

2m11.2593.5

Université de Montréal

**Présence du système de recombinaison spécifique de site  
Xer chez des espèces du groupe des bactéries lactiques**

par  
Shona Teijeiro

Département de microbiologie et immunologie

Faculté de médecine

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maître ès sciences (M.Sc.) en microbiologie et immunologie

fevrier, 1997

©Shona Teijeiro, 1997



W

4

U58

1997

V.081

Identification du jury

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé :

Présence du système de recombinaison spécifique de site  
Xer chez des espèces du groupe des bactéries lactiques

présenté par :

Shona Teijeiro

a été évalué par un jury composé des personnes suivantes :

Dr. Gabriel Drapeau.....président-rapporteur

Dr. George Szatmari.....directeur de recherche

Dr. Jean-Charles Côté.....membre du jury

Mémoire accepté le : ..... 26.08.97 .....

## SOMMAIRE

Chez les procaryotes, la réplication du chromosome peut entraîner la dimérisation des deux nouveaux chromosomes par recombinaison homologe. La fréquence de ce phénomène est particulièrement élevée dans la région du terminus de réplication du chromosome de *E. coli*, notamment aux sites *Hot* et *TRZ*. Si le dimère chromosomique n'est pas résolu avant la partition dans les cellules-filles, il y aura filamentation cellulaire et aberration chromosomique. Pour contrer ce phénomène, l'évolution a mis en place un système de résolution de dimères, soit le système de recombinaison spécifique de site Xer. Chez *E. coli*, ce système comprend les recombinases XerC et XerD et le site chromosomique *dif* situé dans la région du terminus de réplication. Les recombinases s'arriment d'abord aux sites *dif* des chromosomes formant le dimère, puis il y a coupure et échange des brins libérant ainsi les chromosomes à leurs formes monomères.

La répartition de ce système de résolution de dimère est assez bien documenté parmi les bactéries gram négatives. Cependant peu d'études se sont penchées sur les bactéries gram positives. Le but de notre étude est d'essayer de détecter la présence de ce système chez treize espèces de bactéries gram positives du groupe des bactéries lactiques, certaines étant très importantes dans l'industrie alimentaire. Plusieurs techniques ont été employées, dont l'hybridation Southern, l'amplification génique avec amorces dégénérées issues des séquences des recombinases déjà connues, l'amplification génique dite "inverse", ainsi que le séquençage. Nous avons également utilisé un programme informatique afin de dégager la phylogénie et les ressemblances entre les recombinases de gram positives et entre celles-ci et les recombinases déjà connues des bactéries gram négatives.

Par hybridation Southern nous avons réussi à démontrer la présence d'un site analogue au site *dif* de *E. coli* chez *Lactobacillus caseii*. De plus, l'amplification génique avec amorces dégénérées a montré que cette dernière espèce ainsi que les espèces *Lactobacillus plantarum* souche 8014, *Lactobacillus plantarum* souche 14917 et *Lactobacillus bulgaricus* 737 possèdent toutes au moins une recombinase de type Xer. Nous n'avons pas obtenu de séquences pour les autres espèces de bactéries, peut-être à cause d'un manque d'homologie entre les amorces et l'ADN ciblé. Les régions amplifiées de ces quatre souches, correspondant à environ 430 pb du bout C-terminal de la recombinase, ont été séquencées, puis des amorces ont été fabriquées à partir de la séquence de *L. caseii*. Ces amorces devaient amplifier les régions flanquant la séquence connue (amplification "inverse"), permettant ainsi d'obtenir une séquence complète du gène. Nous n'avons pas obtenu de résultats avec cette technique, peut-être à cause de la formation de boucles C-G dans l'ADN simple brin lors de certains cycles de l'amplification. Les quatre séquences ont ensuite été comparées et une phylogénie a été établie. Nous avons trouvé que ces séquences formaient un groupe phylogénique fortement apparenté et qu'elles ressemblaient le plus aux séquences XerD des bactéries gram négatives, ces dernières ressemblances surpassant même celle avec la recombinase XerD de la bactérie gram positive *B. subtilis*. Ces derniers résultats suggèrent la possibilité d'une évolution différente des recombinases XerD à l'intérieur du groupe des gram positives que nous avons examiné.

Mots clés : 1-Recombinaison spécifique de site, 2- *Lactobacillus*, 3- Recombinase XerD, 4- Recombinase XerC, 5-Site *dif*

## TABLE DES MATIERES

Sommaire	i
Table des matières	iii
Liste des tableaux	v
Liste des figures	vi
Liste des abréviations	viii
<b>CHAPITRE I : INTRODUCTION</b>	1
La réplication d'un génome circulaire	1
<i>A- Initiation de la réplication</i>	2
<i>B- L'élongation</i>	3
<i>C- La terminaison de la réplication</i>	4
<i>D- La séparation des chromosomes</i>	7
Les systèmes de recombinaison spécifique de site	8
Le système de recombinaison spécifique de site <i>Xer C-D</i> de <i>E.coli</i>	10
<i>A- Les recombinases XerC et XerD</i>	11
<i>B- Le site génomique dif</i>	12
Mécanisme d'action du système <i>XerC-D</i>	15
Le système <i>XerC-D</i> chez les plasmides	17
Systèmes analogues à <i>Xer</i> chez d'autres bactéries	18
Projet de recherche	19
<b>CHAPITRE II : MATERIEL ET METHODES</b>	21
<b>MATÉRIEL</b>	21
Souches bactériennes utilisées	21
<i>A- Bactéries Gram positives</i>	21
<i>B- Bactéries Gram négatives</i>	22
Vecteurs utilisés	22
<i>A- Vecteurs utilisés pour le séquençage</i>	22
<i>B- Vecteurs utilisés pour les hybridations Southern</i>	22
<i>C- Vecteurs pPS22, pSD105 et pRM130</i>	23
<b>MÉTHODES</b>	24
Cultures bactériennes	24
Les extractions d'ADN génomique des bactéries lactiques	24
Les extractions d'ADN des plasmides-vecteurs	25

Enzymes de restrictions utilisées	26
Electrophorèses	26
Clonage	27
<i>A- Ligations</i>	27
<i>B- Transformation par électroporation</i>	27
<i>C- Transformation chimique</i>	28
Les amplifications géniques (“PCR”)	29
<i>A- Conditions des réactions d’amplification de l’ADN</i>	30
<i>B- Types d’amorces</i>	31
<i>C- Solutions pour l’amplification génique</i>	33
Les hybridations Southern	34
<i>A- Pré-hybridation et hybridation</i>	35
<i>B- Sondes</i>	35
<i>C- Lavages et détection</i>	36
Le séquençage	37
Les analyses statistiques	38
<b>CHAPITRE III : RESULTATS</b>	40
Le site <i>dif</i> chez <i>Lactobacillus casei</i>	40
Les recombinaisons chez les bactéries lactiques	43
<i>A- Les amplifications géniques avec amorces dégénérées</i>	43
<i>B- Les amplifications géniques “inverses”</i>	48
Analyses statistiques des séquences des recombinaisons	49
<i>A- Comparaison des séquences de recombinaisons</i>	50
<i>B- Phylogénies suggérées</i>	52
<b>CHAPITRE IV : DISCUSSION</b>	55
<b>CHAPITRE V : CONCLUSION</b>	60
Bibliographie	x
Annexe I	xvii
Annexe II	xviii
Annexe III	xxii
Remerciements	xv

LISTE DES TABLEAUX

Tableau I :	
Caractéristiques majeures des recombinases résolvases-invertases et $\lambda$ -intégrases	10
Tableau II :	
Liste des espèces et souches gram positives utilisées	21
Tableau III:	
Amorces utilisées lors des amplifications géniques	32
Tableau IV:	
Résultats des PCR avec les amorces A1 et C sur les ADN des bactéries lactiques	46
Tableau V:	
Pourcentage de consensus entre XerC et XerD et les espèces séquencées	51

LISTE DES FIGURES

Figure 1:	
Mécanisme de dimérisation des chromosomes par recombinaison homologue	7
Figure 2:	
Cartes des régions entourant les gènes <i>xerC</i> et <i>xerD</i>	12
Figure 3:	
Séquences en pb des sites <i>dif</i> , <i>cer</i> , <i>psi</i> , <i>clf</i> , ColK et ColE3	13
Figure 4:	
Mécanisme d'action des recombinaases XerC et XerD sur le site <i>dif</i>	15
Figure 5:	
Cartes des régions entourant les gènes <i>xerC</i> de <i>L. leichmannii</i> et <i>codV</i> de <i>B.subtilis</i>	19
Figure 6:	
Vecteur pSHON issu du vecteur pKS492	23
Figure 7:	
Extraction d'ADN de <i>Lactobacillus caseii</i>	40
Figure 8:	
Hybridation Southern entre ADN de <i>L.caseii</i> digéré avec <i>EcoRI</i> , et la sonde <i>dif</i>	41

Figure 9: Hybridation Southern entre ADN de <i>L.caseii</i> digéré avec plusieurs enzymes, et la sonde <i>dif</i>	42
Figure 10: Amplification des témoins pSD105, pRM130 et pPS22 avec les amorces A1 et C	44
Figure 11: Bande amplifiée de 450 pb de <i>L. caseii</i> (PCR avec A1 et C)	47
Figure 12: Arbre phylogénétique des séquences peptidiques	52
Figure 13: Arbre phylogénétique des séquences d'ADN	53

