *in vitro* transposition followed by re-introduction into the screening host. The diversity of the library with transposon is therefore dependent on the number of cosmids extracted at first. Also, because the library with transposon is reintroduced into the screening host, repetition can occur and will require the screening of an extra number of clones to be sure to cover the starting clones.

The number of screened clones required to recover genes of interest is directly linked to the size of the cloned DNA, but also to the frequency of organisms with the desired activity in the biomass from which DNA is extracted. To increase this frequency, biomass can be subjected to conditions favoring microorganisms harboring the desired traits, such as carbon or nitrogen sources, pH and temperature. However, the enrichment step can reduce the genetic diversity by promoting fast-growing and culturable members of a microbial consortia (Daniel 2005). If conducted too extensively, it can even decrease the

probability of discovering new genes (Elend *et al.* 2006). Nevertheless, well monitored enrichment combined with a diversified starting material is an efficient strategy to increase the number of positive clones in a metagenomic library screen (Knetsch *et al.* 2003; Rees *et al.* 2003; Daniel 2004; Gabor *et al.* 2004b; Meilleur *et al.* 2009). Also, the cultivation step generally facilitates the isolation of high quality DNA. In the current study, the DNA used for the construction of the metagenomic library was extracted from a biomass enriched for microorganisms capable of growing at a high temperature and a pH between 7 and 8.5. Also, inoculums were from various sources, insuring a good starting diversity.

The metagenomic library was first introduced in the common host *E. coli*. The strain used, *E. coli* RosettaBlue (DE3), is a *recA1* T7 RNAPol-producing strain. With this strategy, approximately 2000 clones were screened for esterolytic/lipolytic and proteolytic activity. From a small number of clones, compared to most metagenomic screening studies (Lorenz et Eck 2005), 17 colonies showed a cleared zone on tributyrin agar. The high success rate obtained in this functional screening experiment showed that the enrichment step was properly conducted. The use of ground beef extract was likely a good source of carbon for the selection of bacteria producing lipolytic enzymes. Also, the cosmid pFX583 used for the construction of the metagenomic library and the *E. coli* RosettaBlue (DE3)/pLysS strain allowed efficient detection of esterase/lipase enzymes, whose expression could be driven by foreign promoters and the T7 promoter of the vector. Despite the great success of the lipolytic enzymes screen, no proteolytic activities were detected for the same number of clones.

Although *E. coli* is the most employed host for functional screening (Daniel 2004), using an alternative host proved to be useful to expand the range of detected genes (Wang *et al.* 2000; Courtois *et al.* 2003; Martinez *et al.* 2004; Li *et al.* 2005; Wexler *et al.* 2005). Bacteria with a different codon usage and a higher protein secretion capacity than that of *E. coli* are of particular interest, as these characteristics can greatly influence the result of function-based screens. Because of the bifunctionality of the pFX583 vector, it was possible to screen the library in *S. lividans* 10T7. Like *E. coli* RosettaBlue(DE3), *S. lividans* 10T7 allows

T7RNApol-directed transcription (Lussier *et al.* 2010b). This Gram-positive host also has a very high secretion capacity and can produce the foreign protein directly into the culture medium (Binnie *et al.* 1997). Unfortunately, screening in *S. lividans* 10T7 gave results similar to *E. coli* and also failed to detect proteolytic activity. This disappointing result is contrasting with the high number of positive hits obtained in the lipolytic activity screen. Although proteolytic enzymes are present in all bacteria, intracellular proteases are highly specific (Bond et Butler 1987) and may not be detected on skimmed milk agar, even if the cells are lysed. On the other hand, extracellular proteases are mostly produced as inactive precursors that require to be activated by limited proteolysis (Wandersman 1989). The activation step could be suboptimal in a heterologous host and therefore affect the detection of the enzyme. Also, the skimmed milk agar method may not be sensitive enough for the detection of small amount of proteases. Still, only a small number of clones were screened compared to most metagenomic studies (Steele *et al.* 2009) and a more extensive screen could possibly allow the isolation of a protease encoding gene.

From the lipolytic activity screen in *E. coli*, a new lipase encoding gene was isolated. LipF5-11 showed only small similarity with available amino acid sequences, the highest identity being of 42 % for a predicted lipase from *Streptomyces pristinaespiralis*. The amino acid sequence of LipF5-11 has all the characteristics of a true lipase, featuring the conserved penta-peptide Ala-Xaa-Ser-Xaa-Gly typical of lipases from *Bacillus* (Arpigny et Jaeger 1999). A high-throughput microplate enzymatic assay was conducted on culture supernatants to evaluate substrate specificity, optimal pH and thermostability. LipF5-11 had the highest activity for pNP-acyl esters with long carbon chains (> 10), indicating that it is probably a lipase and not an esterase (Verger 1997; Jaeger *et al.* 1999). However, a natural substrate should be used to confirm the true substrate specificity of LipF5-11 (Sigurgísladóttir *et al.* 1993). In accordance with the conditions used for biomass enrichment, LipF5-11 showed the highest activity at pH 8. It also retained more than 50 % of its activity after a pre-incubation of two hours up to 60 °C. Downstream of *lipF5-11*, sequence analysis of the cosmid

revealed a putative PpiC-type peptidyl-prolyl cis-trans isomerase and lipase chaperone genes arranged in an operon-type structure. Although the expression of *lipF5-11* was sufficient to allow detection and partial characterization, the produced amount was still very low (result not shown). This could be due to improper folding or secretion in *E. coli*, possibly because the isomerase and the lipase chaperone were not well expressed or non-functional. For future purification and characterization, the productivity problem of LipF5-11 will have to be addressed.

In this study, biomass enrichment combined with the use of the vector pFX583 allowed a high yield of lipolytic clones. For the first time, a bifunctional cosmid vector allowing T7RNAPol-directed transcription has been used for the construction and the screening of a metagenomic library. The cosmid pFX583 has the potential to increase the transcription of the foreign genes and can be used in *E. coli* and *S. lividans*. Those two characteristics can truly enhance the gene detection frequency from metagenomic library screening. The approach used in this study led to the identification of a new lipase, a type of biocatalyst with great industrial potential.

## 5.10 Acknowledgments

This work was supported by a Strategic Grant from the Natural Sciences and Engineering Council of Canada. The authors thank Raymonde Jetté and July Dorion-Thibaudeau for their technical assistance.

## **6. DISCUSSION**

La recherche présentée dans cette thèse a été réalisée dans le cadre général d'une étude métagénomique visant à identifier de nouvelles enzymes d'intérêt industriel. Le principal objectif de ce projet de recherche consistait à développer de nouveaux hôtes d'expression bactériens, ainsi que des vecteurs multifonctionnels pouvant être utilisés dans le criblage fonctionnel de banques métagénomiques et l'expression de gènes recombinants. Afin de démontrer leur utilité, le vecteur pFX583 et la souche *S. lividans* 10T7 ont été utilisés pour la construction et le criblage d'une banque métagénomique.

## 6.1 Systèmes d'expression

Il peut être difficile de choisir rationnellement la méthode à employer pour la production d'une protéine recombinante donnée. De plus, une approche fonctionnant pour une protéine ne sera pas nécessairement adéquate pour une autre. Tout de même, basé sur l'expression de plus de 10000 gènes, le *Structural Genomics Consortium* (SGC) a suggéré une stratégie consensus qui devrait être priorisée (Graslund *et al.* 2008). Ainsi, le SGC préconise l'utilisation d'*E. coli* comme hôte d'expression, notamment la souche BL21(DE3) et ses dérivés, combiné à un vecteur permettant la transcription dirigée par l'ARNpol T7. L'approche décrite par le SGC n'est cependant pas infaillible et il peut être nécessaire d'utiliser un hôte d'expression alternatif pour obtenir une production adéquate (Graslund *et al.* 2008). C'est donc à titre de systèmes d'expression alternatifs que les souches de *S. lividans* et de *B. subtilis*, ainsi que leurs vecteurs, ont été développés.

### 6.1.1 Souches

Pour l'expression de gènes recombinants, le principal intérêt de *S. lividans* et *B. subtilis* repose sur leur grande capacité de sécrétion (Binnie *et al.* 1997; Schallmey *et al.* 2004). Contrairement à *E. coli*, les protéines sécrétées par ces hôtes Gram-positif sont dirigées directement dans le milieu de culture. La

sécrétion des protéines recombinantes présente plusieurs avantages. Cela permet d'obtenir des protéines extracellulaires sous leur forme mature, c'est-à-dire, sans leur peptide signal. Aussi, le milieu extracellulaire contient un moins grand nombre de protéines, ce qui facilite le processus de purification. Le fait de pouvoir accéder aux protéines sans devoir lyser les cellules simplifie aussi la purification. De plus, la production des protéines dans le milieu de culture permet d'effectuer des essais enzymatiques directement avec le surnageant de culture et est avantageuse pour le criblage fonctionnel de banques génomiques.

Le développement de systèmes d'expression T7 chez *S. lividans* et *B. subtilis* visait à combiner la capacité de sécrétion de ces hôtes à la grande efficacité de transcription de l'ARNpol T7. La construction d'un hôte d'expression T7 requiert bien évidemment de pouvoir produire l'ARNpol T7. Pour ce faire, le gène codant pour l'ARNpol T7 (gène *I*) peut être porté par un plasmide ou bien intégré dans le chromosome. Dans le présent projet de recherche, l'option de l'intégration chromosomique a été retenue. Le gène *I* a été intégré par recombinaison site-spécifique chez *S. lividans* et par double recombinaison homologue chez *B. subtilis*. L'intégration chromosomique permet d'assurer un nombre constant de copies du gènes par cellule et demeure plus stable que l'option plasmidique (Lomovskaya *et al.* 1980; Vazquez-Cruz *et al.* 1996). Le premier défi rencontré a été d'exprimer le gène *I* chez *S. lividans*. En effet, le gène *I* comporte quatre codons TTA<sub>leu</sub>, connu pour affecter l'expression de gènes recombinants chez cet hôte (Leskiw *et al.* 1993; Kataoka *et al.* 1999; Rebets *et al.* 2006; O'Rourke *et al.* 2009). Les quatre codons TTA<sub>leu</sub> ont donc été remplacés par le codon CTC<sub>leu</sub>, rendant possible la production d'ARNpol T7 chez *S. lividans*. Ce problème ne s'est pas présenté chez *B. subtilis* du fait que cet organisme ne comporte pas de biais dans l'utilisation des codons (Schumann 2007). Chez *S. lividans* comme chez *B. subtilis*, le gène *I* a été placé sous le contrôle de promoteurs inductibles.

### **6.1.2 Vecteurs**

L'objectif visé était de développer des vecteurs d'expression bi-fonctionnels, pouvant être utilisés à la fois chez *E. coli* et chez *S. lividans* ou *B. subtilis*. Le vecteur pFX583 est fonctionnel chez *E. coli* et *S. lividans*, tel que démontré par l'expression de *xlnA* chez les deux organismes. Dans le cas de pFX64, il peut être maintenu chez *E. coli* et *B. subtilis*, mais sa fonctionnalité n'a été démontrée que pour *B. subtilis*. Pour que pFX64 puisse être utilisé chez *E. coli* et *B. subtilis*, une séquence RBS fonctionnelle chez les deux organismes devrait être incluse dans le vecteur ou fournie par le gène cloné. Outre l'expression de gènes recombinants, les vecteurs pFX583 et pFX64 ont aussi été construits pour être utilisés dans la construction et le criblage de banques métagénomiques. À cette fin, une séquence *cos* a été introduite dans les deux vecteurs pour qu'ils puissent être utilisés comme cosmide. Très peu de vecteurs bi-fonctionnels ont été décrits dans la littérature (Sullivan *et al.* 1984; Doekel *et al.* 2002) et aucun ne présente les caractéristiques de pFX583 et pFX64.

### **6.1.3 Performances des systèmes d'expression**

En combinaison avec la souche *S. lividans* 10T7, le vecteur pFX583 a permis l'expression inductible d'un gène codant pour une xylanase extracellulaire. L'ajout de thiostreptone au milieu de culture induit la production d'ARNpol T7 qui transcrit ensuite spécifiquement le gène cloné dans le vecteur. La fonctionnalité du vecteur a aussi été testée chez *E. coli*. Ainsi, en utilisant le vecteur pFX583, il est possible d'évaluer aisément l'expression d'un gène chez deux hôtes différents. La souche *B. subtilis* WBT7 et le vecteur pFX64 offrent les mêmes possibilités. La fonctionnalité du système a été démontrée en exprimant un gène codant pour une  $\alpha$ -amylase (*amyE*). L'addition de xylose permet d'induire la production d'ARNpol T7 et de ce fait l'expression du gène cloné. La quantité d' $\alpha$ -amylase produite était tout de même relativement faible. Chez *S. lividans* comme chez *B. subtilis*, les protéines ont été produites dans le milieu extracellulaire. Ce

projet de recherche a permis de développer deux systèmes d'expression au fort potentiel et ouvre la voie à l'utilisation de l'ARNpol T7 chez *S. lividans* et *B. subtilis*.