LISTE DES ABREVIATIONS

A:	Adénine
a.a.:	Acide aminé
ADN:	Acide desoxyribonucléique
Ap:	Ampicilline
BHI:	“Brain and Heart Infusion”
C:	Cytosine
C°:	Degré Celsius
cm:	Centimètre
dATP:	Déoxyadénosine 5'-triphosphate
dCTP:	Déoxycytosine 5'-triphosphate
dGTP:	Déoxyguanosine 5'-triphosphate
dNTP:	Déoxynueléotide 5'-triphosphate
dTTP:	Déoxythymidine 5'-triphosphate
digUTP:	Digoxigénine uridine 5'-triphosphate
D.O.:	Densité optique
EDTA:	Acide éthylène-diamine-tetra-acétique
EtOH:	Ethanol
G:	Guanine
g:	Gramme
H:	Histidine
I:	Isoleucine
kpb:	Kilo paire de bases
kD:	Kilodalton
KOAc:	Acétate de Potassium
LB:	Luria Broth
min:	Minute
µg:	Microgramme
µL:	Microlitre
mg:	Milligramme
MgCl <sub>2</sub> :	Chlorure de magnésium
mL:	Millilitre
MRS:	De Man, Rogosa, Sharpe
NaCl:	Chlorure de sodium
NaOH:	Hydroxide de sodium

NBT:	Nitroblue Tetrazolium
NEB :	New England Biolabs
pb:	Paire de bases
PCR:	“Polymerase Chain Reaction”
Q:	Glutamine
R:	Arginine
RPM:	Révolution par minute
sec:	Seconde
SDS:	Sodium dodecyl sulfate
T:	Thymine
TAE:	Tris-acétate
TB:	Terrific Broth
TE:	Tris-EDTA
TEN:	Tris-EDTA-NaCl
U:	Unité
UPGMA:	Unweighted pair group method using arithmetic averages
UV:	Ultraviolet
X-Pho:	5-Bromo-4-Chloro-3-Indolyl phosphate
Y:	Tyrosine

# CHAPITRE 1

## INTRODUCTION

Chez les eucaryotes comme chez les procaryotes la division cellulaire est précédée par la réplication et la ségrégation des chromosomes vers les pôles de la cellule. Quoique le fonctionnement de leur système génétique soit le même, les procaryotes et les eucaryotes diffèrent sur un grand nombre de points, dont l'aspect et le nombre de chromosomes. Ainsi, chez les procaryotes le chromosome est unique et circulaire alors que chez les eucaryotes les chromosomes sont multiples et linéaires. Ces différences influencent la manière dont l'ADN se réplique ainsi que l'intégrité du chromosome. Ainsi, la forme linéaire est plus susceptible aux attaques des exonucléases alors que la forme circulaire peut souffrir de dimérisation via la recombinaison homologe (Leslie & Sherratt, 1995). Les procaryotes et les eucaryotes ont donc évolués différemment pour mettre en place des mécanismes optimaux de réplication et de ségrégation des chromosomes dans les cellules-filles.

### **La réplication d'un génome circulaire**

Chez *Escherichia coli*, la réplication du chromosome exige la participation de 25 à 30 produits de gènes différents qui s'organisent à l'intérieur de la machinerie réplivative (Marians, 1996). La réplication et la séparation des chromosomes se fait en quatre étapes soit: A- l'initiation de la réplication, B- l'élongation, C- la terminaison de la réplication, D- la séparation des chromosomes. La ségrégation des chromosomes vers les pôles cellulaires et la constriction cellulaire (formation d'un septum divisant la cellules en deux) est initiée

tout de suite après la terminaison de la réplication (Schmid et von Freiseleben, 1996).

#### *A- Initiation de la réplication*

L'initiation de la réplication du chromosome de *E.coli* se fait à partir d'une origine fixe appelée *oriC* situé à la minute 84 sur la carte génétique (Kayashi *et al*, 1989). A partir de ce point deux fourches de réplication avancent, chacune dans une direction, vers la région du terminus de réplication située à l'opposé d'*oriC* (Kuempel *et al*, 1991).

Pour être fonctionnel *oriC* doit contenir un minimum de 258 pb. On y retrouve cinq séquences répétitives non palindromiques de 9 pb appelées boîtes DnaA sur lesquelles vont se fixer les produits du gène *dnaA*. A l'intérieur de ces boîtes se trouvent 9 à 14 séquences nucléiques GATC (Messer et Weigel, 1996), ce qui représente un taux 10 fois plus élevé que celui auquel on s'attendrait pour une distribution aléatoire de cette séquence (Marians, 1996). Les sites GATC sont des sites de méthylation par la méthyltransferase Dam (Mathews, 1996).

La protéine DnaA est fondamentale dans l'initiation de la réplication et dans la formation et l'organisation du replisome et du primosome (Messer et Weigel, 1996). La première étape dans le mécanisme d'initiation est la méthylation des sites GATC par l'enzyme Dam. Le statut méthylatoire des sites GATC régule l'expression du gène *dnaA* (situé à 42 kpb d'*oriC*), leur méthylation accroissant la synthèse de la protéine DnaA. Les molécules de DnaA se fixent ensuite sur les boîtes DnaA et donnent lieu à une structure nucléoprotéique secondaire qui va induire le début de la réplication (Mathews, 1996). La transcription des gènes *mioC* et *gid* situés de part et d'autre d'*oriC* semble également être

un stimulus pour ce dernier (Cashel *et al*, 1996). De plus, les protéines HUa et HUb, homologues aux histones d'eucaryotes, ainsi que les protéines IHF et FIS s'arriment à *oriC* et induisent une déformation de l'ADN qui stimulerait également la réplication (Messer et Weigel, 1996; Schmid et von Freiesleben, 1996).

Après l'arrimage aux boîtes DnaA, les molécules de DnaA commencent à dérouler l'ADN sur 13 mers, puis l'hélicase DnaB (associée à DnaC) déroule le duplex parental tout en relarguant DnaC. L'activation de l'hélicase est conditionnelle à la présence de DnaA. L'hélicase interagit ensuite avec la primase DnaG et DnaA est relarguée. La fonction de la primase DnaG associée à l'hélicase (les deux forment le primosome) est de produire des amorces d'ARN de 10 à 12 pb essentielles à l'initiation de la synthèse des fragments d'Okazaki. Le primosome requiert les protéines DnaC, DnaT, PriA, PriB et PriC pour son fonctionnement (Kadner, 1996; Marians, 1996; Messer et Weigel, 1996).

### *B- L'élongation*

Une fois que les amorces d'ARN sont produites, l'ADN Polymérase III HE s'assemble à chaque fourche et le processus d'élongation des brins commence. La polymérase est précédée de l'hélicase DnaB qui déroule l'ADN et de l'ADN Gyrase (topoisomérase de type II) qui enlève les super tours positifs de l'ADN et provoque l'apparition de super tours négatifs. Ces super tours négatifs activent alors le processus de réplication (Marians, 1996; Messer et Weigel, 1996).

L'élongation de l'ADN se fait différemment selon la polarité du brin synthétisé. Ainsi le brin 5'---3' ("leading strand") est synthétisé de façon continue par la polymérase, alors que

le brin 3'---5' (brin retardé) est synthétisé de façon discontinue en utilisant les fragments d'Okazaki. Ces fragments sont absolument nécessaires puisque la polymérase ne peut synthétiser l'ADN que dans le sens 5'---3'. Après synthèse de fragments d'environ 2 kpb, les amorces d'ARN sont enlevées par la RNase HI et l'Exonucléase 5'---3' de la Polymérase I. Les brèches entre les fragments sont ensuite colmatées soit par la Polymérase I ou la Polymérase II. Finalement les bouts de chaque fragment sont joints entre eux par l'ADN ligase. Tous ces mécanismes permettent au brin retardé de devenir un brin continu (Marians, 1996; Messer et Weigel, 1996).

### *C- La terminaison de la réplication*

Cette étape de la réplication est moins bien connue que celle de l'initiation et de l'élongation. Après un certain temps d'élongation, les deux fourches répliquatives arrivent, chacune dans une direction, dans une zone nommée le terminus de réplication où elles seront freinées. Les délimitations exactes du terminus de réplication sont difficiles à calculer puisque, à l'exception du site de recombinaison spécifique de site *dif*, l'entièreté du terminus peut subir une délétion sans affecter significativement la viabilité de la cellule (Hill, 1996). Cependant on peut définir le terminus de réplication comme la région opposée à *oriC* entre les minutes 27 à 36 de la carte génétique de *E.coli* qui comprend les sites les plus importants impliqués dans la terminaison de la réplication, soit ceux de l'arrêt des fourches répliquatives (sites *Ter*), de la décaténation et de la recombinaison des nouveaux chromosomes (*TRZ* et *dif*) (McMaken *et al*, 1987; Kuempel *et al*, 1991).

Les sites *TerA*, *TerB*, *TerC*, *TerD*, *TerE* et *TerF* sont des sites terminateurs qui

piègent les fourches de réplication. Les sites *TerA*, *TerD* et *TerE* arrêtent la fourche arrivant dans le sens anti-horaire et les sites *TerB*, *TerC* et *TerF* arrêtent la fourche arrivant dans le sens horaire. Un septième site, *TerG*, présentant de fortes homologues avec les autres sites *Ter*, a récemment été découvert mais son rôle comme terminateur n'est pas encore clair. Les sites *TerE*, *TerF* et *TerG* sont situés à l'extérieur de la région du terminus. Les sites *TerE* et *TerF* ne joueraient pas de rôle important dans la réplication normale mais entreraient en jeu dans des conditions exceptionnelles lorsque les autres terminateurs manquent à leur fonction (Hill, 1996).