Les performances des deux systèmes d'expression laissent cependant place à amélioration. Pour ce faire, il serait essentiel d'identifier les étapes limitantes du processus d'expression. Pour obtenir un niveau de transcription maximum, l'ARNpol T7 n'est requise qu'en faible concentration (Studier et Moffatt 1986). Cependant, une production insuffisante d'ARNpol T7 pourrait avoir un effet direct sur la quantité de protéines recombinantes produites. Il serait donc intéressant d'augmenter le nombre de copies du gène *l* dans le chromosome des souches d'expression et vérifier si la productivité du système est accrue. À l'inverse, la productivité pourrait être affectée par une surcharge de la machinerie cellulaire causée par la grande activité de synthèse de l'ARNpol T7 (Golomb et Chamberlin 1974). Il a été démontré chez *E. coli* que la surexpression de gènes peut mener à une inhibition de la croissance et par le fait même affecter la productivité du système (Dong *et al.* 1995). La surexpression de gènes recombinants augmenterait la dégradation des ARNr et diminuerait le pool de ribosomes disponibles pour la traduction des ARNm de l'hôte. La productivité des deux systèmes d'expression pourrait aussi être limitée au niveau de processus post-traductionnel. Chez *S. lividans* et *B. subtilis*, les protéines destinées à la sécrétion doivent être maintenues par des chaperonnes dans une conformation apte à la translocation (Anne et Van Mellaert 1993; Li *et al.* 2004). Si une protéine recombinante n'interagit pas adéquatement avec les chaperonnes, à cause d'un niveau de production trop élevé, elle peut former des agrégats ou être tout simplement dégradée. Dans une étude réalisée chez *B. subtilis*, l'analyse de la distribution d'une protéine recombinante a révélé que la fraction sécrétée ne représentait que 23 % de la production totale (Wu *et al.* 1993). La majorité des protéines demeuraient à l'intérieur de la cellule sous forme insoluble. En utilisant une souche produisant de façon constitutive les chaperonnes DnaK et GroE, il a été possible d'augmenter à 60 % le taux de protéines sécrétées (Wu *et al.* 1998). Une stratégie similaire pourrait être utilisée avec les systèmes de *S. lividans* et *B.*

*subtilis* pour tenter d'augmenter la quantité de protéines sécrétées. Une autre approche simple pour favoriser la sécrétion des protéines serait de réduire la température d'incubation durant la période d'induction. Avec le système d'expression T7 d'*E. coli*, l'induction à basse température (15-25 °C) diminue les risques de formation d'agrégats de protéines insolubles (Graslund *et al.* 2008). À ces températures, le métabolisme cellulaire ralenti donnerait le temps aux protéines recombinantes nouvellement formées de se replier adéquatement (Vera *et al.* 2007)

## 6.2 Construction et criblage d'une banque métagénomique

L'étude métagénomique visait à identifier de nouvelles enzymes, telles que des lipases, des protéases et diverses glycohydrolases, ayant une activité optimale à haute température et à pH neutre ou moyennement alcalin. Pour ce faire, l'approche de criblage fonctionnel a été retenue. Ce mode de criblage permet de repérer des gènes d'intérêt n'ayant aucune homologie avec des séquences connues. Le criblage fonctionnel nécessite cependant que les gènes clonés soient adéquatement exprimés dans un hôte hétérologue, ce qui est en fait la principale limitation de cette approche. Les banques métagénomiques à grands inserts sont le plus souvent construites à l'aide de fosmides ou de cosmides, et sont généralement criblées chez *E. coli* (Simon et Daniel 2009; Steele *et al.* 2009). Le choix d'utiliser *E. coli* pour le criblage fonctionnel n'est que très peu rationnel et s'explique principalement par l'absence d'alternatives. Il a pourtant été démontré que la diversification des hôtes dans le criblage fonctionnel permettait d'élargir la diversité des enzymes détectées (Li *et al.* 2005; Wang *et al.* 2006). Les souches et les vecteurs développés dans ce projet de recherche sont donc de nouveaux outils tout à fait pertinents pour la construction et le criblage de banques métagénomiques.

### 6.2.1 Vecteurs pFX583 et pFX64

Les vecteurs pFX583 et pFX64 peuvent être utilisés dans la construction de banques métagénomiques à titre de vecteurs plasmidiques ou cosmidiques. Les banques à petits et grands inserts présentent tous deux des avantages et des désavantages (Daniel 2005). Dans les deux cas, la présence d'un promoteur T7 dans les vecteurs pFX583 et pFX64 permet d'utiliser l'ARNpol T7 pour transcrire l'insert d'ADN. Grâce à sa grande activité de synthèse, l'ARNpol T7 permet d'obtenir un niveau de transcription très élevé. Ceci est particulièrement utile pour le criblage fonctionnel, car la quantité d'enzymes produites doit être suffisamment élevée pour que l'activité puisse être détectée.

Les cosmides sont avantageux de par la grandeur des fragments d'ADN qu'ils peuvent accueillir. Par contre, comme les insertions sont de grandes tailles, la transcription des gènes clonés dépend généralement de leur promoteur natif. Ces derniers ne sont pas nécessairement fonctionnels chez l'hôte hétérologue, ce qui peut bien évidemment réduire le nombre de gènes pouvant être détectés. Les vecteurs pFX583 et pFX64 offrent la possibilité d'utiliser plus d'un hôte, ce qui peut en partie régler ce problème (Wang *et al.* 2000; Courtois *et al.* 2003; Martinez *et al.* 2004; Li *et al.* 2005; Wexler *et al.* 2005). La transcription dirigée par l'ARNpol T7 peut aussi augmenter la probabilité d'expression des gènes clonés. À partir du promoteur T7 contenu dans les vecteurs, l'ARNpol T7 a la capacité de générer de très grands transcrits et est généralement peu sensible à la présence de terminateurs bactériens (McAllister *et al.* 1981).

Le vecteur pFX583 a été utilisé avec succès pour la construction et le criblage d'une banque métagénomique. Le criblage fonctionnel chez *E. coli* a permis la détection d'un grand nombre de clones possédant une activité lipolytique. De ces clones, une nouvelle lipase a été isolée et caractérisée partiellement. La banque a aussi été criblée pour des protéases chez *E. coli* et *S. lividans* sans néanmoins permettre la détection d'un clone positif. Ce dernier résultat s'explique difficilement, considérant le nombre élevé de clones détectés

dans le criblage d'enzymes lipolytiques. Des tentatives d'explication sont présentées dans la section 5.9.

### 6.2.2 Hôtes de criblage alternatifs

Bien qu'*E. coli* ait été utilisée avec succès pour le criblage fonctionnel de plusieurs banques métagénomiques, l'utilisation d'hôtes alternatifs ne peut qu'accroître la diversité de gènes détectés. Les souches *S. lividans* 10T7 et *B. subtilis* WBT7 peuvent être utilisées, respectivement, avec les vecteurs pFX583 et pFX64 pour le criblage de banques préalablement construites chez *E. coli*. La grande capacité de sécrétion de *S. lividans* et *B. subtilis* rend ces hôtes particulièrement intéressants pour le criblage fonctionnel. Dans la plupart des méthodes de criblage, les enzymes doivent en effet être produites dans le milieu de culture pour pouvoir être détectées. Idéalement la banque devrait pouvoir être transférée d'*E. coli* à *S. lividans* ou *B. subtilis* par conjugaison, pour ainsi conserver la diversité de la banque d'un hôte à l'autre. La conjugaison est rendue possible par la séquence *oriT* présente dans les vecteurs pFX583 et pFX64. Pouvoir effectuer la conjugaison à haut débit représente toute fois un défi technique de taille.

## 7. CONCLUSION

Dans le cadre de ce projet de recherche appliquée, de nouveaux outils moléculaires ont été développés pour l'expression de gènes recombinants et l'analyse métagénomique. De nombreux domaines de recherche requièrent la production de protéines recombinantes, faisant des systèmes d'expression des outils indispensables. La métagénomique permet d'accéder à l'énorme diversité génétique microbienne et est une approche puissante pour la découverte de nouvelles enzymes. Ces biocatalyseurs sont d'une grande utilité aussi bien pour la recherche que pour l'industrie biotechnologique.

Basé sur le système d'expression T7 d'*E. coli*, deux vecteurs multifonctionnels ont été construits, pFX583 et pFX64. Ces vecteurs peuvent être utilisés à la fois pour la production de protéines recombinantes et pour la construction et le criblage fonctionnel de banques métagénomiques. Deux hôtes d'expression compatibles avec le système T7 ont aussi été développés, soit *S. lividans* 10T7 et *B. subtilis* WBT7. Ces bactéries Gram-positif présentent des caractéristiques particulièrement intéressantes pour l'expression de gènes recombinants et le criblage fonctionnel de banques métagénomiques. Le vecteur pFX583 a été utilisé avec succès pour la construction et le criblage fonctionnel d'une banque métagénomique. Une nouvelle lipase a ainsi pu être isolée et caractérisée partiellement.

La fonctionnalité du système d'expression T7 de *B. subtilis* a pu être démontrée, mais des améliorations seront nécessaires pour permettre la sécrétion efficace de protéines recombinantes. Le système de *S. lividans* offre des performances beaucoup plus intéressantes et a fait le sujet d'une publication dans la revue *Applied and Environmental Microbiology*. De nombreux laboratoires de recherche travaillant avec les *Streptomyces* et/ou en métagénomique se sont intéressés à ce système. La souche *S. lividans* 10T7 et le vecteur pFX583 ont été transférés à des universités américaines, européennes et asiatiques.

## RÉFÉRENCES

- Aakvik, T., Degnes, K.F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Volker, U., Ellingsen, T.E. et Valla, S. (2009). "A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species." FEMS Microbiol Lett **296**(2): 149-158.
- Abrahmsen, L., Moks, T., Nilsson, B., Hellman, U. et Uhlen, M. (1985). "Analysis of signals for secretion in the staphylococcal protein A gene." EMBO J **4**(13B): 3901-3906.
- Alonso, J.C. (1989). "DNA replication of plasmids from Gram-positive bacteria in *Bacillus subtilis*. Plasmid pUB110 as a model system." Microbiologia **5**(1): 5-12.
- Amann, R.I., Ludwig, W. et Schleifer, K.H. (1995a). "Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation." Microbiol Rev **59**(1): 143-169.
- Amann, R.I., Ludwig, W. et Schleifer, K.H. (1995b). "Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation." Microbiol Rev **59**(1): 143-169.
- Ambulos, N.P., Jr., Mongkolsuk, S., Kaufman, J.D. et Lovett, P.S. (1985). "Chloramphenicol-induced translation of *cat-86* mRNA requires two cis-acting regulatory regions." J Bacteriol **164**(2): 696-703.
- Ambulos, N.P., Jr., Smith, T., Mulbry, W. et Lovett, P.S. (1990). "CUG as a mutant start codon for *cat-86* and *xylE* in *Bacillus subtilis*." Gene **94**(1): 125-128.

Anne, J. et Van Mellaert, L. (1993). "*Streptomyces lividans* as host for heterologous protein production." FEMS Microbiol Lett **114**(2): 121-128.

Arpigny, J.L. et Jaeger, K.E. (1999). "Bacterial lipolytic enzymes: classification and properties." Biochem J **343**: 177-183.

Atlung, T., Christensen, B.B. et Hansen, F.G. (1999). "Role of the rom protein in copy number control of plasmid pBR322 at different growth rates in *Escherichia coli* K-12." Plasmid **41**(2): 110-119.

Ayadi, D.Z., Chouayekh, H., Mhiri, S., Zerria, K., Fathallah, D.M. et Bejar, S. (2007). "Expression by *Streptomyces lividans* of the rat alpha integrin CD11b A-domain as a secreted and soluble recombinant protein." J Biomed Biotechnol **2007**(1): 54327.

Bailey, C.R., Bruton, C.J., Butler, M.J., Chater, K.F., Harris, J.E. et Hopwood, D.A. (1986). "Properties of in vitro recombinant derivatives of pJV1, a multi-copy plasmid from *Streptomyces phaeochromogenes*." J Gen Microbiol **132**(8): 2071-2078.

Band, L. et Henner, D.J. (1984). "*Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression." DNA **3**(1): 17-21.

Baneyx, F. (1999). "Recombinant protein expression in *Escherichia coli*." Curr Opin Biotechnol **10**(5): 411-421.