Les sites *Ter* nécessitent le produit du gène *tus* (ou *tau*) pour leur fonction d'inhibition (Pelletier *et al*, 1988; Kuempel *et al*, 1989; 1991; Kayashi *et al*, 1989; Hill, 1996). Le gène *tus* est situé près de *TerB*. La protéine Tus ("Terminus utilizing substance") est constituée de 309 acides aminés qui s'arrime sous forme de monomère aux sites *Ter*. La forme monomère fait que les faces de la protéine exposées aux fourches répliquatives ne seront pas les mêmes. Ainsi une des faces sera permissive au passage d'une fourche répliquative dans une direction seulement. Ceci empêche les fourches de sortir du terminus pour se diriger vers l'origine de réplication. La cible de la protéine Tus est l'hélicase DnaB: la protéine Tus arrimée à un site *Ter* arrête la fourche répliquative en empêchant l'hélicase de dérouler l'ADN. Un autre mécanisme d'arrêt de la réplication serait tout simplement dû à la collision des deux fourches répliquatives. Les proportions relatives des arrêts répliquatifs dû à l'un ou l'autre de ces mécanismes n'est cependant pas connu (Hill, 1996).

Le terminus de réplication est reconnu pour sa très haute incidence de recombinaison homologue (ou générale). La recombinaison homologue est un échange génétique entre

brins d'ADN présentant de très fortes homologies entre eux. Les brins d'ADN échangés peuvent être très longs (plus grand que 1 Kpb). Les enzymes majeures de la recombinaison homologue sont RecA et RecBCD. L'enzyme RecBCD est impliquée dans la présynapse de la recombinaison, c'est-à-dire dans les réactions préparant l'ADN pour l'échange des brins (exposition de la partie 3' des brins d'ADN simple brin), alors que l'enzyme RecA est responsable de la synapse recombinatoire (catalyse du pairage des brins homologues et échange des brins) et est également impliquée dans la postsynapse (maturation des hétéroduplexes intermédiaires) (Lloyd et Brooks Low, 1996).

Les hauts taux de recombinaison homologue observés dans le terminus sont situés au niveau des sites *Hot* et du site *TRZ* ("Terminus Recombination Zone") (Cornet *et al*, 1994; Rothfield, 1994, Hill, 1996). Les sites *Hot* (A, B, C, D, E, F, G, H) sont majoritairement compris entre les sites *TerA* et *TerB*. La forte hausse du taux de recombinaison à ces sites serait explicable par le fait que les fourches répliquatives sont arrêtées (sur les sites *Ter*) et qu'il y a proximité de deux séquences identiques d'ADN. Le site *TRZ* est situé à la minute 33.8, près de *TerC* et du site de recombinaison spécifique de site *dif*. Ce site constitue une région d'hyper-recombinaison maximale, complètement indépendant des sites *dif*, *Hot* et *Ter*. Cette hyper-recombinaison s'expliquerait partiellement en terme de caténation entre les deux chromosomes. Ainsi les liens de caténation accumulés tout au long de la répllication seraient "poussés" vers la dernière région à être répliquée (*TRZ*) donnant ainsi lieu à de l'hyper-recombinaison (Hill, 1996). Il est difficile d'entrevoir un avantage à cette hyper-recombinaison et il est fort probable que cette caractéristique ne soit que la conséquence de la répllication d'un génome circulaire (Hill, 1996).

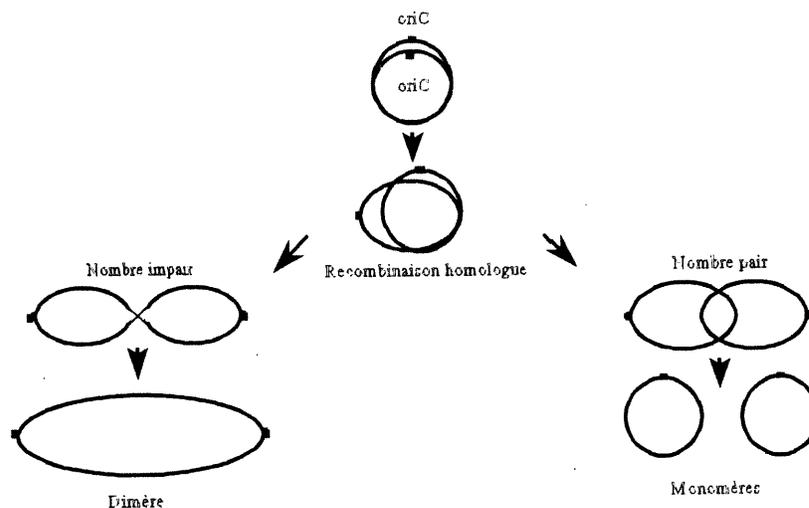
### D- La séparation des chromosomes

La séparation normale des chromosomes peut être bloquée par deux mécanismes indépendants qui engendrent la formation de liens entre les chromosomes. Ainsi, la caténation des deux nouveaux chromosomes est fréquente au niveau du terminus et est solutionnée par la Gyrase et la Topoisomérase IV (Rothfield, 1994). De plus, le phénomène de la recombinaison homologue en nombre impair produit des chromosomes liés sous forme de dimères (Kuempel, 1991; Blakely *et al*, 1993; Leslie & Sherratt, 1995).

La figure I illustre ce dernier mécanisme. La dimérisation peut aussi se produire entre plasmides multicopies (Summers, 1991; 1993).

Figure 1

#### Mécanisme de dimérisation des chromosomes par recombinaison homologue



Si, lors de la division cellulaire, les réplicons sont sous la forme di ou multimérique,

le matériel génétique ne sera pas distribué de façon équitable entre les cellules-filles (Blakely *et al*, 1993). Ceci mène à des problèmes de stabilité de population lorsqu'il s'agit de plasmides et à des problèmes de filamentation cellulaire et d'aberration génétique s'il s'agit de chromosomes (Sherratt *et al*, 1993). Il est donc important que les réplicons soient sous la forme monomérique avant leur ségrégation vers les pôles cellulaires (Blakely *et al*, 1993). Le système de recombinaison spécifique du site Xer résout les multimères et permet ainsi une séparation et une ségrégation normale des chromosomes dans les cellules-filles.

### **Les systèmes de recombinaison spécifique de site**

Il existe deux grands groupes de systèmes de recombinaison spécifique de site, soit le groupe des resolvases-invertases et celui des  $\lambda$ -intégrases. Avant de nous pencher sur le système qui nous concerne, soit le système  $\lambda$ -intégrase Xer, nous allons rapidement passer en revue les différences et les ressemblances majeures entre les deux grands groupes de recombinaisons.

La recombinaison spécifique de site se distingue de la recombinaison homologue par le fait que les brins sont échangés non pas grâce de la reconnaissance de fortes homologies mais plutôt par la reconnaissance de très courtes séquences spécifiques. De plus cette recombinaison spécifique de site est dite conservatrice puisqu'il n'y a ni synthèse ni dégradation d'ADN (Stark *et al*, 1992).

La recombinaison spécifique de site sert la fonction de réorganisation de l'ADN chez un grand nombre d'espèces eucaryotes et procaryotes. Par exemple chez les bactéries on retrouve le système Hin de *Salmonella* qui permet l'expression des gènes de flagelline par

inversion d'une séquence d'ADN (Sadowski *et al*, 1986). Chez les virus, on retrouve le système LoxP/Cre du phage P1 qui permet le maintien du phage dans la cellule via la résolution par excision de dimères (Kilby *et al*, 1993). Chez les levures, le système FIp/Frt permet la production de copies multiples de plasmides par inversion de segments d'ADN (Kilby *et al*, 1993). L'ADN est excisé si les sites sont orientés dans la même direction alors qu'il est inversé si les sites sont orientés dans des directions opposées (Sadowski, 1986).

Les recombinases du type résolvas-invertases sont plus petites que les recombinases de type  $\lambda$ -intégrases, elles sont composées de deux domaines fonctionnels, soit un domaine d'arrimage à l'ADN (C-terminal) et un domaine contenant le site catalytique de la protéine (N-terminal). Chez les  $\lambda$ -intégrases, il n'y a qu'un domaine fonctionnel contenant le site catalytique au bout C-terminal de la protéine (Sherratt *et al*, 1993).

On observe également que les résolvas-invertases sont beaucoup plus conservées que les  $\lambda$ -intégrases. De plus, lors de l'activité de coupure et d'échange des brins, les résolvas-invertases utilisent une sérine conservée dans le site catalytique, alors que chez les  $\lambda$ -intégrases il s'agit d'une tyrosine (Sherratt *et al*, 1993). Le tableau I présente les caractéristiques majeures des résolvas-invertases et des  $\lambda$ -intégrases.

Tableau I

Caractéristiques majeures des recombinases résolvases-invertases et  $\lambda$ -intégrases

CARACTERISTIQUES	RESOLVASES-INVERTASES	$\lambda$ -INTEGRASES
Exemple de systèmes	Réolvases : Tn3, $\gamma\delta$ Invertases: P1, Mu, e14, <i>Salmonella typhimurium</i>	$\lambda$ , plasmide 2 $\mu$ m, Fimbriae, résolution des réplicons
Facteurs accessoires	Réolvases: aucun Invertases: Fis agissant au site <i>sis</i>	Facteurs sont fréquemment utilisés
Site de recombinaison	Réolvases: de 100 à 130 pb Invertases: 30 pb	De 30 à 240 pb
Site d'enjambement	Environ 30 pb	Environ 30 pb
Région centrale du site	2 pb	6 à 8 pb
Acide aminé catalytique conservé	Sérine	Tyrosine

Tableau tiré de la maîtrise de Stéphane Sirois ("Détection des recombinases XerC et XerD chez les *Enterobacteriaceae*" : mémoire présenté au département de microbiologie et immunologie de l'Université de Montréal en 1994).