Barnard, G.C., Henderson, G.E., Srinivasan, S. et Gerngross, T.U. (2004). "High level recombinant protein expression in *Ralstonia eutropha* using T7 RNA polymerase based amplification." Protein Expr Purif **38**(2): 264-271.

Bell, P.J., Sunna, A., Gibbs, M.D., Curach, N.C., Nevalainen, H. et Bergquist, P.L. (2002). "Prospecting for novel lipase genes using PCR." Microbiology **148**(Pt 8): 2283-2291.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitzsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrell, B.G., Parkhill, J. et Hopwood, D.A. (2002). "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)." Nature **417**(6885): 141-147.

Berks, B.C., Palmer, T. et Sargent, F. (2005). "Protein targeting by the bacterial twin-arginine translocation (Tat) pathway." Curr Opin Microbiol **8**(2): 174-181.

Berks, B.C., Sargent, F. et Palmer, T. (2000). "The Tat protein export pathway." Mol Microbiol **35**(2): 260-274.

Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M. et Simonet, P. (2005). "High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction." J Microbiol Methods **62**(1): 1-11.

Bey, B.S., Fichot, E.B., Dayama, G., Decho, A.W. et Norman, R.S. (2010). "Extraction of high molecular weight DNA from microbial mats." Biotechniques **49**(3): 631-640.

Bhandari, P. et Gowrishankar, J. (1997). "An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer." J Bacteriol **179**(13): 4403-4406.

Bibb, M.J., Janssen, G.R. et Ward, J.M. (1985). "Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*." Gene **38**(1-3): 215-226.

Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. et Schoner, B.E. (1992). "Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp." Gene **116**(1): 43-49.

Binnie, C., Cossar, J.D. et Stewart, D.I. (1997). "Heterologous biopharmaceutical protein expression in *Streptomyces*." Trends Biotechnol **15**(8): 315-320.

Bolivar, F. (1979). "Molecular cloning vectors derived from the CoLE1 type plasmid pMB1." Life Sci **25**(10): 807-817.

Bond, J.S. et Butler, P.E. (1987). "Intracellular proteases." Annu Rev Biochem **56**: 333-364.

Bongers, R.S., Veening, J.W., Van Wieringen, M., Kuipers, O.P. et Kleerebezem, M. (2005). "Development and characterization of a subtilin-regulated expression system in *Bacillus subtilis*: strict control of gene expression by addition of subtilin." Appl Environ Microbiol **71**(12): 8818-8824.

Borodina, I., Krabben, P. et Nielsen, J. (2005). "Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism." Genome Res **15**(6): 820-829.

Braaz, R., Wong, S.L. et Jendrossek, D. (2002). "Production of PHA depolymerase A (PhaZ5) from *Pauicimonas lemoignei* in *Bacillus subtilis*." FEMS Microbiol Lett **209**(2): 237-241.

Brennan, Y., Callen, W.N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernandez, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G.P., Mathur, E.J., Short, J.M., Robertson, D.E. et Steer, B.A. (2004). "Unusual microbial xylanases from insect guts." Appl Environ Microbiol **70**(6): 3609-3617.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T. et Eggert, T. (2006). "Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria." J Mol Biol **362**(3): 393-402.

Bron, S., Holsappel, S., Venema, G. et Peeters, B.P. (1991a). "Plasmid deletion formation between short direct repeats in *Bacillus subtilis* is stimulated by single-stranded rolling-circle replication intermediates." Mol Gen Genet **226**(1-2): 88-96.

Bron, S., Meijer, W., Holsappel, S. et Haima, P. (1991b). "Plasmid instability and molecular cloning in *Bacillus subtilis*." Res Microbiol **142**(7-8): 875-883.

Brosius, J., Erfle, M. et Storella, J. (1985). "Spacing of the -10 and -35 regions in the *tac* promoter. Effect on its *in vivo* activity." J Biol Chem **260**(6): 3539-3541.

Bruckner, R., Shoseyov, O. et Doi, R.H. (1990). "Multiple active forms of a novel serine protease from *Bacillus subtilis*." Mol Gen Genet **221**(3): 486-490.

Buell, G., Schulz, M.F., Selzer, G., Chollet, A., Movva, N.R., Semon, D., Escanez, S. et Kawashima, E. (1985). "Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (IGF-I)." Nucleic Acids Res **13**(6): 1923-1938.

Bunterngsook, B., Kanokratana, P., Thongaram, T., Tanapongpipat, S., Uengwetwanit, T., Rachdawong, S., Vichitsoonthonkul, T. et Eurwilaichitr, L. (2010). "Identification and characterization of lipolytic enzymes from a peat-swamp forest soil metagenome." Biosci Biotechnol Biochem **74**(9): 1848-1854.

Carter, A.D., Morris, C.E. et McAllister, W.T. (1981). "Revised transcription map of the late region of bacteriophage T7 DNA." J Virol **37**(2): 636-642.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W. et Cedergren, R. (1997). "On the evolution of the single-subunit RNA polymerases." J Mol Evol **45**(6): 671-681.

Chamberlin, M., McGrath, J. et Waskell, L. (1970). "New RNA polymerase from *Escherichia coli* infected with bacteriophage T7." Nature **228**(5268): 227-231.

Chamberlin, M. et Ring, J. (1973). "Characterization of T7-specific ribonucleic acid polymerase. II. Inhibitors of the enzyme and their application to the study of the enzymatic reaction." J Biol Chem **248**(6): 2245-2250.

Chater, K.F. (2006). "*Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics." Philos Trans R Soc Lond B Biol Sci **361**(1469): 761-768.

Chater, K.F. et Chandra, G. (2008). "The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*." J Microbiol **46**(1): 1-11.

Chauhan, V., Singh, A., Waheed, S.M., Singh, S. et Bhatnagar, R. (2001). "Constitutive expression of protective antigen gene of *Bacillus anthracis* in *Escherichia coli*." Biochem Biophys Res Commun **283**(2): 308-315.

Chen, H., Bjerknes, M., Kumar, R. et Jay, E. (1994). "Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs." Nucleic Acids Res **22**(23): 4953-4957.

Chen, P.T., Shaw, J.F., Chao, Y.P., David Ho, T.H. et Yu, S.M. (2010). "Construction of chromosomally located T7 expression system for production of heterologous secreted proteins in *Bacillus subtilis*." J Agric Food Chem **58**(9): 5392-5399.

Chiu, M.L., Folcher, M., Katoh, T., Puglia, A.M., Vohradsky, J., Yun, B.S., Seto, H. et Thompson, C.J. (1999). "Broad spectrum thiopeptide recognition specificity of the *Streptomyces lividans* TipAL protein and its role in regulating gene expression." J Biol Chem **274**(29): 20578-20586.

Cieslinski, H., Bialkowskaa, A., Tkaczuk, K., Dlugolecka, A., Kur, J. et Turkiewicz, M. (2009). "Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library." Pol J Microbiol **58**(3): 199-204.

Claessen, D., de Jong, W., Dijkhuizen, L. et Wosten, H.A. (2006). "Regulation of *Streptomyces* development: reach for the sky!" Trends Microbiol **14**(7): 313-319.

Clewell, D.B., Yagi, Y., Dunny, G.M. et Schultz, S.K. (1974). "Characterization of three plasmid deoxyribonucleic acid molecules in a strain of *Streptococcus faecalis*: identification of a plasmid determining erythromycin resistance." J Bacteriol **117**(1): 283-289.

Combes, P., Till, R., Bee, S. et Smith, M.C. (2002). "The *Streptomyces* genome contains multiple pseudo-*attB* sites for the (phi)C31-encoded site-specific recombination system." J Bacteriol **184**(20): 5746-5752.

Conrad, B., Bashkirov, V.I. et Hofemeister, J. (1992). "Imprecise excision of plasmid pE194 from the chromosomes of *Bacillus subtilis* pE194 insertion strains." J Bacteriol **174**(21): 6997-7002.

Conrad, B., Savchenko, R.S., Breves, R. et Hofemeister, J. (1996). "A T7 promoter-specific, inducible protein expression system for *Bacillus subtilis*." Mol Gen Genet **250**(2): 230-236.

Côté, A. et Shareck, F. (2010). "Expression and characterization of a novel heterologous moderately thermostable lipase derived from metagenomics in *Streptomyces lividans*." J Ind Microbiol Biotechnol **37**(9): 883-891.

Cottrell, M.T., Wood, D.N., Yu, L. et Kirchman, D.L. (2000). "Selected chitinase genes in cultured and uncultured marine bacteria in the alpha- and gamma-subclasses of the proteobacteria." Appl Environ Microbiol **66**(3): 1195-1201.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P. et Pernodet, J.L. (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." Appl Environ Microbiol **69**(1): 49-55.

Couto, G.H., Glogauer, A., Faoro, H., Chubatsu, L.S., Souza, E.M. et Pedrosa, F.O. (2010). "Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast." Genet Mol Res **9**(1): 514-523.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R. et Wittwer, P. (2005). "Metagenomic gene discovery: past, present and future." Trends Biotechnol **23**(6): 321-329.

Daegelen, P., Studier, F.W., Lenski, R.E., Cure, S. et Kim, J.F. (2009). "Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3)." J Mol Biol **394**(4): 634-643.

Dahl, M.K., Degenkolb, J. et Hillen, W. (1994). "Transcription of the *xyl* operon is controlled in *Bacillus subtilis* by tandem overlapping operators spaced by four base-pairs." J Mol Biol **243**(3): 413-424.

Dahl, M.K., Schmiedel, D. et Hillen, W. (1995). "Glucose and glucose-6-phosphate interaction with Xyl repressor proteins from *Bacillus spp.* may contribute to regulation of xylose utilization." J Bacteriol **177**(19): 5467-5472.

Dairi, T., Nakano, T., Mizukami, T., Aisaka, K., Hasegawa, M. et Katsumata, R. (1995). "Conserved organization of genes for biosynthesis of chlortetracycline in *Streptomyces* strains." Biosci Biotechnol Biochem **59**(7): 1360-1361.

Dalbey, R.E., Lively, M.O., Bron, S. et van Dijl, J.M. (1997). "The chemistry and enzymology of the type I signal peptidases." Protein Sci **6**(6): 1129-1138.

Daniel, R. (2004). "The soil metagenome--a rich resource for the discovery of novel natural products." Curr Opin Biotechnol **15**(3): 199-204.

Daniel, R. (2005). "The metagenomics of soil." Nat Rev Microbiol **3**(6): 470-478.

Davanloo, P., Rosenberg, A.H., Dunn, J.J. et Studier, F.W. (1984). "Cloning and expression of the gene for bacteriophage T7 RNA polymerase." Proc Natl Acad Sci U S A **81**(7): 2035-2039.

Davis, G.D., Elisee, C., Newham, D.M. et Harrison, R.G. (1999). "New fusion protein systems designed to give soluble expression in *Escherichia coli*." Biotechnol Bioeng **65**(4): 382-388.

de Boer, H.A., Comstock, L.J. et Vasser, M. (1983). "The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters." Proc Natl Acad Sci U S A **80**(1): 21-25.

De Keersmaeker, S., Vrancken, K., Van Mellaert, L., Lammertyn, E., Anne, J. et Geukens, N. (2006). "Evaluation of TatABC overproduction on Tat- and Sec-dependent protein secretion in *Streptomyces lividans*." Arch Microbiol **186**(6): 507-512.

de Smit, M.H. et van Duin, J. (1990). "Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis." Proc Natl Acad Sci U S A **87**(19): 7668-7672.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. et Diaz-Orejas, R. (1998). "Replication and control of circular bacterial plasmids." Microbiol Mol Biol Rev **62**(2): 434-464.

Delbrück, M. (1945). "Interference between bacterial viruses: III. The mutual exclusion effect and the depressor effect." J Bacteriol **50**(2): 151-170.

Demain, A.L. et Vaishnav, P. (2009). "Production of recombinant proteins by microbes and higher organisms." Biotechnol Adv **27**(3): 297-306.

Demerec, M. et Fano, U. (1945). "Bacteriophage-resistant mutants in *Escherichia coli*." Genetics **30**(2): 119-136.

Denis, F. et Brzezinski, R. (1991). "An improved aminoglycoside resistance gene cassette for use in Gram-negative bacteria and *Streptomyces*." FEMS Microbiol Lett **65**(3): 261-264.

Denis, F. et Brzezinski, R. (1992). "A versatile shuttle cosmid vector for use in *Escherichia coli* and actinomycetes." Gene **111**(1): 115-118.

Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D. et Derewenda, Z.S. (1994). "Crystal structure, at 2.6-A resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of beta-1,4-D-glycanases." J Biol Chem **269**(33): 20811-20814.