**Le système de recombinaison spécifique de site XerC-D de *E. coli***

Puisque la division cellulaire normale exige des formes chromosomiques monomères, il est logique de penser que l'évolution ait mis en place un système assurant la résolution des formes di ou multimérique. Ce système est celui de la recombinaison spécifique de site Xer: des enzymes de recombinaison spécialisées reconnaissent des sites spécifiques sur l'ADN, les rapprochent, puis clivent, échangent et recollent les brins d'ADN rendant ainsi la molécule dimérique à ses formes monomères originales (Stark *et al*, 1989).

Dans le cas de *E. coli*, les enzymes XerC et XerD du système de recombinaison Xer agissent au niveau du site génomique *dif*. Une mutation dans l'une ou l'autre de ces enzymes ou sur le site *dif* provoquera une hausse de la proportion de cellules filamenteuses et/ou génétiquement aberrantes dans la population (Blakely *et al*, 1991; Kuempel, 1991; Blakely

*et al*, 1993). Lorsque la molécule dimérique est un plasmide, ces mêmes recombinaisons agissent de façon concurrente à des protéines accessoires (ArgR et PepA) sur le site de recombinaison plasmidique (*cer* du plasmide ColE1) (Summers et Sherratt, 1988; Summers, 1989).

#### *A- Les recombinaisons XerC et XerD*

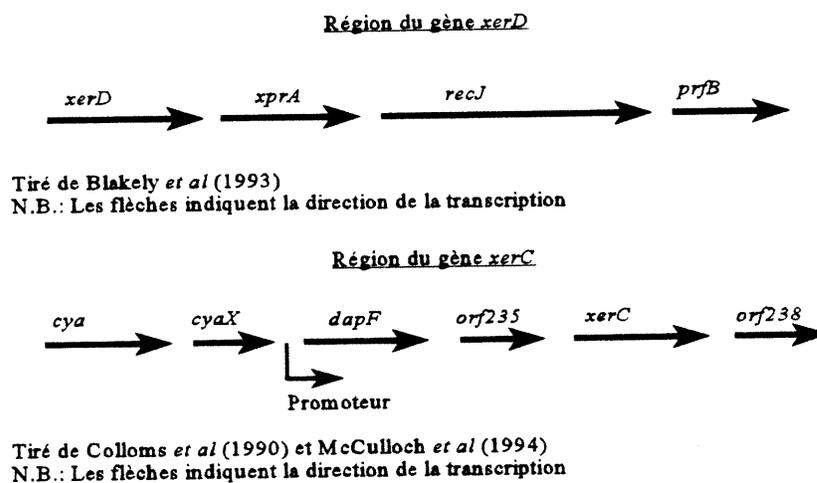
Les protéines XerC et XerD sont des membres de la famille des recombinaisons  $\lambda$ -intégrases (Stark, 1992). Cette famille de recombinaisons sert à l'excision ou à l'inversion de segments d'ADN jouant des rôles très variés.

Les recombinaisons  $\lambda$ -intégrases sont caractérisées par une longueur variant entre 280 et plus de 350 acides aminés ainsi que par de fortes homologies dans les domaines I et II de leur partie C-terminale. Cependant on ne trouve que 4 acides aminés complètement conservés chez toutes les lambda-intégrases, soit une arginine dans le domaine I et une arginine, une histidine et une tyrosine dans le domaine II (voir l'annexe III)(Colloms *et al*, 1990; Blakely *et al*, 1993; Sherratt *et al*, 1993; Colloms, 1996). Ces résidus jouent un rôle important dans l'échange des brins, ainsi la tyrosine du domaine II serait impliquée dans le clivage initial de l'ADN par la protéine (attaque phosphodiester), alors que les résidus du domaine I seraient impliqués dans le clivage et le transfert des brins(Colloms, 1996).

La séquence en acides aminés des protéines XerC et XerD est homologue à 37 % (Blakely *et al*, 1993; Sherratt *et al*, 1993). XerC est codée par une séquence de 1.2 kpb (Colloms *et al*, 1990; Blakely *et al*, 1991) située à la minute 85 sur le chromosome de *E. coli*, soit très près d' *oriC*. Son poids moléculaire est de 33.8 Kd (Colloms *et al*, 1990).

XerD située à la minute 63 (Lovett et Kolodner, 1991), est codée par une séquence de 1.5 Kpb et a un poids moléculaire similaire à XerC (Blakely *et al*, 1993). Ces deux enzymes agissent en coopération sur le site génomique *dif*. La figure 2 présente les cartes des régions entourant les gènes *xerC* et *xerD* de *E. coli*.

Figure 2  
Cartes des régions entourant les gènes *xerC* et *xerD*



### B- Le site génomique *dif*

Le site *dif* (*deletion induced filamentation*) est situé à la minute 33.6 sur la carte génétique du chromosome de *E. coli* soit entre les sites *TerA* et *TerC* (Kuempel *et al*, 1991, Leslie & Sherratt, 1995). Le site *dif* comprend 28 pb dont une région centrale de 6 pb flanquée de chaque côté par des séquences de 11 pb (Cornet *et al*, 1994). Le côté droit du segment appelé “bras droit” lie la protéine XerD, alors que le côté gauche, le “bras gauche”, lie la protéine XerC (Sherratt *et al*, 1993; Blakely *et al*, 1993; ). Quoique présentant une séquence symétrique de 6 pb, les deux “bras” sont néanmoins différents et se distinguent

notamment par le fait que le bras gauche est systématiquement moins conservé que le bras droit. Ceci a un effet sur l'arrimage des protéines à l'ADN, beaucoup moins fort avec XerC qu'avec XerD. Les bras du site *dif* présentent de fortes homologies avec ceux des sites *cer* (ColE1), *psi* (pSC101), *clf* (CloDF13), ColK et ColE3 (Blakely & Sherratt, 1994). La figure 3 présente les séquences en pb de ces sites.

Figure 3

Séquences en pb des sites *dif*, *cer*, *psi*, *clf*, Colk et ColE3

Site	XerC	Région centrale	XerD
<i>cer</i>	5' GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT 3'
<i>psi</i>	5' GGTGCGCGCAA	GATCCA	TTATGTAAAC 3'
<i>clf</i>	5' GGTACCGATAA	GGGATG	TTATGGTAAAT 3'
Colk	5' GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT 3'
ColE3	5' GGTGCGTACAA	CGGGAG	TTATGGTAAAT 3'
<i>dif</i>	5' GGTGCGCATAA	TGTATA	TTATGTAAAT 3'

Tableau tiré de Blakely et Sherratt (1994) p.5615

La séquence en pb de la région centrale ne semble pas jouer de rôle significatif dans le fonctionnement du système. Des mutations peuvent donc être introduites sans conséquences dans la région centrale à la condition que les deux sites à recombinaison contiennent les mêmes changements. En effet, deux sites (par exemple *cer* et *dif*) dont les régions centrales ne sont pas homologues ne pourront se recombinaison. Plutôt que la

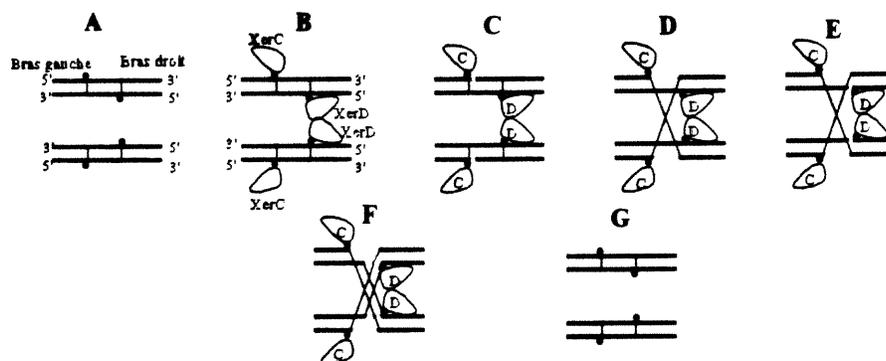
séquence, il semblerait que ce soit le nombre de pb dans cette région qui en influence la fonction (McCulloch *et al*, 1994). De plus, ce nombre de pb influence également la direction de la recombinaison spécifique de site. Ainsi une région de 6 pb (*dif*) donnera lieu à de la recombinaison non sélective, c'est-à-dire autant inter qu'intra-moléculaire, alors qu'une région de 8 pb (par exemple le site *cer* du plasmide ColE1) associée aux facteurs accessoires donnera lieu à de la recombinaison sélective intramoléculaire seulement (Blakely *et al*, 1993; Sherratt *et al*, 1993).

La délétion du site *dif* entraîne l'apparition de cellules filamenteuses après plusieurs générations. Cette filamentation serait en partie causée par l'induction du système SOS. Des expériences de délétion du site *dif* utilisant des mutants de suppression du système SOS montrent que la filamentation est quand même présente. Ceci démontre que la délétion de *dif* à elle seule est suffisante pour induire le phénotype filamenteux. La proportion de cellules filamenteuses dans ces populations est d'environ 10 % ce qui correspond à la fréquence estimée d'échange entre chromatides soeurs (Kuempel *et al*, 1991). De plus, des expériences sur la translocation de *dif* à d'autres positions sur le chromosome, induit l'apparition du phénotype filamenteux indiquant donc que la recombinaison spécifique de site doit se faire dans la région du terminus pour la ségrégation chromosomique normale (Cornet *et al*, 1994). Les études de Leslie et Sherratt (1995) montrent également que l'orientation inverse de *dif* dans le terminus n'affecte pas la ségrégation chromosomique.

## Mécanismes d'action du système XerC-D

La plupart des expériences sur le fonctionnement du système Xer ont été faites en utilisant le site *cer*, *psi* ou le site *dif*, inséré dans un plasmide. En général, les résultats obtenus pour l'un ou l'autre de ces sites seront donc considérés comme équivalents puisque les règles régissant l'action des enzymes XerC ou XerD sont sensées être les mêmes peu importe le site ciblé. La figure 4 illustre l'activité des recombinases sur le site *dif*.

Figure 4  
Mécanisme d'action des recombinases XerC et XerD sur le site *dif*



La première étape dans l'activité du système est la formation du lien entre l'ADN et les protéines XerC et XerD (figure 4-B). Chacun des deux sites reçoit deux monomères des recombinases, soit un monomère XerC pour le bras gauche et un monomère XerD pour le bras droit (Sherratt *et al*, 1993). L'ancrage à l'ADN requiert la coopération des deux enzymes. Ainsi la protéine XerD s'accroche la première et aide l'arrimage de XerC par modification de la conformation tridimensionnelle de l'ADN (Blakely, 1993; Blakely & Sherratt, 1994). Ce n'est que lorsque les quatre monomères sont fixés et que les deux sites

sont correctement alignés que la recombinaison peut se faire.

La deuxième étape dans le mécanisme est le clivage et l'échange des brins d'ADN (figure 4-C à 4-F). C'est ici qu'interviennent les résidus catalytiques conservés des recombinases. Les deux résidus R et le résidu H interviennent dans le clivage et la ligation, alors que le résidu Y du bout C-terminal devient le nucléophile qui sera lié de façon covalente à l'ADN lors de la réaction d'échange des brins (Stark *et al*, 1992; Colloms *et al*, 1996). Des expériences *in vitro* sur le site *cer* montrent que le clivage des brins se fait entre le dernier nucléotide du site de l'arrimage protéique et le premier nucléotide de la région centrale (Arciszewska & Sherratt, 1995). De plus, Colloms *et al* (1996) montrent que pour le site *psi* le premier échange se fait via XerC entre les deux brins 5'---3' de chaque site (figure 4-C) alors que XerD catalyse le deuxième échange, soit celui des brins 3'---5' (figure 4-E) (Colloms *et al*, 1996). Les deux paires d'échanges seraient temporellement et spatialement séparées par un intermédiaire jonction de Holliday (Arciszewska *et al*, 1995). Un résultat surprenant ressort cependant des études de Colloms *et al* (1996): contrairement à ce que l'on croyait, l'enzyme XerD ne serait pas impliquée dans le deuxième échange de brins des sites *cer*.

Le site *dif*, contrairement à ses homologues plasmidiques, ne montre pas de sélectivité de résolution intra-moléculaire lorsque cloné dans un vecteur. Si *dif* occasionne de la recombinaison inter-moléculaire on peut alors se demander pourquoi il n'y a pas de filamentation cellulaire dans les populations sauvages. Leslie et Sherratt (1995) avancent l'hypothèse que la formation d'une jonction de Holliday entre deux chromatides soeurs lors de la ségrégation créerait un changement de conformation local qui induirait un événement

de résolution libérant ainsi les chromosomes monomères.

### **Le système XerC-D chez les plasmides**

Un grand nombre de plasmides (R1, ColE1, ColK, CloDF13, pSC101, ColA, pMB1, NPT16) possèdent des séquences analogues à *dif* qui permettent la résolution de dimères ou multimères en monomères via les recombinases XerC et XerD (Cornet *et al*, 1994). L'utilisation d'un tel système par les plasmides leur permet une certaine stabilité de population à travers les générations. Nous allons brièvement décrire le système plasmidique ayant été le plus étudié, soit celui du plasmide ColE1 (site *cer*).

La principale différence entre plasmides et chromosomes vient du fait que la recombinaison spécifique de site du plasmide nécessite non seulement les enzymes Xer mais aussi les protéines accessoires ArgR (le répresseur arginine) et PepA (l'aminopeptidase A) (Stirling *et al*, 1988). La recombinaison au site *cer* est également hautement sélective, ne permettant que la recombinaison intra-moléculaire.

Le site plasmidique comprend une séquence de 280 pb de long. En plus des sites d'ancrage des recombinases, le site comporte une séquence de 250 pb nécessaire aux facteurs accessoires ArgR et PepA (Summers et Sherratt, 1988; Kilby *et al*, 1993). Les séquences d'arrimage des recombinases sont analogues à celles du chromosome sauf en ce qui a trait au nombre de paires de bases dans la région centrale, celui-ci est de 8 dans *cer* alors qu'il est de 6 dans *dif*.

### Système analogue à Xer chez d'autres bactéries

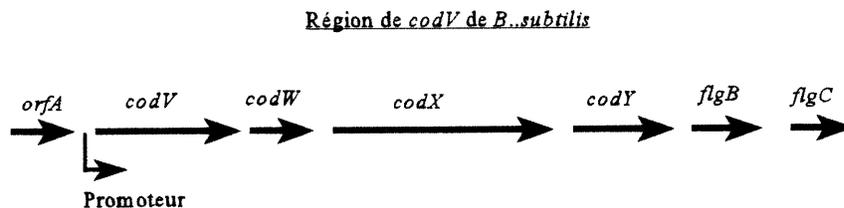
Plusieurs recherches ont mis en évidence, directement ou indirectement, la présence d'un système analogue à Xer chez plusieurs espèces bactériennes. Ainsi, Höfte *et al* (1994) ont découvert un gène analogue à *xerC* chez *Pseudomonas aeruginosa*. Ce gène, appelé *sss*, produit une protéine dont la séquence présente une identité de 50 % avec XerC. De plus le gène en amont de *sss* présente des homologies avec le gène de la diaminopimelate épimérase (*dapF*) qui, chez *E. coli*, fait partie du multicistron contenant *xerC*. Ces mêmes chercheurs ont également mis en évidence la complémentation de mutants XerC<sup>-</sup> de *E. coli* par la protéine Sss et vice-versa chez les mutants Sss<sup>-</sup> de *P.aeruginosa*. Les études de Fleischmann *et al* (1995) montrent que l'espèce *Haemophilus influenzae* possèdent également des séquences codant pour des recombinaisons homologues à XerC et XerD de *E. coli*.

Les études de Sirois et Szatmari (1995) sur des bactéries Gram négatives de la famille des entérobactériacées ont démontré par hybridation Southern la présence de recombinaisons analogues à XerC et XerD chez *Morganella morganii*, *Salmonella typhimurium*, *Serratia liquefaciens*, *Shigella flexneri* et *Klebsiella pneumoniae*. Ces auteurs ont également mis en évidence la présence d'activité recombinatoire de type Xer chez ces mêmes espèces.

Dans le groupe des gram positifs peu de travail a été fait. Chez *Lactobacillus leichmannii* Becker et Brendel (1996) ont mis en évidence un gène présentant une identité de 38.8 % à *xerC*. Chez *Bacillus subtilis*, Slack *et al* (1995) ont trouvé un gène, *codV*, dont la protéine présenterait une identité de 39% avec XerD. Un fait intéressant ressort

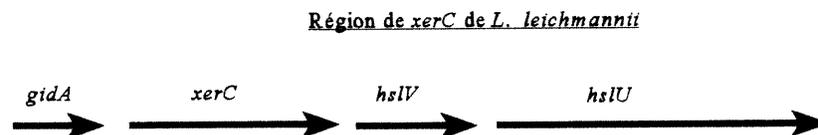
également de ces études : chez ces deux espèces, les gènes des recombinares se retrouvent adjacents à des gènes codant pour des protéines de chocs thermiques (“Heat Shock Proteins”). La figure 5 présente les cartes des régions entourant les gènes *xerC* de *L. leichmannii* et *codV* de *B.subtilis*.

Figure 5  
Cartes des régions entourant les gènes *xerC* de *L. leichmannii* et *codV* de *B.subtilis*



Tiré de Slack et al (1995)

N.B.: les flèches indiquent la direction de la transcription  
Les gènes *codW* et *codX* sont les gènes de choc thermique



Tiré de Becker et Brendel (1996)

N.B.: les flèches indiquent la direction de la transcription  
Les gènes *hslV* et *hslU* sont les gènes de choc thermique

### Projet de recherche

Dans ce mémoire nous tenterons de vérifier si un système analogue au système Xer (recombinares XerC et XerD et site *dif*) est présent chez 13 espèces de bactéries gram positives du groupe des bactéries lactiques, plus particulièrement des genres *Lactobacillus*

et *Lactococcus*. Certaines de ces bactéries sont très importantes dans les industries de produits laitiers. L'espèce *Streptococcus salivarius* a aussi été étudiée puisque des recherches récentes (Moyer *et al*, 1996) portant sur des gènes de la sous-unité rRNA l'ont classée très proche parente de l'espèce *Lactobacillus casei*. Les principales techniques que nous avons utilisées sont l'hybridation Southern, l'amplification génique, le clonage et le séquençage. De plus, nous avons utilisé un programme informatique permettant de comparer les différentes séquences et d'en dégager d'éventuelles phylogénies.

## CHAPITRE 2

### MATERIEL ET METHODES

#### MATÉRIEL

##### Souches bactériennes utilisées

##### *A- Bactéries gram positives*

Les espèces de bactéries lactiques gram positives utilisées ont été choisies en fonction de leurs disponibilité. Le tableau II en présente la liste et certaines caractéristiques particulières.