Digilio, F.A., Morra, R., Pedone, E., Bartolucci, S. et Rossi, M. (2003). "High-level expression of *Aliciclobacillus acidocaldarius* thioredoxin in *Pichia pastoris* and *Bacillus subtilis*." Protein Expr Purif **30**(2): 179-184.

Distler, J., Braun, C., Ebert, A. et Piepersberg, W. (1987). "Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene." Mol Gen Genet **208**(1-2): 204-210.

Doekele, S., Eppelmann, K. et Marahiel, M.A. (2002). "Heterologous expression of nonribosomal peptide synthetases in *B. subtilis*: construction of a bi-functional *B. subtilis/E. coli* shuttle vector system." FEMS Microbiol Lett **216**(2): 185-191.

Dong, H., Nilsson, L. et Kurland, C.G. (1995). "Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction." J Bacteriol **177**(6): 1497-1504.

Doolittle, W.F. (1999). "Phylogenetic classification and the universal tree." Science **284**(5423): 2124-2129.

Drepper, T., Arvani, S., Rosenau, F., Wilhelm, S. et Jaeger, K.E. (2005). "High-level transcription of large gene regions: a novel T(7) RNA-polymerase-based system for expression of functional hydrogenases in the phototrophic bacterium *Rhodobacter capsulatus*." Biochem Soc Trans **33**(Pt 1): 56-58.

Drepper, T., Eggert, T., Hummel, W., Leggewie, C., Pohl, M., Rosenau, F., Wilhelm, S. et Jaeger, K.E. (2006). "Novel biocatalysts for white biotechnology." Biotechnol J **1**(7-8): 777-786.

Duan, C.J., Xian, L., Zhao, G.C., Feng, Y., Pang, H., Bai, X.L., Tang, J.L., Ma, Q.S. et Feng, J.X. (2009). "Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens." J Appl Microbiol **107**(1): 245-256.

Dubé, E., Shareck, F., Hurtubise, Y., Daneault, C. et Beauregard, M. (2008). "Homologous cloning, expression, and characterisation of a laccase from *Streptomyces coelicolor* and enzymatic decolourisation of an indigo dye." Appl Microbiol Biotechnol **79**(4): 597-603.

Dubendorff, J.W. et Studier, F.W. (1991). "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor." J Mol Biol **219**(1): 45-59.

Dubnau, D., Davidoff-Abelson, R., Scher, B. et Cirigliano, C. (1973). "Fate of transforming deoxyribonucleic acid after uptake by competent *Bacillus subtilis*: phenotypic characterization of radiation-sensitive recombination-deficient mutants." J Bacteriol **114**(1): 273-286.

Dubnau, D., Goldthwaite, C., Smith, I. et Marmur, J. (1967). "Genetic mapping in *Bacillus subtilis*." J Mol Biol **27**(1): 163-185.

Dunn, J.J. et Studier, F.W. (1983). "Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements." J Mol Biol **166**(4): 477-535.

Dupont, C., Daigneault, N., Shareck, F., Morosoli, R. et Kluepfel, D. (1996). "Purification and characterization of an acetyl xylan esterase produced by *Streptomyces lividans*." Biochem J **319** ( Pt 3): 881-886.

Earl, A.M., Losick, R. et Kolter, R. (2008). "Ecology and genomics of *Bacillus subtilis*." Trends Microbiol **16**(6): 269-275.

Ebanks, R., Dupont, M., Shareck, F., Morosoli, R., Kluepfel, D. et Dupont, C. (2000). "Development of an *Escherichia coli* expression system and thermostability screening assay for libraries of mutant xylanase." J Ind Microbiol Biotechnol **25**(6): 310-314.

Elend, C., Schmeisser, C., Leggewie, C., Babiak, P., Carballeira, J.D., Steele, H.L., Reymond, J.L., Jaeger, K.E. et Streit, W.R. (2006). "Isolation and biochemical characterization of two novel metagenome-derived esterases." Appl Environ Microbiol **72**(5): 3637-3645.

Elroy-Stein, O. et Moss, B. (1990). "Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells." Proc Natl Acad Sci U S A **87**(17): 6743-6747.

Entcheva, P., Liebl, W., Johann, A., Hartsch, T. et Streit, W.R. (2001). "Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia." Appl Environ Microbiol **67**(1): 89-99.

Fan, Q., Huang, F., Leadlay, P.F. et Spencer, J.B. (2008). "The neomycin biosynthetic gene cluster of *Streptomyces fradiae* NCIMB 8233: genetic and biochemical evidence for the roles of two glycosyltransferases and a deacetylase." Org Biomol Chem **6**(18): 3306-3314.

Faraldo, M.L., de Pedro, M.A. et Berenguer, J. (1991). "Cloning and expression in *Escherichia coli* of the structural gene coding for the monomeric protein of the S layer of *Thermus thermophilus* HB8." J Bacteriol **173**(17): 5346-5351.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A., Yakimov, M.M., Timmis, K.N. et Golyshin, P.N. (2005). "Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin." Chem Biol **12**(8): 895-904.

Fischer, J. (1996). Entwicklung eines regulierbaren Expressionssystems zur effizienten Synthese rekombinanter Proteine in *Streptomyces lividans*. Stuttgart, University of Stuttgart. Ph.D.: 163.

Fuerst, T.R., Niles, E.G., Studier, F.W. et Moss, B. (1986). "Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase." Proc Natl Acad Sci U S A **83**(21): 8122-8126.

Fukushima, T., Yamamoto, H., Atrihi, A., Foster, S.J. et Sekiguchi, J. (2002). "A polysaccharide deacetylase gene (*pdaA*) is required for germination and for production of muramic delta-lactam residues in the spore cortex of *Bacillus subtilis*." J Bacteriol **184**(21): 6007-6015.

Gabor, E.M., Alkema, W.B. et Janssen, D.B. (2004a). "Quantifying the accessibility of the metagenome by random expression cloning techniques." Environ Microbiol **6**(9): 879-886.

Gabor, E.M., de Vries, E.J. et Janssen, D.B. (2004b). "Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases." Environ Microbiol **6**(9): 948-958.

Gamer, M., Frode, D., Biedendieck, R., Stammen, S. et Jahn, D. (2009). "A T7 RNA polymerase-dependent gene expression system for *Bacillus megaterium*." Appl Microbiol Biotechnol **82**(6): 1195-1203.

Gartner, D., Degenkolb, J., Ripperger, J.A., Allmansberger, R. et Hillen, W. (1992). "Regulation of the *Bacillus subtilis* W23 xylose utilization operon: interaction of the Xyl repressor with the *xyl* operator and the inducer xylose." Mol Gen Genet **232**(3): 415-422.

Gartner, D., Geissendorfer, M. et Hillen, W. (1988). "Expression of the *Bacillus subtilis* *xyl* operon is repressed at the level of transcription and is induced by xylose." J Bacteriol **170**(7): 3102-3109.

Gauthier, C., Li, H. et Morosoli, R. (2005). "Increase in xylanase production by *Streptomyces lividans* through simultaneous use of the Sec- and Tat-dependent protein export systems." Appl Environ Microbiol **71**(6): 3085-3092.

Geissendorfer, M. et Hillen, W. (1990). "Regulated expression of heterologous genes in *Bacillus subtilis* using the *Tn10* encoded *tet* regulatory elements." Appl Microbiol Biotechnol **33**(6): 657-663.

Gilbert, M., Morosoli, R., Shareck, F. et Kluepfel, D. (1995). "Production and secretion of proteins by streptomycetes." Crit Rev Biotechnol **15**(1): 13-39.

Glauert, A.M. et Hopwood, D.A. (1961). "The fine structure of *Streptomyces violaceoruber* (*S. coelicolor*). III. The walls of the mycelium and spores." J Biophys Biochem Cytol **10**: 505-516.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. et Riggs, A.D. (1979). "Expression in *Escherichia coli* of chemically synthesized genes for human insulin." Proc Natl Acad Sci U S A **76**(1): 106-110.

Golomb, M. et Chamberlin, M. (1974). "Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis." J Biol Chem **249**(9): 2858-2863.

Gowrishankar, J. et Manna, D. (1996). "How is osmotic regulation of transcription of the *Escherichia coli proU* operon achieved? A review and a model." Genetica **97**(3): 363-378.

Grant, W.D. et Heaphy, S. (2010). "Metagenomics and recovery of enzyme genes from alkaline saline environments." Environ Technol **31**(10): 1135-1143.

Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhen-Paganon, S., Park, H.W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K.,

Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang, M., Montelione, G.T., Stuart, D.I., Owens, R.J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K. et Gunsalus, K.C. (2008). "Protein production and purification." Nat Methods **5**(2): 135-146.

Gray, C.H. et Tatum, E.L. (1944). "X-Ray Induced growth factor requirements in bacteria." Proc Natl Acad Sci U S A **30**(12): 404-410.

Greenfield, L., Boone, T. et Wilcox, G. (1978). "DNA sequence of the *araBAD* promoter in *Escherichia coli* B/r." Proc Natl Acad Sci U S A **75**(10): 4724-4728.

Gren, E.J. (1984). "Recognition of messenger RNA during translational initiation in *Escherichia coli*." Biochimie **66**(1): 1-29.

Grodbberg, J. et Dunn, J.J. (1988). "*ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification." J Bacteriol **170**(3): 1245-1253.

Gronenborn, B. (1976). "Overproduction of phage lambda repressor under control of the *lac* promotor of *Escherichia coli*." Mol Gen Genet **148**(3): 243-250.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R. et Osburne, M.S. (1998). "Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability." Gene **209**(1-2): 95-103.

Gryczan, T.J., Hahn, J., Contente, S. et Dubnau, D. (1982). "Replication and incompatibility properties of plasmid pE194 in *Bacillus subtilis*." J Bacteriol **152**(2): 722-735.

- Guimond, J. et Morosoli, R. (2008). "Identification of *Streptomyces lividans* proteins secreted by the twin-arginine translocation pathway following growth with different carbon sources." Can J Microbiol **54**(7): 549-558.
- Guzman, L.M., Belin, D., Carson, M.J. et Beckwith, J. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." J Bacteriol **177**(14): 4121-4130.
- Hager, P.W. et Rabinowitz, J.C. (1985a). "Inefficient translation of T7 late mRNA by *Bacillus subtilis* ribosomes. Implications for species-specific translation." J Biol Chem **260**(28): 15163-15167.
- Hager, P.W. et Rabinowitz, J.C. (1985b). Translational specificity in *Bacillus subtilis*. The molecular biology of the bacilli, vol II. Dubnau, D. New York, Academic Press.
- Hamraeus, G., Karhumaa, K. et Rutberg, B. (2002). "A 5' stem-loop and ribosome binding but not translation are important for the stability of *Bacillus subtilis aprE* leader mRNA." Microbiology **148**(Pt 6): 1795-1803.
- Handelsman, J. (2004). "Metagenomics: application of genomics to uncultured microorganisms." Microbiol Mol Biol Rev **68**(4): 669-685.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. et Goodman, R.M. (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." Chem Biol **5**(10): R245-249.
- Hardeman, F. et Sjoling, S. (2007). "Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment." FEMS Microbiol Ecol **59**(2): 524-534.

Hartl, B., Wehrl, W., Wiegert, T., Homuth, G. et Schumann, W. (2001). "Development of a new integration site within the *Bacillus subtilis* chromosome and construction of compatible expression cassettes." J Bacteriol **183**(8): 2696-2699.

Harwood, C.R. (1992). "*Bacillus subtilis* and its relatives: molecular biological and industrial workhorses." Trends Biotechnol **10**(7): 247-256.

Harwood, C.R. et Cranenburgh, R. (2008). "*Bacillus* protein secretion: an unfolding story." Trends Microbiol **16**(2): 73-79.

Haryoung Poo, J.J.S., Seung-Pyo Hong, Yoon-Ho Choi, Sun Woo Yun, Jin-Ho Kim, Sang Chul Lee, Sung-Goo Lee and Moon Hee Sung (2002). "Novel high-level constitutive expression system, pHCE vector, for a convenient and cost-effective soluble production of human tumor necrosis factor- $\alpha$ " Biotechnol Lett **24**(14): 1185-1189.

Hatanaka, T., Usuki, H., Arima, J., Uesugi, Y., Yamamoto, Y., Kumagai, Y., Yamasato, A. et Mukaihara, T. (2010) "Extracellular production and characterization of two *Streptomyces* L-asparaginases." Appl Biochem Biotechnol DOI: 10.1007/s12010-010-9087-9.

Haug, I., Weissenborn, A., Brodde, D., Bentley, S., Kieser, T. et Altenbuchner, J. (2003). "*Streptomyces coelicolor* A3(2) plasmid SCP2\*: deductions from the complete sequence." Microbiology **149**(Pt 2): 505-513.