Tableau II  
Liste des espèces et souches gram positives utilisées

ESPECE	SOUCHE/PROVENANCE	CARACTERISTIQUES
1- <i>Lactobacillus bulgaricus</i>	INRA (France) no 737	
2- <i>Lactobacillus caseii</i>	Souche de l'Uni. de Montréal	Pousse en aérobie
3- <i>Lactobacillus fermentii</i>	INRA (France) BP83	Guérie de plasmides
4- <i>Lactobacillus fermentii</i>	INRA (France) 1.16 A	
5- <i>Lactobacillus plantarum</i>	ATCC 8014	
6- <i>Lactobacillus plantarum</i>	ATCC 14917	
7- <i>Lactococcus lactis</i> ssp <i>cremoris</i>	AM2. Souche de l'Uni. Laval	
8- <i>Lactococcus lactis</i> ssp <i>cremoris</i>	H22. Souche de l'Uni. Laval	
9- <i>Lactococcus lactis</i> ssp <i>cremoris</i>	INRA (France) MG 1363	Guérie de plasmides
10- <i>Lactococcus lactis</i> ssp <i>lactis</i>	INRA (France) no 763	
11- <i>Lactococcus lactis</i> ssp <i>lactis</i>	INRA (France) no 1667	
12- <i>Lactococcus lactis</i> ssp <i>lactis</i>	INRA (France) IL 1403	Guérie de plasmides
13- <i>Streptococcus salivarius</i>	ATCC 13419	

### *B- Bactéries Gram négatives*

1-*Escherichia coli* DH5  $\alpha$  préparé pour la transformation chimique (*endA1 hsdR17 r<sup>-</sup>m<sup>+</sup> supE44 thi-1 recA1 gyrA nal<sup>r</sup> relA1  $\Delta$ (lacZYA-arg<sup>f</sup>) U169 lacZ $\Delta$ M15*)

2-*Escherichia coli* chimio-compétentes Novablue lot no 1191 de la Cie Novagen (*endA1 hsdR17 (r<sub>k12</sub><sup>-</sup> m<sub>k12</sub><sup>+</sup>) supE44 thi-1 recA1 gyrA96 relA1 lac[F proA+B+lac<sup>I</sup> Z $\Delta$ M15:: Tn10 (tet<sup>r</sup>)*]

### **Vecteurs utilisés**

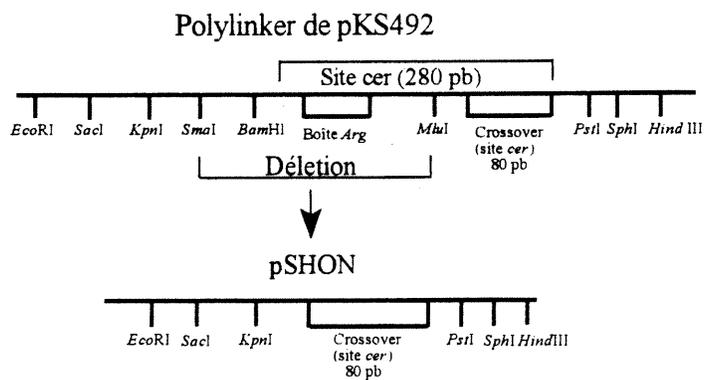
#### *A- Vecteurs utilisés pour le séquençage:*

Le vecteur pT7blue (R) de la Cie Novagen a été utilisé pour les ligations d'inserts à séquencer. Ce vecteur, issu du plasmide pUC19, permet le clonage de fragment purifiés provenant d'amplifications géniques et leur séquençage subséquent. Le vecteur pTZ/PC conçu par D. Tessier et D. Thomas de l'Institut de Recherche en Biotechnologie du Conseil National de Recherches du Canada a également été utilisé lors de séquençage. Ce vecteur issu du plasmide pTZ19R a été conçu pour cloner directement des fragments d'ADN amplifié par la polymérase Taq et pour leur séquençage (contient l'origine de réplication du bactériophage  $\phi$ 1) (Tessier et Thomas, communication personnelle).

#### *B- Vecteur utilisé pour les hybridations Southern:*

Le vecteur pKS492 issu du pUC18 a également été utilisé. Il contient le site plasmidique *cer* inséré dans le "polylinker". Ce vecteur a été digéré avec les enzymes *Sma* I (NEB) et *Mlu* I (NEB) afin d'éliminer le site de fixation de ArgR d'environ 200 pb. Le vecteur a ensuite été religué. Le plasmide résultant (pSHON) contient un site *cer* de 80 pb contenant une séquence homologue à *dif*. La figure 6 présente la construction du pSHON.

Figure 6  
Vecteur pSHON issu du vecteur pKS492



Le vecteur pSHON a été utilisé lors des expériences de reconnaissance du site *dif* de *Lactobacillus caseii* par hybridation Southern. Les vecteurs pTZ/PC contenant les inserts séquencés de *Lactobacillus caseii* et de *Lactobacillus plantarum* ATCC 14917 ont également été utilisés comme sondes dans certaines hybridations Southern.

#### C- Vecteurs pPS22, pSD105 et pRM130

Le vecteur pPS22 construit par Slack *et al* (1995) à un poids moléculaire de 7 Kpb contient la séquence de 1018 pb du gène *codV* de *Bacillus subtilis*. Le vecteur pSD105, construit par Colloms et collaborateurs (1990), contient la séquence de 1.2 kpb du gène *xerC* de *E.coli*. Enfin, le vecteur pRM130 (Blakely *et al*, 1993) contient la séquence de 1.5 kpb du gène *xerD* de *E.coli*. Ils ont tous servi pour tester les amorces A1 et C qui ont été utilisées lors de certains PCR.

## MÉTHODES

### Cultures bactériennes

Les bactéries du genre *Lactobacillus* et *Lactococcus* ont été incubées dans du bouillon MRS (de Man, Rogosa, Sharpe) de la compagnie Oxoid, additionné de lactose à 2%, à 30 C° en jarre anaérobie pour une durée de deux jours.

Les bactéries du genre *Streptococcus salivarius* ont été incubées dans du bouillon BHI (Brain & Heart Infusion de la compagnie Oxoid), à 37 C° pour toute une nuit.

Selon les expériences les différentes souches de *Escherichia coli* ont été incubées soit dans du bouillon LB, du bouillon SOB ou du bouillon TB (Terrific Broth) de DIFCO, additionné ou non d'ampicilline (50µg/mL), à 37 C° pour toute une nuit.

### Les extractions d'ADN génomique des bactéries lactiques

Après plusieurs essais nous nous sommes aperçus que les procédés habituels d'extraction d'ADN pour les bactéries gram négatives et gram positives ne fonctionnaient pas avec les bactéries lactiques, entre autre parce que celles-ci possèdent une paroi très riche en sucres résistant à de nombreux procédés de lyse. Il a donc été nécessaire de mettre au point un nouveau protocole d'extraction d'ADN:

Après incubation nocturne en jarre anaérobie, des cultures de 5 mL ont été resuspendues dans une solution de 25 % de glycine dans du bouillon MRS pour une durée de 3 heures à 30 C° en anaérobie (tubes vissés très pleins). La glycine en haute concentration inhibe la formation de la paroi cellulaire (Holo et Nes, 1989). Par la suite les

cultures ont été centrifugées à 14000 RPM pour ensuite être lavées dans 500  $\mu$ L de 100mM Tris pH8 (Maniatis *et al*, 1982). Le culot a été incubé à 37° C x 60 min. dans 200  $\mu$ L de glucose à 25 % dans du TE, contenant également 10  $\mu$ L de mutanolysine 1 U/mL (SIGMA) et 100  $\mu$ L de lysozyme 50 mg/mL(Boehringer/Mannheim). L'ADN a ensuite été extrait selon le protocole et en utilisant les produits d'extraction Genomix de la compagnie Talent.

### **Les extractions d'ADN des plasmides- vecteurs**

L'ADN des vecteurs a été extrait avec les produits d'extraction Plasmix de la compagnie Talent. Les extractions d'ADN plasmidique vouées aux séquençages faits dans notre laboratoire ont été faites par lyse alcaline. Le protocole est le suivant : des cellules DH5 alpha transformées avec le vecteur pT7-blue contenant l'insert à séquencer ont été inoculées et incubées dans du bouillon LB à 37 C° pour toute une nuit. On a recueilli 1.5 mL du bouillon qu'on a centrifugé dans un tube Eppendorf. Le surnageant a été enlevé et le culot lavé avec 500  $\mu$ L de TEN (Maniatis, *et al*, 1982) puis recentrifugé. Une fois le surnageant enlevé, le culot a été resuspendu dans 50  $\mu$ L de TE-sucrose (50mM Tris pH 7.5, 10% sucrose, 10mM EDTA, 0.1% Triton X100 de la Compagnie BioRad et 2 mg/ mL de lysozyme de la compagnie Boehringer Mannheim). Après un mélange au vortex de quelques secondes, on a laissé incubé 5 minutes à la température de la pièce puis on a rajouté 200  $\mu$ L d'une solution de NaOH 0.2 N /SDS 1% (SDS de la compagnie BioRad). Le tout a été incubé à 37 C° pendant 15 minutes. On a ensuite neutralisé avec 150  $\mu$ L de 5 M KOAc froid. Le tube a été mélangé par inversion puis incubé 10 min sur glace. On a par la suite centrifugé à 14 000 RPM pendant 5 minutes à 4 C°. Le surnageant a été recueilli puis on a

procédé à une extraction phénol-chloroforme et une précipitation à l'éthanol tel que décrit dans Maniatis *et al* (1982). Le culot d'ADN obtenu a alors été resuspendu dans 60  $\mu$ L de TE avec 20  $\mu$ g/mL de RNase. Une journée avant le séquençage on a rajouté 5  $\mu$ L d'une solution de 2M NaOH/2mM EDTA à l'ADN et on a laisser incuber à 37 C° pendant 15 minutes. On a ensuite ajouté 5  $\mu$ L de 2mM d'acétate de sodium glacé et 150  $\mu$ L d'éthanol à 95 %. Le tube a été agité pendant quelques secondes au vortex puis placé à -70 C° pendant 20 minutes. Le tube a ensuite été centrifugé à 14 000 RPM pendant 10 minutes. Le surnageant a été enlevé puis le culot séché à l'air. On a finalement resuspendu dans 40  $\mu$ L d'eau milliQ stérile.

#### **Enzymes de restriction utilisées.**

Lorsque des digestions d'ADN étaient nécessaires, les enzymes de restriction utilisées et leurs tampons étaient ceux de la compagnie New England Biolabs et les protocoles suivis étaient ceux prescrits par cette compagnie.

#### **Electrophorèses**

Les électrophorèses ont été faites en utilisant de l'agarose de la compagnie ICN, dissous dans du tampon TAE fait selon le protocole de Maniatis (1982). Les concentrations des gels d'électrophorèse ont varié de 0.7% à 0.9%, les plus hautes concentrations étant réservées pour séparer les fragments d'ADN de petites tailles (produits de PCR). Les appareils d'électrophorèse utilisés sont de marque BRL. Deux formats ont été utilisés: le format 6.5 cm par 8.5 cm contenant 8 puits, et le format 11.5 cm par 14 cm, contenant 20

puits. Le colorant utilisé pour la migration de l'ADN était du bleu de bromophénol à .4% dans une solution de glycérol à 25%. La proportion de colorant utilisée était d'environ 2  $\mu\text{L}$  pour chaque 10  $\mu\text{L}$  de solution totale. Après migration en moyenne à 50 volts pour une heure, les gels étaient trempés pendant quelques minutes dans une solution de bromure d'éthidium (concentration = 0.005  $\mu\text{g}/\text{mL}$ ) puis rincés à l'eau avant d'être photographiés à la lumière U.V. (312 nm, UVP) avec un appareil Polaroid MP.4 et un film Polaroid 557.

## **Clonage**

### *A-Ligations*

Pour les ligations dans le vecteur pT7-blue le protocole suivi était celui de la compagnie Novagen. Pour toutes les autres ligations sauf celles qui ont été utilisées lors des PCR inverses, le protocole de ligation utilisé était le suivant: l'ADN de l'insert et l'ADN du vecteur ont été mis en proportion 3:1 en tenant compte de la concentration et du poids moléculaire. On ajoutait ensuite de 10 à 12 U de la T4 DNA ligase de la compagnie NEB et le tampon fourni par cette même compagnie dans une proportion 1:10. Les ligations ont été incubées toute une nuit à 16 C°.

Les protocoles de ligation ayant servi lors des PCR inverses sont expliqués dans la section sur l'amplification génique.

### *B-Transformations par électroporation*

Les cellules électrocompétentes ont été préparées comme suit: 5 mL de bouillon LB ont étéensemencés avec des cellules DH5 $\alpha$  et incubé à 37 C° pendant la nuit avec agitation.

Le lendemain matin, 500 $\mu$ L de cette culture a été utilisé pour ré-ensemencer 5 mL de bouillon LB neuf. La densité optique (D.O. à 550 nm) a été ensuite vérifiée régulièrement. Lorsque la D.O. obtenue était entre 0.5 et 1.0, les cellules étaient mises sur glace puis centrifugées à 4°C à 9000 RPM pendant 15 min. Le culot a ensuite été resuspendu dans 1 volume d'eau stérile froide, puis a été centrifugé à 4°C pendant 15 min à 9000 RPM. Ce même procédé a été répété trois fois, en réduisant le volume de resuspension de moitié à chaque fois. La dernière resuspension a été faite dans 1 mL de glycérol froid stérile à 10 %. Les cellules ont ensuite été conservées à -70°C. Le protocole utilisé pour la transformation par électroporation est le suivant. Quarante  $\mu$ L de cellules DH5 alpha électrocompétentes ont été mélangées sur glace dans une cuvette à électroporation avec 2  $\mu$ L du vecteur pSHON. Après administration de 2.5 KV pendant 4.5 millisecondes dans un appareil Bio Rad *E.coli* Pulser, la solution a été transférée dans un tube Eppendorf contenant 1 mL de bouillon SOC, puis le tout a été incubé avec agitation à 37 C° pendant 1 heure. Après cette période, 100  $\mu$ L de la solution a été étalé sur une boîte de pétri contenant du LB Ap.

#### *C-Transformation chimiques*

Le vecteur pT7-blue ayant servi lors du séquençage à été transformé dans les souches de cellules chimio-compétentes Novablue de la compagnie Novagen. Les protocoles suivis pour transformer les cellules sont ceux de cette compagnie.

Le vecteur pTZ-PC ayant servi lors de certains séquençages a été transformé dans des souches d'*Escherichia coli* DH5  $\alpha$  chimio-compétentes issues de notre laboratoire. Le protocole de transformation suivi était celui décrit par Maniatis *et al* (1982).

### **Les amplifications géniques (PCR)**

L'appareil d'amplification génique utilisé est le modèle Combi Thermal Reactor TR2 de la compagnie Hybaid. Plusieurs programmes ont été utilisés selon le type d'amorce et l'ADN ciblé.

Nous nous sommes servi de l'amplification génique ou PCR (Polymerase Chain Reaction) de deux façons différentes. Selon la première façon, l'ADN amplifié pouvait être la séquence interne d'un gène ou d'un insert dans un vecteur. Selon la deuxième façon, une séquence d'ADN connue était utilisée pour produire des amorces allant dans des directions opposées à la séquence interne originale. Ce deuxième type de PCR, appelé PCR "inverse", permet d'amplifier les séquences adjacentes à la séquence connue. Dans notre cas il s'agissait donc de circulariser un fragment d'ADN contenant la séquence connue flanquée des séquences inconnues nous intéressant. Le protocole de ligation pour les produits qui ont été utilisés dans ces PCR a été fourni par Dr D. Leclerc de l'université McGill. Puisque dans ce cas-ci il s'agissait de favoriser les autoligations, il était important que l'ADN à circulariser soit en faible concentration. Trois dilutions aux concentrations de plus en plus faible d'ADN ont été utilisées lors des ligations précédant les PCR inverses. Deux  $\mu\text{g}$  d'ADN ont d'abord été digérés avec 50 U d'enzyme de restriction avec son tampon correspondant dans un volume total de 50  $\mu\text{L}$  pendant toute une nuit. On a ensuite fait une extraction avec 1 volume de phénol-chloroforme puis l'ADN a été précipité en ajoutant 5  $\mu\text{L}$  de 5M NaCl et 250  $\mu\text{L}$  d'EtOH 95 % froid. Le tube a été mis à - 70 C° pendant 20 à 30 minutes puis nous avons centrifugé à 14000 RPM pendant 10 minutes à 4°C. Le culot

obtenu a ensuite été séché à l'air puis resuspendu dans 50  $\mu\text{L}$  d'eau déionisée stérile. Trois tubes ont été préparés, tous contenaient 1  $\mu\text{L}$  (400 U) de T4 DNA ligase de NEB et 10  $\mu\text{L}$  de tampon de ligation 5X de BRL, les concentrations de l'ADN à l'intérieur des trois tubes étaient les suivants : le premier contenait 5  $\mu\text{L}$  d'ADN et 30  $\mu\text{L}$  d'eau déionisée stérile; le deuxième tube contenait 10  $\mu\text{L}$  d'ADN et 25  $\mu\text{L}$  d'eau déionisée stérile; le troisième tube contenait 35  $\mu\text{L}$  d'ADN sans ajout d'eau. Deux  $\mu\text{L}$  de chacune de ces solutions ont été utilisés dans 3 PCR inverses indépendants.

#### *A- Conditions des réactions d'amplification de l'ADN*

Les différentes conditions pour l'amplification génique ont été mises au point par essai-erreur. La principale différence entre les conditions utilisées se situe au niveau de la température optimale d'hybridation des amorces à la matrice d'ADN (minimum= 37 C°, maximum= 54 C°). Celle-ci était relativement faible lorsqu'il s'agissait d'utiliser des oligonucléotides dégénérés ou lorsque la similarité entre les amorces et l'ADN était inconnue ou réputée faible. Il a également fallu tenir compte des différents points de fusion de l'ADN des amorces. Le temps alloué pour l'hybridation était situé entre 30 sec et 1 min selon les amorces utilisés.

Un autre facteur d'importance est le temps de polymérisation. Ce laps de temps a varié en fonction de la longueur de l'ADN matrice situé entre les deux amorces (minimum= 30 sec, maximum= 2 min). Plus le brin était long plus le temps de polymérisation devait l'être aussi. La température de polymérisation était de 72 C° pour tous les programmes.

La température de dénaturation de l'ADN était située entre 94 C° et 96 C° pour des

périodes de 30 sec. Tous nos programmes comportaient 30 cycles et se terminaient par une période d'élongation de 5 min à 72 C°.

### Programme

Cycles 1 à 30:

Température de dénaturation	94 C° à 96 C°	x 30 sec
Température d'hybridation	37 C° à 54 C°	x 30 sec à 1 min
Température de polymérisation	72 C°	x 30 sec à 1 min

Cycle d'élongation:

Température d'élongation	72 C°	x 5 min
--------------------------	-------	---------

### *B- Types d'amorces*

Deux types d'amorces ont été conçues, soit les amorces faites d'oligonucléotides dégénérés et celles faites avec des oligonucléotides non dégénérés. Le tableau III présente la séquence des amorces utilisées lors des amplifications géniques, leur type (dégénérées ou non dégénérées) et leur provenance.