Haugaard, E.S. et Haugaard, N. (1955). "Degradation of crystalline insulin by subtilisin (a proteinase from *B. subtilis*)."C R Trav Lab Carlsberg Chim **29**(20-21): 350-365.

Heath, C., Hu, X.P., Cary, S.C. et Cowan, D. (2009). "Identification of a novel alkaliphilic esterase active at low temperatures by screening a

metagenomic library from antarctic desert soil." Appl Environ Microbiol **75**(13): 4657-4659.

Henne, A., Daniel, R., Schmitz, R.A. et Gottschalk, G. (1999). "Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate." Appl Environ Microbiol **65**(9): 3901-3907.

Henner, D.J. (1990). "Inducible expression of regulatory genes in *Bacillus subtilis*." Methods Enzymol **185**: 223-228.

Herai, S., Hashimoto, Y., Higashibata, H., Maseda, H., Ikeda, H., Omura, S. et Kobayashi, M. (2004). "Hyper-inducible expression system for streptomycetes." Proc Natl Acad Sci U S A **101**(39): 14031-14035.

Hess, M., Sczyrba, A., Egan, R., Kim, T.W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z. et Rubin, E.M. (2011). "Metagenomic discovery of biomass-degrading genes and genomes from cow rumen." Science **331**(6016): 463-467.

Higuchi, R., Krummel, B. et Saiki, R.K. (1988). "A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions." Nucleic Acids Res **16**(15): 7351-7367.

Hirsh, J. et Schleif, R. (1977). "The *araC* promoter: transcription, mapping and interaction with the *araBAD* promoter." Cell **11**(3): 545-550.

Ho, K.M. et Lim, B.L. (2003). "Co-expression of a prophage system and a plasmid system in *Bacillus subtilis*." Protein Expr Purif **32**(2): 293-301.

Hodgson, D.A. (2000). "Primary metabolism and its control in streptomycetes: a most unusual group of bacteria." Adv Microb Physiol **42**: 47-238.

Hofemeister, B., Konig, S., Hoang, V., Engel, J., Mayer, G., Hansen, G. et Hofemeister, J. (1994). "The gene *amyE*(TV1) codes for a nonglucogenic alpha-amylase from *Thermoactinomyces vulgaris* 94-2A in *Bacillus subtilis*." Appl Environ Microbiol **60**(9): 3381-3389.

Hopwood, D.A. (1999). "Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*." Microbiology **145** ( Pt 9): 2183-2202.

Horinouchi, S. et Weisblum, B. (1982). "Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance." J Bacteriol **150**(2): 815-825.

Hu, P. et Switzer, R.L. (1995). "Evidence for substrate stabilization in regulation of the degradation of *Bacillus subtilis* aspartate transcarbamylase *in vivo*." Arch Biochem Biophys **316**(1): 260-266.

Hu, Y., Fu, C., Huang, Y., Yin, Y., Cheng, G., Lei, F., Lu, N., Li, J., Ashforth, E.J., Zhang, L. et Zhu, B. (2010). "Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment." FEMS Microbiol Ecol **72**(2): 228-237.

Hu, Y., Zhang, G., Li, A., Chen, J. et Ma, L. (2008). "Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach." Appl Microbiol Biotechnol **80**(5): 823-830.

Hurtubise, Y., Shareck, F., Kluepfel, D. et Morosoli, R. (1995). "A cellulase/xylanase-negative mutant of *Streptomyces lividans* 1326 defective in cellobiose and xylobiose uptake is mutated in a gene encoding

a protein homologous to ATP-binding proteins." Mol Microbiol **17**(2): 367-377.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. et Omura, S. (2003). "Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*." Nat Biotechnol **21**(5): 526-531.

Imanaka, T., Fujii, M. et Aiba, S. (1981). "Isolation and characterization of antibiotic resistance plasmids from thermophilic bacilli and construction of deletion plasmids." J Bacteriol **146**(3): 1091-1097.

Iordanescu, S. (1976). "Three distinct plasmids originating in the same *Staphylococcus aureus* strain." Arch Roum Pathol Exp Microbiol **35**(1-2): 111-118.

Iordanescu, S., Surdeanu, M., Della Latta, P. et Novick, R. (1978). "Incompatibility and molecular relationships between small Staphylococcal plasmids carrying the same resistance marker." Plasmid **1**(4): 468-479.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. et Boyer, H.W. (1977). "Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin." Science **198**(4321): 1056-1063.

Jaeger, K.E., Dijkstra, B.W. et Reetz, M.T. (1999). "Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases." Annu Rev Microbiol **53**: 315-351.

Jan, J., Valle, F., Bolivar, F. et Merino, E. (2001). "Construction of protein overproducer strains in *Bacillus subtilis* by an integrative approach." Appl Microbiol Biotechnol **55**(1): 69-75.

Jeon, J.H., Kim, J.T., Kim, Y.J., Kim, H.K., Lee, H.S., Kang, S.G., Kim, S.J. et Lee, J.H. (2009). "Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome." Appl Microbiol Biotechnol **81**(5): 865-874.

Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.H., Couloux, A., Lee, S.W., Yoon, S.H., Cattolico, L., Hur, C.G., Park, H.S., Segurens, B., Kim, S.C., Oh, T.K., Lenski, R.E., Studier, F.W., Daegelen, P. et Kim, J.F. (2009). "Genome sequences of Escherichia coli B strains REL606 and BL21(DE3)." J Mol Biol **394**(4): 644-652.

Jiang, Z., Wang, H., Ma, Y. et Wei, D. (2006). "Characterization of two novel lipase genes isolated directly from environmental sample." Appl Microbiol Biotechnol **70**(3): 327-332.

Jurgen, B., Schweder, T. et Hecker, M. (1998). "The stability of mRNA from the *gsiB* gene of *Bacillus subtilis* is dependent on the presence of a strong ribosome binding site." Mol Gen Genet **258**(5): 538-545.

Juteau, P., Tremblay, D., Ould-Moulaye, C.B., Bisaillon, J.G. et Beaudet, R. (2004). "Swine waste treatment by self-heating aerobic thermophilic bioreactors." Water Res **38**(3): 539-546.

Kang, Y., Son, M.S. et Hoang, T.T. (2007). "One step engineering of T7-expression strains for protein production: increasing the host-range of the T7-expression system" Protein Expr Purif **55**(2): 325-333.

Kataoka, M., Kosono, S. et Tsujimoto, G. (1999). "Spatial and temporal regulation of protein expression by *bldA* within a *Streptomyces lividans* colony." FEBS Lett **462**(3): 425-429.

Katz, E., Thompson, C.J. et Hopwood, D.A. (1983). "Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*." J Gen Microbiol **129**(9): 2703-2714.

Katzke, N., Arvani, S., Bergmann, R., Circolone, F., Markert, A., Svensson, V., Jaeger, K.E., Heck, A. et Drepper, T. (2009). "A novel T7 RNA polymerase dependent expression system for high-level protein production in the phototrophic bacterium *Rhodobacter capsulatus*." Protein Expr Purif.

Katzke, N., Arvani, S., Bergmann, R., Circolone, F., Markert, A., Svensson, V., Jaeger, K.E., Heck, A. et Drepper, T. (2010). "A novel T7 RNA polymerase dependent expression system for high-level protein production in the phototrophic bacterium *Rhodobacter capsulatus*." Protein Expr Purif **69**(2): 137-146.

Keggins, K.M., Lovett, P.S. et Duvall, E.J. (1978). "Molecular cloning of genetically active fragments of *Bacillus* DNA in *Bacillus subtilis* and properties of the vector plasmid pUB110." Proc Natl Acad Sci U S A **75**(3): 1423-1427.

Kelsall, S.R. et Kulski, J.K. (1995). "Expression of the major capsid protein of human papillomavirus type 16 in *Escherichia coli*." J Virol Methods **53**(1): 75-90.

Kendall, K.J. et Cohen, S.N. (1988). "Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties." J Bacteriol **170**(10): 4634-4651.

Khan, S.A. (2005). "Plasmid rolling-circle replication: highlights of two decades of research." Plasmid **53**(2): 126-136.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. et Hopwood, D.A. (2000). Practical Streptomyces genetics. Norwich, The John Innes Foundation.

Kieser, T., Hopwood, D.A., Wright, H.M. et Thompson, C.J. (1982). "pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors." Mol Gen Genet **185**(2): 223-228.

Kim, H.C. et Kim, D.M. (2009). "Methods for energizing cell-free protein synthesis." J Biosci Bioeng **108**(1): 1-4.

Kim, L., Mogk, A. et Schumann, W. (1996). "A xylose-inducible *Bacillus subtilis* integration vector and its application." Gene **181**(1-2): 71-76.

Kim, S.J., Lee, C.M., Han, B.R., Kim, M.Y., Yeo, Y.S., Yoon, S.H., Koo, B.S. et Jun, H.K. (2008). "Characterization of a gene encoding cellulase from uncultured soil bacteria." FEMS Microbiol Lett **282**(1): 44-51.

Kim, Y.H., Kwon, E.J., Kim, S.K., Jeong, Y.S., Kim, J., Yun, H.D. et Kim, H. (2010). "Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library." Biochem Biophys Res Commun **393**(1): 45-49.

Kim, Y.J., Choi, G.S., Kim, S.B., Yoon, G.S., Kim, Y.S. et Ryu, Y.W. (2006). "Screening and characterization of a novel esterase from a metagenomic library." Protein Expr Purif **45**(2): 315-323.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. et Daniel, R. (2003). "Construction and screening of metagenomic libraries derived from

enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*." Appl Environ Microbiol **69**(3): 1408-1416.

Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S.K., Dreesen, O., Errington, J., Fillinger, S., Foster, S.J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Le Coq, D., Masson, A., Mauel, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pragai, Z., Pooley, H.M., Rapoport, G., Rawlins, J.P., Rivas, L.A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H., Scanlan, E., Schumann, W., Seegers, J.F., Sekiguchi, J., Sekowska, A., Seror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U. et Ogasawara, N. (2003). "Essential *Bacillus subtilis* genes." Proc Natl Acad Sci U S A **100**(8): 4678-4683.

Kodama, T., Endo, K., Ara, K., Ozaki, K., Kakeshita, H., Yamane, K. et Sekiguchi, J. (2007a). "Effect of *Bacillus subtilis spo0A* mutation on cell wall lytic enzymes and extracellular proteases, and prevention of cell lysis." J Biosci Bioeng **103**(1): 13-21.

Kodama, T., Endo, K., Sawada, K., Ara, K., Ozaki, K., Kakeshita, H., Yamane, K. et Sekiguchi, J. (2007b). "*Bacillus subtilis* AprX involved in degradation

of a heterologous protein during the late stationary growth phase." J Biosci Bioeng **104**(2): 135-143.

Kölb, A., Busby, S., Buc, H., Garges, S. et Adhya, S. (1993). "Transcriptional regulation by cAMP and its receptor protein." Annu Rev Biochem **62**: 749-795.

Komeda, H., Hori, Y., Kobayashi, M. et Shimizu, S. (1996). "Transcriptional regulation of the *Rhodococcus rhodochrous* J1 *nitA* gene encoding a nitrilase." Proc Natl Acad Sci U S A **93**(20): 10572-10577.

Kontinen, V.P., Saris, P. et Sarvas, M. (1991). "A gene (*prsA*) of *Bacillus subtilis* involved in a novel, late stage of protein export." Mol Microbiol **5**(5): 1273-1283.

Kontinen, V.P. et Sarvas, M. (1993). "The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion." Mol Microbiol **8**(4): 727-737.

Korn-Wendisch, F. et Kutzner, H.J. (1992). The family *Streptomycetaceae*. The Prokaryotes. Balows, A., Trüper, H.G., Dworkin, M., Harder, W. et Schleifer, K.H. New York, Springer: 921-995.

Kouker, G. et Jaeger, K.E. (1987). "Specific and sensitive plate assay for bacterial lipases." Appl Environ Microbiol **53**(1): 211-213.

Krsek, M. et Wellington, E.M. (1999). "Comparison of different methods for the isolation and purification of total community DNA from soil." J Microbiol Methods **39**(1): 1-16.

Kuderova, A., Nanak, E., Truksa, M. et Brzobohaty, B. (1999). "Use of rifampicin in T7 RNA polymerase-driven expression of a plant enzyme: rifampicin improves yield and assembly." Protein Expr Purif **16**(3): 405-409.

Kues, U. et Stahl, U. (1989). "Replication of plasmids in Gram-negative bacteria." Microbiol Rev **53**(4): 491-516.

Kumar, C.C. et Novick, R.P. (1985). "Plasmid pT181 replication is regulated by two countertranscripts." Proc Natl Acad Sci U S A **82**(3): 638-642.

Lacatena, R.M., Banner, D.W., Castagnoli, L. et Cesareni, G. (1984). "Control of initiation of pMB1 replication: purified Rop protein and RNA I affect primer formation *in vitro*." Cell **37**(3): 1009-1014.

Lacatena, R.M. et Cesareni, G. (1983). "Interaction between RNA1 and the primer precursor in the regulation of CoIE1 replication." J Mol Biol **170**(3): 635-650.

Lacey, R.W. et Chopra, I. (1974). "Genetic studies of a multi-resistant strain of *Staphylococcus aureus*." J Med Microbiol **7**(2): 285-297.

Laemmli, U.K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

Lagodich, A.V., Shtaniuk Iu, V., Prozorov, A.A. et Titok, M.A. (2004). "The replication system of plasmids from *Bacillus subtilis* environmental isolates." Mol Biol (Mosk) **38**(3): 437-441.

Lam, K.H., Chow, K.C. et Wong, W.K. (1998). "Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins." J Biotechnol **63**(3): 167-177.

Lammertyn, E., Van Mellaert, L., Schacht, S., Dillen, C., Sablon, E., Van Broekhoven, A. et Anne, J. (1997). "Evaluation of a novel subtilisin inhibitor gene and mutant derivatives for the expression and secretion of mouse tumor necrosis factor alpha by *Streptomyces lividans*." Appl Environ Microbiol **63**(5): 1808-1813.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGgettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. et Higgins, D.G. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lawlor, E.J., Baylis, H.A. et Chater, K.F. (1987). "Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2)." Genes Dev **1**(10): 1305-1310.

Lee, C., Kim, J., Shin, S.G. et Hwang, S. (2006a). "Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*." J Biotechnol **123**(3): 273-280.

Lee, C.C., Kibblewhite-Accinelli, R.E., Wagschal, K., Robertson, G.H. et Wong, D.W. (2006b). "Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library." Extremophiles **10**(4): 295-300.

Lee, D.G., Jeon, J.H., Jang, M.K., Kim, N.Y., Lee, J.H., Kim, S.J., Kim, G.D. et Lee, S.H. (2007). "Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library." Biotechnol Lett **29**(3): 465-472.

Lee, J.K., Edwards, C.W. et Hulett, F.M. (1991). "Bacillus licheniformis APase I gene promoter: a strong well-regulated promoter in *B. subtilis*." J Gen Microbiol **137**(5): 1127-1133.

Lee, M.H., Hong, K.S., Malhotra, S., Park, J.H., Hwang, E.C., Choi, H.K., Kim, Y.S., Tao, W. et Lee, S.W. (2010). "A new esterase EstD2 isolated from plant rhizosphere soil metagenome." Appl Microbiol Biotechnol **88**(5): 1125-1134.

Lee, P.C., Park, I.Y., Kim, M.S. et Kim, S.C. (1998). "Effect of the replication mode of a plasmid on the stability of multimeric endoxylanase genes in *Bacillus subtilis*." J Biotechnol **62**(3): 177-185.

Lee, S.J., Kim, D.M., Bae, K.H., Byun, S.M. et Chung, J.H. (2000). "Enhancement of secretion and extracellular stability of staphylokinase in *Bacillus subtilis* by *wprA* gene disruption." Appl Environ Microbiol **66**(2): 476-480.

Leggewie, C., Henning, H., Schmeisser, C., Streit, W.R. et Jaeger, K.E. (2006). "A novel transposon for functional expression of DNA libraries." J Biotechnol **123**(3): 281-287.

Leloup, L., Haddaoui el, A., Chambert, R. et Petit-Glatron, M.F. (1997). "Characterization of the rate-limiting step of the secretion of *Bacillus subtilis* alpha-amylase overproduced during the exponential phase of growth." Microbiology **143** ( Pt 10): 3295-3303.

Leskela, S., Wahlstrom, E., Kontinen, V.P. et Sarvas, M. (1999). "Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the Lgt gene." Mol Microbiol **31**(4): 1075-1085.

Leskiw, B.K., Mah, R., Lawlor, E.J. et Chater, K.F. (1993). "Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2)." J Bacteriol **175**(7): 1995-2005.

Lesuisse, E., Schanck, K. et Colson, C. (1993). "Purification and preliminary characterization of the extracellular lipase of *Bacillus subtilis* 168, an extremely basic pH-tolerant enzyme." Eur J Biochem **216**(1): 155-160.

Li, W., Zhou, X. et Lu, P. (2004). "Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*." Res Microbiol **155**(8): 605-610.

Li, Y., Wexler, M., Richardson, D.J., Bond, P.L. et Johnston, A.W. (2005). "Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes." Environ Microbiol **7**(12): 1927-1936.

Liu, K., Wang, J., Bu, D., Zhao, S., McSweeney, C., Yu, P. et Li, D. (2009). "Isolation and biochemical characterization of two lipases from a metagenomic library of China Holstein cow rumen." Biochem Biophys Res Commun **385**(4): 605-611.

Lomovskaya, N.D., Chater, K.F. et Mkrtumian, N.M. (1980). "Genetics and molecular biology of *Streptomyces* bacteriophages." Microbiol Rev **44**(2): 206-229.

Loomis, W.F., Jr. et Magasanik, B. (1966). "Nature of the effector of catabolite repression of beta-galactosidase in *Escherichia coli*." J Bacteriol **92**(1): 170-177.

Lorenz, P. et Eck, J. (2005). "Metagenomics and industrial applications." Nat Rev Microbiol **3**(6): 510-516.

Lorenz, P., Liebeton, K., Niehaus, F. et Eck, J. (2002). "Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space." Curr Opin Biotechnol **13**(6): 572-577.

Lussier, F.X., Chambenoit, O., Côté, A., Hupé, J.F., Denis, F., Juteau, P., Beaudet, R. et Shareck, F. (2010a) "Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector." J Ind Microbiol Biotechnol DOI: 10.1007/s10295-010-0915-2.

Lussier, F.X., Denis, F. et Shareck, F. (2010b). "Adaptation of the highly productive T7 expression system to *Streptomyces lividans*." Appl Environ Microbiol **76**(3): 967-970.

Maarit Niemi, R., Heiskanen, I., Wallenius, K. et Lindström, K. (2001). "Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia." Journal of Microbiological Methods **45**(3): 155-165.

Makoff, A.J. et Oxer, M.D. (1991). "High level heterologous expression in *E. coli* using mutant forms of the *lac* promoter." Nucleic Acids Res **19**(9): 2417-2421.

Margot, P. et Karamata, D. (1996). "The *wprA* gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease." Microbiology **142** ( Pt 12): 3437-3444.

Martinez, A., Kolvek, S.J., Yip, C.L., Hopke, J., Brown, K.A., MacNeil, I.A. et Osburne, M.S. (2004). "Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing

environmental libraries and detecting heterologous natural products in multiple expression hosts." Appl Environ Microbiol **70**(4): 2452-2463.

Mayer, M.P. (1995). "A new set of useful cloning and expression vectors derived from pBlueScript." Gene **163**(1): 41-46.

McAllister, W.T. et Carter, A.D. (1980). "Regulation of promoter selection by the bacteriophage T7 RNA polymerase in vitro." Nucleic Acids Res **8**(20): 4821-4837.

McAllister, W.T., Morris, C., Rosenberg, A.H. et Studier, F.W. (1981). "Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection." J Mol Biol **153**(3): 527-544.

McAllister, W.T. et Raskin, C.A. (1993). "The phage RNA polymerases are related to DNA polymerases and reverse transcriptases." Mol Microbiol **10**(1): 1-6.

McLaughlin, J.R., Murray, C.L. et Rabinowitz, J.C. (1981). "Plasmid-directed expression of *Staphylococcus aureus* beta-lactamase by *Bacillus subtilis* in vitro." J Biol Chem **256**(21): 11272-11283.

Meijer, W.J., van der Lelie, D., Venema, G. et Bron, S. (1995). "Effects of the generation of single-stranded DNA on the maintenance of plasmid pMV158 and derivatives in different *Bacillus subtilis* strains." Plasmid **33**(2): 79-89.

Meilleur, C., Hupe, J.F., Juteau, P. et Shareck, F. (2009). "Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library." J Ind Microbiol Biotechnol **36**(6): 853-861.

Mo, X.C., Chen, C.L., Pang, H., Feng, Y. et Feng, J.X. (2010). "Identification and characterization of a novel xylanase derived from a rice straw degrading enrichment culture." Appl Microbiol Biotechnol **87**(6): 2137-2146.

Moffatt, B.A., Dunn, J.J. et Studier, F.W. (1984). "Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase." J Mol Biol **173**(2): 265-269.

Mogk, A., Mayer, M.P. et Deuerling, E. (2002). "Mechanisms of protein folding: molecular chaperones and their application in biotechnology." Chembiochem **3**(9): 807-814.

Moks, T., Abrahmsen, L., Holmgren, E., Bilich, M., Olsson, A., Uhlen, M., Pohl, G., Sterky, C., Hultberg, H., Josephson, S. et al (1987). "Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification." Biochemistry **26**(17): 5239-5244.

Morosoli, R., Shareck, F. et Kluepfel, D. (1997). "Protein secretion in streptomycetes." FEMS Microbiol Lett **146**(2): 167-174.

Munson, L.M., Stormo, G.D., Niece, R.L. et Reznikoff, W.S. (1984). "*lacZ* translation initiation mutations." J Mol Biol **177**(4): 663-683.

Murakami, T., Holt, T.G. et Thompson, C.J. (1989). "Thiostrepton-induced gene expression in *Streptomyces lividans*." J Bacteriol **171**(3): 1459-1466.

Muth, G., Wohlleben, W. et Puhler, A. (1988). "The minimal replicon of the *Streptomyces ghanaensis* plasmid pSG5 identified by subcloning and Tn5 mutagenesis." Mol Gen Genet **211**(3): 424-429.

Natale, P., Bruser, T. et Driessen, A.J. (2008). "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms." Biochim Biophys Acta **1778**(9): 1735-1756.

Nguyen, H.D., Nguyen, Q.A., Ferreira, R.C., Ferreira, L.C., Tran, L.T. et Schumann, W. (2005). "Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability." Plasmid **54**(3): 241-248.

Nisole, A., Lussier, F.X., Morley, K.L., Shareck, F., Kazlauskas, R.J., Dupont, C. et Pelletier, J.N. (2006). "Extracellular production of *Streptomyces lividans* acetyl xylan esterase A in *Escherichia coli* for rapid detection of activity." Protein Expr Purif **46**(2): 274-284.

Noda, S., Ito, Y., Shimizu, N., Tanaka, T., Ogino, C. et Kondo, A. (2010). "Over-production of various secretory-form proteins in *Streptomyces lividans*." Protein Expr Purif **73**(2): 198-202.

O'Rourke, S., Wietzorek, A., Fowler, K., Corre, C., Challis, G.L. et Chater, K.F. (2009). "Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*." Mol Microbiol **71**(3): 763-778.

Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D. et Schleif, R. (1980). "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**(6): 3346-3350.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. et Horinouchi, S. (2008). "Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350." J Bacteriol **190**(11): 4050-4060.

Olmos-Soto, J. et Contreras-Flores, R. (2003). "Genetic system constructed to overproduce and secrete proinsulin in *Bacillus subtilis*." Appl Microbiol Biotechnol **62**(4): 369-373.

Osborne, A.R., Rapoport, T.A. et van den Berg, B. (2005). "Protein translocation by the Sec61/SecY channel." Annu Rev Cell Dev Biol **21**: 529-550.

Palva, I. (1982). "Molecular cloning of alpha-amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*." Gene **19**(1): 81-87.

Palva, I., Lehtovaara, P., Kaariainen, L., Sibakov, M., Cantell, K., Schein, C.H., Kashiwagi, K. et Weissmann, C. (1983). "Secretion of interferon by *Bacillus subtilis*." Gene **22**(2-3): 229-235.

Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L. et Feng, J.X. (2009). "Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase." Curr Microbiol **58**(4): 404-408.

Park, E.J., Kim, K.H., Abell, G.C., Kim, M.S., Roh, S.W. et Bae, J.W. (2010) "Metagenomic analysis of the viral community in fermented foods." Appl Environ Microbiol DOI: AEM.01859-10 [pii] 10.1128/AEM.01859-10.

Paul, E.A. et Clark, F.E. (1989). Soil microbiology and biochemistry. San Diego, Academic Press.

Pettersson, E., Lundeberg, J. et Ahmadian, A. (2009). "Generations of sequencing technologies." Genomics **93**(2): 105-111.

Phan, T.T., Nguyen, H.D. et Schumann, W. (2006). "Novel plasmid-based expression vectors for intra- and extracellular production of recombinant proteins in *Bacillus subtilis*." Protein Expr Purif **46**(2): 189-195.

Phan, T.T. et Schumann, W. (2007). "Development of a glycine-inducible expression system for *Bacillus subtilis*." J Biotechnol **128**(3): 486-499.

Pimienta, E., Ayala, J.C., Rodriguez, C., Ramos, A., Van Mellaert, L., Vallin, C. et Anne, J. (2007). "Recombinant production of *Streptococcus equisimilis* streptokinase by *Streptomyces lividans*." Microb Cell Fact **6**: 20.

Podzelinska, K., Latimer, R., Bhattacharya, A., Vining, L.C., Zechel, D.L. et Jia, Z. (2010). "Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate bond." J Mol Biol **397**(1): 316-331.

Polisky, B., Bishop, R.J. et Gelfand, D.H. (1976). "A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria." Proc Natl Acad Sci U S A **73**(11): 3900-3904.

Ranjan, R., Grover, A., Kapardar, R.K. et Sharma, R. (2005). "Isolation of novel lipolytic genes from uncultured bacteria of pond water." Biochem Biophys Res Commun **335**(1): 57-65.

Raynal, A., Friedmann, A., Tophile, K., Guerineau, M. et Pernodet, J.L. (2002). "Characterization of the *attP* site of the integrative element pSAM2 from *Streptomyces ambofaciens*." Microbiology **148**(Pt 1): 61-67.

Rebets, Y.V., Ostash, B.O., Fukuhara, M., Nakamura, T. et Fedorenko, V.O. (2006). "Expression of the regulatory protein LndI for landomycin E production in *Streptomyces globisporus* 1912 is controlled by the availability of tRNA for the rare UUA codon." FEMS Microbiol Lett **256**(1): 30-37.

Reddy, C.A., Beveridge, T.J., Breznak, J.A., Marzluf, G., Schmidt, T.M. et Snyder, L.R. (2007). Methods for General and Molecular Microbiology. Washington, DC, ASM Press.

Rees, H.C., Grant, S., Jones, B., Grant, W.D. et Heaphy, S. (2003). "Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries." Extremophiles 7(5): 415-421.

Riggs, A.D., Suzuki, H. et Bourgeois, S. (1970). "*lac* repressor-operator interaction. I. Equilibrium studies." J Mol Biol 48(1): 67-83.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J. et Goodman, R.M. (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." Appl Environ Microbiol 66(6): 2541-2547.

Roose-Amsaleg, C.L., Garnier-Sillam, E. et Harry, M. (2001). "Extraction and purification of microbial DNA from soil and sediment samples." Applied Soil Ecology 18(1): 47-60.

Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J. et Studier, F.W. (1987). "Vectors for selective expression of cloned DNAs by T7 RNA polymerase." Gene 56(1): 125-135.

Rufo, G.A., Jr., Sullivan, B.J., Sloma, A. et Pero, J. (1990). "Isolation and characterization of a novel extracellular metalloprotease from *Bacillus subtilis*." J Bacteriol 172(2): 1019-1023.

Rygus, T. et Hillen, W. (1991). "Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-

utilization operon." Applied Microbiology and Biotechnology **35**(5): 594-599.

Ryqus, T. et Hillen, W. (1992). "Catabolite repression of the *xyl* operon in *Bacillus megaterium*." J Bacteriol **174**(9): 3049-3055.

Ryqus, T., Scheler, A., Allmansberger, R. et Hillen, W. (1991). "Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus megaterium* encoded regulon for xylose utilization." Arch Microbiol **155**(6): 535-542.

Saida, F., Uzan, M., Odaert, B. et Bontems, F. (2006). "Expression of highly toxic genes in *E. coli*: special strategies and genetic tools." Curr Protein Pept Sci **7**(1): 47-56.

saiSree, L., Reddy, M. et Gowrishankar, J. (2001). "IS186 insertion at a hot spot in the *lon* promoter as a basis for *lon* protease deficiency of *Escherichia coli* B: identification of a consensus target sequence for IS186 transposition." J Bacteriol **183**(23): 6943-6946.

Sambrook, J., Fritsch, E.F. et Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. New-York, Cold Spring Harbor Laboratory press.

Schaerlaekens, K., Lammertyn, E., Geukens, N., De Keersmaeker, S., Anne, J. et Van Mellaert, L. (2004). "Comparison of the Sec and Tat secretion pathways for heterologous protein production by *Streptomyces lividans*." J Biotechnol **112**(3): 279-288.

Schallmey, M., Singh, A. et Ward, O.P. (2004). "Developments in the use of *Bacillus* species for industrial production." Can J Microbiol **50**(1): 1-17.

Schein, C.H. et Noteborn, M.H.M. (1988). "Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature." Nat Biotech **6**(3): 291-294.

Schloss, P.D. et Handelsman, J. (2003). "Biotechnological prospects from metagenomics." Curr Opin Biotechnol **14**(3): 303-310.

Schmidt, F.R. (2004). "Recombinant expression systems in the pharmaceutical industry." Appl Microbiol Biotechnol **65**(4): 363-372.

Schmitt-John, T. et Engels, J.W. (1992). "Promoter constructions for efficient secretion expression in *Streptomyces lividans*." Appl Microbiol Biotechnol **36**(4): 493-498.

Schumann, W. (2007). "Production of recombinant proteins in *Bacillus subtilis*." Adv Appl Microbiol **62**: 137-189.

Schumann, W. et Ferreira, L.C.S. (2004). "Production of recombinant proteins in *Escherichia coli*." Genet. Mol. Biol. **27**(3): 442-453.

Servin-Gonzalez, L. (1993). "Relationship between the replication functions of *Streptomyces* plasmids pJV1 and pIJ101." Plasmid **30**(2): 131-140.

Sharma, S., Khan, F.G. et Qazi, G.N. (2010). "Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas." Appl Microbiol Biotechnol **86**(6): 1821-1828.

Sharp, J.S. et Bechhofer, D.H. (2003). "Effect of translational signals on mRNA decay in *Bacillus subtilis*." J Bacteriol **185**(18): 5372-5379.

Shepard, H.M., Yelverton, E. et Goeddel, D.V. (1982). "Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome binding sites." DNA **1**(2): 125-131.

Shine, J. et Dalgarno, L. (1974). "The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites." Proc Natl Acad Sci U S A **71**(4): 1342-1346.

Shokri, A., Sanden, A.M. et Larsson, G. (2003). "Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*." Appl Microbiol Biotechnol **60**(6): 654-664.

Sianidis, G., Pozidis, C., Becker, F., Vrancken, K., Sjoeholm, C., Karamanou, S., Takamiya-Wik, M., van Mellaert, L., Schaefer, T., Anne, J. et Economou, A. (2006). "Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*." J Biotechnol **121**(4): 498-507.

Sigurgísladóttir, S., Konráðsdóttir, M., Jónsson, Á., Kristjánsson, J.K. et Matthiasson, E. (1993). "Lipase activity of thermophilic bacteria from icelandic hot springs." Biotechnology Letters **15**(4): 361-366.

Simon, C. et Daniel, R. (2009). "Achievements and new knowledge unraveled by metagenomic approaches." Appl Microbiol Biotechnol **85**(2): 265-276.

Skarstad, K. et Boye, E. (1988). "Perturbed chromosomal replication in *recA* mutants of *Escherichia coli*." J Bacteriol **170**(6): 2549-2554.

Sloma, A., Rufo, G.A., Jr., Rudolph, C.F., Sullivan, B.J., Theriault, K.A. et Pero, J. (1990). "Bacillopeptidase F of *Bacillus subtilis*: purification of the protein and cloning of the gene." J Bacteriol **172**(3): 1470-1477.

Sloma, A., Rufo, G.A., Jr., Theriault, K.A., Dwyer, M., Wilson, S.W. et Pero, J. (1991). "Cloning and characterization of the gene for an additional extracellular serine protease of *Bacillus subtilis*." J Bacteriol **173**(21): 6889-6895.

Smokvina, T., Boccard, F., Pernodet, J.L., Friedmann, A. et Guerineau, M. (1991). "Functional analysis of the *Streptomyces ambofaciens* element pSAM2." Plasmid **25**(1): 40-52.

Soetaert, W. et Vandamme, E. (2006). "The impact of industrial biotechnology." Biotechnol J **1**(7-8): 756-769.

Sorensen, H.P. et Mortensen, K.K. (2005). "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*." J Biotechnol **115**(2): 113-128.

Stanley, N.R., Palmer, T. et Berks, B.C. (2000). "The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*." J Biol Chem **275**(16): 11591-11596.

Steele, H.L., Jaeger, K.E., Daniel, R. et Streit, W.R. (2009). "Advances in recovery of novel biocatalysts from metagenomes." J Mol Microbiol Biotechnol **16**(1-2): 25-37.

Streit, W.R., Daniel, R. et Jaeger, K.E. (2004). "Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms." Curr Opin Biotechnol **15**(4): 285-290.

Strohl, W.R. et Connors, N.C. (1992). "Significance of anthraquinone formation resulting from the cloning of actinorhodin genes in heterologous streptomycetes." Mol Microbiol **6**(2): 147-152.

Studier, F.W. (1991). "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system." J Mol Biol **219**(1): 37-44.

Studier, F.W., Daegelen, P., Lenski, R.E., Maslov, S. et Kim, J.F. (2009). "Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes." J Mol Biol **394**(4): 653-680.

Studier, F.W. et Moffatt, B.A. (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." J Mol Biol **189**(1): 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. et Dubendorff, J.W. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Methods Enzymol **185**: 60-89.

Sullivan, M.A., Yasbin, R.E. et Young, F.E. (1984). "New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments." Gene **29**(1-2): 21-26.

Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. et Davies, G.J. (1997). "The *Streptomyces lividans* family 12 endoglucanase: construction of the catalytic core, expression, and X-ray structure at 1.75 Å resolution." Biochemistry **36**(51): 16032-16039.

Surzycki, S. (2000). Basic Techniques in Molecular Biology. Berlin, Springer.

Sweet, C.R. (2003). "Expression of recombinant proteins from *lac* promoters." Methods Mol Biol **235**: 277-288.

Takano, E., White, J., Thompson, C.J. et Bibb, M.J. (1995). "Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp." Gene **166**(1): 133-137.

Tang, K., Utairungsee, T., Kanokratana, P., Sriprang, R., Champreda, V., Eurwilaichitr, L. et Tanapongpipat, S. (2006). "Characterization of a novel cyclomaltodextrinase expressed from environmental DNA isolated from Bor Khleung hot spring in Thailand." FEMS Microbiol Lett **260**(1): 91-99.

Tatum, E.L. (1945). "X-Ray induced mutant strains of *Escherichia coli*." Proc Natl Acad Sci U S A **31**(8): 215-219.

Tatum, E.L. et Lederberg, J. (1947). "Gene recombination in the bacterium *Escherichia coli*." J Bacteriol **53**(6): 673-684.

Taylor, R.G., Walker, D.C. et McInnes, R.R. (1993). "*E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing." Nucleic Acids Res **21**(7): 1677-1678.

Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **72**(2): 211-222.

Thompson, C.J., Ward, J.M. et Hopwood, D.A. (1980). "DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species." Nature **286**(5772): 525-527.

Thomson, C.A., Delaquis, P.J. et Mazza, G. (1999). "Detection and measurement of microbial lipase activity: a review." Crit Rev Food Sci Nutr **39**(2): 165-187.

Thorpe, H.M., Wilson, S.E. et Smith, M.C. (2000). "Control of directionality in the site-specific recombination system of the *Streptomyces* phage phiC31." Mol Microbiol **38**(2): 232-241.

Thuy Le, A.T. et Schumann, W. (2007). "A novel cold-inducible expression system for *Bacillus subtilis*." Protein Expr Purif **53**(2): 264-269.

Titok, M., Suski, C., Dalmais, B., Ehrlich, S.D. et Jannière, L. (2006). "The replicative polymerases PolC and DnaE are required for theta replication of the *Bacillus subtilis* plasmid pBS72." Microbiology **152**(Pt 5): 1471-1478.

Titok, M.A., Chapuis, J., Selezneva, Y.V., Lagodich, A.V., Prokulevich, V.A., Ehrlich, S.D. et Janniere, L. (2003). "*Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector." Plasmid **49**(1): 53-62.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S. et van Dijl, J.M. (2000). "Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome." Microbiol Mol Biol Rev **64**(3): 515-547.

Torsvik, V., Daae, F.L., Sandaa, R.A. et Ovreas, L. (1998). "Novel techniques for analysing microbial diversity in natural and perturbed environments." J Biotechnol **64**(1): 53-62.

Torsvik, V., Goksoyr, J. et Daae, F.L. (1990). "High diversity in DNA of soil bacteria." Appl Environ Microbiol **56**(3): 782-787.

Torsvik, V. et Ovreas, L. (2002). "Microbial diversity and function in soil: from genes to ecosystems." Curr Opin Microbiol **5**(3): 240-245.

Torsvik, V., Sørheim, R. et Goksøyr, J. (1996). "Total bacterial diversity in soil and sediment communities—A review." J Ind Microbiol Biotechnol **17**(3): 170-178.

Tran, L., Wu, X.C. et Wong, S.L. (1991). "Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*." J Bacteriol **173**(20): 6364-6372.

Turner, P., Holst, O. et Karlsson, E.N. (2005). "Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration." Protein Expr Purif **39**(1): 54-60.

Ueda, Y., Taguchi, S., Nishiyama, K., Kumagai, I. et Miura, K. (1993). "Effect of a rare leucine codon, TTA, on expression of a foreign gene in *Streptomyces lividans*." Biochim Biophys Acta **1172**(3): 262-266.

Uozumi, T., Ozaki, A., Beppu, T. et Arima, K. (1980). "New cryptic plasmid of *Bacillus subtilis* and restriction analysis of other plasmids found by general screening." J Bacteriol **142**(1): 315-318.

Vagner, V., Dervyn, E. et Ehrlich, S.D. (1998). "A vector for systematic gene inactivation in *Bacillus subtilis*." Microbiology **144** (Pt 11): 3097-3104.

Vallin, C., Ramos, A., Pimienta, E., Rodriguez, C., Hernandez, T., Hernandez, I., Del Sol, R., Rosabal, G., Van Mellaert, L. et Anne, J. (2006). "*Streptomyces* as host for recombinant production of *Mycobacterium tuberculosis* proteins." Tuberculosis (Edinb) **86**(3-4): 198-202.

Van Mellaert, L., Mei, L., Lammertyn, E., Schacht, S. et Anne, J. (1998). "Site-specific integration of bacteriophage VWB genome into *Streptomyces*

*venezuelae* and construction of a VWB-based integrative vector." Microbiology **144** ( Pt 12): 3351-3358.

Vapnek, D., Alton, N.K., Bassett, C.L. et Kushner, S.R. (1976). "Amplification in *Escherichia coli* of enzymes involved in genetic recombination: construction of hybrid ColE1 plasmids carrying the structural gene for exonuclease I." Proc Natl Acad Sci U S A **73**(10): 3492-3496.

Vazquez-Cruz, C., Ochoa-Sanchez, J.C. et Olmedo-Alvarez, G. (1996). "Pulse-field gel-electrophoretic analysis of the amplification and copy-number stability of an integrational plasmid in *Bacillus subtilis*." Appl Microbiol Biotechnol **46**(1): 55-60.

Vellanoweth, R.L. et Rabinowitz, J.C. (1992). "The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* *in vivo*." Mol Microbiol **6**(9): 1105-1114.

Vera, A., Gonzalez-Montalban, N., Aris, A. et Villaverde, A. (2007). "The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures." Biotechnol Bioeng **96**(6): 1101-1106.

Verger, R. (1997). "Interfacial activation' of lipases: facts and artifacts." Trends in Biotechnology **15**(1): 32-38.

Vincent, P., Shareck, F., Dupont, C., Morosoli, R. et Kluepfel, D. (1997). "New alpha-L-arabinofuranosidase produced by *Streptomyces lividans*: cloning and DNA sequence of the *abfB* gene and characterization of the enzyme." Biochem J **322** ( Pt 3): 845-852.

Voelkerding, K.V., Dames, S.A. et Durtschi, J.D. (2009). "Next-generation sequencing: from basic research to diagnostics." Clin Chem **55**(4): 641-658.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E. et Streit, W.R. (2003). "Prospecting for novel biocatalysts in a soil metagenome." Appl Environ Microbiol **69**(10): 6235-6242.

von Heijne, G. (1984). "Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells." EMBO J **3**(10): 2315-2318.

von Heijne, G. (1985). "Signal sequences. The limits of variation." J Mol Biol **184**(1): 99-105.

von Heijne, G. (1986). "A new method for predicting signal sequence cleavage sites." Nucleic Acids Res **14**(11): 4683-4690.

von Heijne, G. et Abrahmsen, L. (1989). "Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts." FEBS Lett **244**(2): 439-446.

Vrancken, K., De Keersmaecker, S., Geukens, N., Lammertyn, E., Anne, J. et Van Mellaert, L. (2007). "*pspA* overexpression in *Streptomyces lividans* improves both Sec- and Tat-dependent protein secretion." Appl Microbiol Biotechnol **73**(5): 1150-1157.

Wandersman, C. (1989). "Secretion, processing and activation of bacterial extracellular proteases." Mol Microbiol **3**(12): 1825-1831.

Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T. et Charles, T.C. (2006). "Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants." Appl Environ Microbiol **72**(1): 384-391.

Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J. et Davies, J. (2000). "Novel natural products from soil DNA libraries in a streptomycete host." Org Lett **2**(16): 2401-2404.

Wang, K., Li, G., Yu, S.Q., Zhang, C.T. et Liu, Y.H. (2010a). "A novel metagenome-derived beta-galactosidase: gene cloning, overexpression, purification and characterization." Appl Microbiol Biotechnol **88**(1): 155-165.

Wang, X.J., Yan, Y.J., Zhang, B., An, J., Wang, J.J., Tian, J., Jiang, L., Chen, Y.H., Huang, S.X., Yin, M., Zhang, J., Gao, A.L., Liu, C.X., Zhu, Z.X. et Xiang, W.S. (2010b). "Genome sequence of the milbemycin-producing bacterium *Streptomyces bingchengensis*." J Bacteriol **192**(17): 4526-4527.

Wanner, B.L., Kodaira, R. et Neidhardt, F.C. (1977). "Physiological regulation of a decontrolled *lac* operon." J Bacteriol **130**(1): 212-222.

Wanner, B.L., Kodaira, R. et Neidhardt, F.C. (1978). "Regulation of *lac* operon expression: reappraisal of the theory of catabolite repression." J Bacteriol **136**(3): 947-954.

Waschkowitz, T., Rockstroh, S. et Daniel, R. (2009). "Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries." Appl Environ Microbiol **75**(8): 2506-2516.

Weber, M., Teeling, H., Huang, S., Waldmann, J., Kassabgy, M., Fuchs, B.M., Klindworth, A., Klockow, C., Wichels, A., Gerdts, G., Amann, R. et Glockner, F.O. (2010) "Practical application of self-organizing maps to

interrelate biodiversity and functional data in NGS-based metagenomics."  
ISME J DOI: ismej2010180 [pii]  
10.1038/ismej.2010.180.

Wei, P., Bai, L., Song, W. et Hao, G. (2009). "Characterization of two soil metagenome-derived lipases with high specificity for *p*-nitrophenyl palmitate." Arch Microbiol **191**(3): 233-240.

Westers, L., Westers, H. et Quax, W.J. (2004). "*Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism." Biochim Biophys Acta **1694**(1-3): 299-310.

Wexler, M., Bond, P.L., Richardson, D.J. et Johnston, A.W. (2005). "A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase." Environ Microbiol **7**(12): 1917-1926.

White, J.H. et Richardson, C.C. (1987). "Gene 18 protein of bacteriophage T7. Overproduction, purification, and characterization." J Biol Chem **262**(18): 8845-8850.

Whitman, W.B., Coleman, D.C. et Wiebe, W.J. (1998). "Prokaryotes: the unseen majority." Proc Natl Acad Sci U S A **95**(12): 6578-6583.

Whittier, R.F. et Chase, J.W. (1981). "DNA repair in *E. coli* strains deficient in single-strand DNA binding protein." Mol Gen Genet **183**(2): 341-347.

Wickner, W., Driessens, A.J. et Hartl, F.U. (1991). "The enzymology of protein translocation across the *Escherichia coli* plasma membrane." Annu Rev Biochem **60**: 101-124.

Willetts, N.S., Clark, A.J. et Low, B. (1969). "Genetic location of certain mutations conferring recombination deficiency in *Escherichia coli*." J Bacteriol **97**(1): 244-249.

Wilson, C.J., Zhan, H., Swint-Kruse, L. et Matthews, K.S. (2007). "The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding." Cell Mol Life Sci **64**(1): 3-16.

Wong, S.L. (1995). "Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins." Curr Opin Biotechnol **6**(5): 517-522.

Wright, F. et Bibb, M.J. (1992). "Codon usage in the G+C-rich *Streptomyces* genome." Gene **113**(1): 55-65.

Wright, M. (1971). "Mutants of *Escherichia coli* lacking endonuclease I, ribonuclease I, or ribonuclease II." J Bacteriol **107**(1): 87-94.

Wu, S.C. et Wong, S.L. (2002). "Engineering of a *Bacillus subtilis* strain with adjustable levels of intracellular biotin for secretory production of functional streptavidin." Appl Environ Microbiol **68**(3): 1102-1108.

Wu, S.C., Ye, R., Wu, X.C., Ng, S.C. et Wong, S.L. (1998). "Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones." J Bacteriol **180**(11): 2830-2835.

Wu, S.C., Yeung, J.C., Duan, Y., Ye, R., Szarka, S.J., Habibi, H.R. et Wong, S.L. (2002). "Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: effects of molecular chaperones and a wall-bound protease on antibody fragment production." Appl Environ Microbiol **68**(7): 3261-3269.

Wu, X.C., Lee, W., Tran, L. et Wong, S.L. (1991). "Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases." J Bacteriol **173**(16): 4952-4958.

Wu, X.C., Ng, S.C., Near, R.I. et Wong, S.L. (1993). "Efficient production of a functional single-chain antidigoxin antibody via an engineered *Bacillus subtilis* expression-secretion system." Biotechnology (N Y) **11**(1): 71-76.

Xie, G., Chain, P.S., Lo, C.C., Liu, K.L., Gans, J., Merritt, J. et Qi, F. (2010). "Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing." Mol Oral Microbiol **25**(6): 391-405.

Xue, G.P., Johnson, J.S. et Dalrymple, B.P. (1999). "High osmolarity improves the electro-transformation efficiency of the Gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*." Journal of Microbiological Methods **34**(3): 183-191.

Yang, M.Y., Ferrari, E. et Henner, D.J. (1984). "Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation." J Bacteriol **160**(1): 15-21.

Yang, S., Huang, H., Zhang, R., Huang, X., Li, S. et Yuan, Z. (2001). "Expression and purification of extracellular penicillin G acylase in *Bacillus subtilis*." Protein Expr Purif **21**(1): 60-64.

Yansura, D.G. et Henner, D.J. (1984). "Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*." Proc Natl Acad Sci U S A **81**(2): 439-443.

Yasbin, R.E., Wilson, G.A. et Young, F.E. (1975). "Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells." J Bacteriol **121**(1): 296-304.

Ye, R., Kim, J.H., Kim, B.G., Szarka, S., Sihota, E. et Wong, S.L. (1999). "High-level secretory production of intact, biologically active staphylokinase from *Bacillus subtilis*." Biotechnol Bioeng **62**(1): 87-96.

Yeh, C.M., Chang, H.K., Hsieh, H.M., Yoda, K., Yamasaki, M. et Tsai, Y.C. (1997). "Improved translational efficiency of subtilisin YaB gene with different initiation codons in *Bacillus subtilis* and alkalophilic *Bacillus* YaB." J Appl Microbiol **83**(6): 758-763.

Youssef, N.H., Couger, M.B. et Elshahed, M.S. (2010). "Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere." PLoS One **5**(8): e12414.

Yuan, G. et Wong, S.L. (1995). "Isolation and characterization of *Bacillus subtilis* *groE* regulatory mutants: evidence for *orf39* in the *dnaK* operon as a repressor gene in regulating the expression of both *groE* and *dnaK*." J Bacteriol **177**(22): 6462-6468.

Zaman, S., Radnedge, L., Richards, H. et Ward, J.M. (1993). "Analysis of the site for second-strand initiation during replication of the *Streptomyces* plasmid pIJ101." J Gen Microbiol **139**(4): 669-676.

Zhang, T. et Han, W.J. (2009). "Gene cloning and characterization of a novel esterase from activated sludge metagenome." Microb Cell Fact **8**: 67.

Zhou, J., Bruns, M.A. et Tiedje, J.M. (1996). "DNA recovery from soils of diverse composition." Appl Environ Microbiol **62**(2): 316-322.

Zhu, Y., Wang, L., Du, Y., Wang, S., Yu, T. et Hong, B. (2011). "Heterologous expression of human interleukin-6 in *Streptomyces lividans* TK24 using novel secretory expression vectors." Biotechnol Lett 33(2): 253-261.