Tableau III  
Amorces utilisées lors des amplifications géniques

Oligonucléotides dégénérés

Amorce	Séquence	Provenance
A1	5' GNGAYMRNGCNHTNYTNGARB T 3'	GSD Oligos
C	5' RTGNGTRTADATYTGNGT 3'	GSD Oligos

Oligonucléotides non dégénérés

Deb	5' ACAACTTCCGATACCCGCAA 3'	Biosource International
Fin	5' TACCACCCAAATCTACACGCATAT 3'	Biosource International
Fin2	5' CGGGTGGTTCAGGAGTTA C 3'	Biosource International
Fin3	5' TACCACCCAAATCTACACGC 3'	Dépt. de Biochimie, U de M
XdRINV	5' CTGTTGGAAAATGGCG 3'	Dépt. de Biochimie, U de M
XdFINV	5' CTGCTTGGGGACTGATT 3'	Dépt. de Biochimie, U de M
Forward	5' AGCGGATAACAATTTACACAGGA 3'	Appligene
Reverse	5' GCGCCAGGGTTTTCCCAGTCACGAC 3'	Appligene

Légende : B= C ou G ou T; D= A ou G ou T; H= A ou C ou T; M= A ou C; N= A ou C ou G ou T; R= A ou G; Y= C ou T; Dépt de Biochimie,U de M= Département de Biochimie de l'Université de Montréal.

Les deux oligonucléotides dégénérés A1 et C fabriqués par la compagnie General Synthesis and Diagnostics Oligos, ont été élaborés à partir des séquences connues des gènes de recombinaisons (*xerC* et *xerD* de *E. coli*; les gènes analogues à *xerC* et *xerD* de *Haemophilus influenzae*, le gène *sss* de *Pseudomonas* et le gène *codV* de *B.subtilis*. Nous n'avons pas tenu compte de la séquence de *L.leichmanii* puisqu'elle n'avait pas encore été découverte à ce moment là). Nous avons tenu compte, dans la conception des amorces, des variations en acides nucléiques rencontrées d'une espèce à une autre ou d'un type de recombinaison à un autre. Ces séquences se retrouvent dans des régions très conservées des gènes, très près ou dans la séquence codant pour son site actif (voir l'annexe II pour une

présentation détaillée des séquences connues des recombinaises et l'annexe I pour les séquences utilisées dans la fabrication des amorces A1 et C). L'amplification de la séquence contenue entre les amorces A1 et C des recombinaises connues devrait donner un fragment d'environ 450 pb. Ces deux amorces ont servi à la détection de gènes homologues inconnus dans les espèces de bactéries lactiques.

Les amorces non dégénérées *fin*, *fin2*, *fin3* ont toutes été utilisées séparément en conjonction avec l'amorce *deb*. Les amorces *XdRINV* et *XdFINV* ont été utilisées ensemble. Ces amorces ont été fabriquées à partir du brin de 450 pb de *Lactobacillus caseii* que nous avons séquencé (voir l'annexe II). Elles se retrouvent encore une fois dans ou près de la séquence codant pour le site actif de la protéine, mais leur orientation est vers l'extérieur de la séquence. Toutes ces amorces ont servis lors de PCR inverses sur différents ADN bactériens.

Enfin, les amorces Forward et Reverse ont été utilisées lors d'amplifications géniques de séquences clonées dans des vecteurs. Ainsi a été créée la sonde *cer* pour l'hybridation Southern avec le site *dif*.

### *C- Solutions pour l'amplification génique*

Toutes les solutions d'amplifications se retrouvent dans un volume final total de 50  $\mu\text{L}$ . Les enzymes, tampons et nucléotides utilisés sont de la compagnie Boehringer-Mannheim. Les éléments présents dans les solutions d'hybridations sont les suivants: 1  $\mu\text{L}$  de l'enzyme Taq polymérase 1U/ $\mu\text{L}$ , le tampon 10 X de la Taq avec  $\text{MgCl}_2$  (5  $\mu\text{L}$ ), les nucléotides dATP, dGTP, dCTP et dTTP à raison de 2mM par volume final, l'ADN à amplifier en volume variable selon l'expérience, les amorces également en volume variable,

et de l'eau déionisée stérile en quantité suffisante pour atteindre le volume final de 50  $\mu$ L. A cette solution a été rajouté 50  $\mu$ L d'huile minérale afin d'empêcher l'évaporation lors des cycles thermiques.

### **Les hybridations Southern**

Toutes les hybridations Southern effectuées dans le cadre de cette maîtrise ont utilisé les produits de la compagnie Boehringer-Mannheim pour hybridation avec de la Digoxigénine. L'avantage de ce type d'hybridation est qu'elle n'utilise pas de radioactivité et que la même solution d'hybridation peut être utilisée plusieurs fois. Après plusieurs essais, le protocole de base de cette compagnie et les protocoles pour hybridation Southern de Maniatis *et al* (1982) ont été adaptés pour mieux servir nos besoins expérimentaux.

La première étape de l'hybridation consistait en une digestion totale de l'ADN ciblé avec une ou des enzymes de restriction (NEB, 1 U par  $\mu$ g d'ADN digéré), suivie d'une migration électrophorétique en gel d'agarose (0.8%).

Après la prise de photo, l'ADN contenu dans le gel a été dénaturé pendant 1/2 heure avec agitation dans 80 mL d'une solution contenant 1.5 M NaCl et 0.5 M NaOH. Après avoir été rincé à l'eau distillée, le gel a été mis pendant 30 min. avec agitation dans 80 mL d'une solution neutralisante composée de 1 M Tris-HCl (pH 7.5), 1.5 M NaCl.

Nous avons ensuite procédé au transfert par capillarité de l'ADN du gel à une membrane de nylon Boehringer Mannheim chargé positivement. Le transfert a été effectué grâce à un montage dans lequel le gel était placé sur un papier Whatman 3MM dont les bouts baignaient dans environ 100 mL de 20 X SSC (Maniatis *et al*, 1982). La membrane de nylon a été mise par dessus le gel et recouverte à son tour d'un papier Whatman 3MM

de même dimension que celle-ci. Plusieurs couches de papiers absorbants d'une épaisseur d'environ 10 cm couvraient le tout. Le montage a ensuite été entouré de papier cellophane puis on a rajouté un poids d'environ 250g sur le dessus des papiers absorbants. Nous avons laissé le transfert se faire pendant toute une nuit à la température de la pièce.

Une fois le transfert terminé, la membrane a été trempée 5 minutes dans une solution de 6 X SSC puis laissé sécher à l'air. L'ADN a ensuite été fixé par UV dans un appareil de marque Stratagene UV Stratalinker.

#### *A- Pré-hybridation et hybridation*

La membrane a été pré-hybridée dans un bac dans 25 mL d'une solution contenant 5 X SSC, 2% de solution *Blocking Reagent* de Boehringer-Mannheim, 0.1% de N-Lauryl-Sarcosine, 0.02% SDS et 50% de formamide déionisée (Aldrich Chemical Company) à une température de 42 C° pendant 4 heures avec agitation dans un four d'hybridation Boekel. Nous avons ensuite hybridé la membrane à une sonde d'ADN pendant toute une nuit à 42 C° avec agitation. La solution d'hybridation contenant la sonde était de même composition que celle de pré-hybridation sauf pour l'ajout de la sonde.

#### *B-Sondes*

Plusieurs hybridations utilisant plusieurs sondes différentes ont été faites, soit les hybridations utilisant le site *cer* tronqué du plasmide pSHON et les hybridations utilisant les inserts séquencés de 400-450 pb de *Lactobacillus plantarum* et *Lactobacillus casei* contenus dans le vecteur pTZ-PC. Les sondes ont été préparées par amplification génique selon le protocole légèrement modifié de Boehringer Mannheim dont le principe de base est de fixer la digoxigénine, via la dig UTP, à plusieurs endroits dans le nouveau brin d'ADN.

Le protocole utilisé est le suivant : 1 µL d'une dilution 1:10 de l'ADN du vecteur

contenant la sonde est ajouté à une solution contenant un mélange de 36  $\mu\text{L}$  d'eau déionisée stérile, 5  $\mu\text{L}$  de tampon 10 X additionné de  $\text{MgCl}_2$  pour la polymérase Taq (Boehringer Mannheim), 1  $\mu\text{L}$  de l'amorce Forward et 1  $\mu\text{L}$  de l'amorce Reverse lorsqu'il s'agissait du vecteur pSHON; ou, dans les cas du vecteurs pTZ-PC portant les inserts de *L. plantarum* et *L. caseii*, 5  $\mu\text{L}$  de l'amorce A1 et 5  $\mu\text{L}$  de l'amorce C; 1 U de l'enzyme Taq polymerase de Boehringer Mannheim et finalement 5  $\mu\text{L}$  d'une solution de 20  $\mu\text{L}$  contenant 2  $\mu\text{L}$  de dATP 10mM, 2  $\mu\text{L}$  de dGTP 10mM, 2  $\mu\text{L}$  de dCTP 10mM, 1.27  $\mu\text{L}$  de dTTP 10mM, 5.73  $\mu\text{L}$  d'eau déionisée stérile et finalement 7  $\mu\text{L}$  de digUTP (concentration ) (tous les dNTP proviennent de la compagnie Boehringer Mannheim). On obtient ainsi une sonde identique à la matrice originale d'ADN sauf à certains endroits où une thymine aura été remplacée par une digUTP.

Préalablement à l'hybridation, la sonde a été dénaturée pendant 10 minutes dans un bain à 100 C° puis rapidement refroidie dans un bain d'eau glacée avant d'être ajoutée à la solution d'hybridation.

#### *C- Lavage et détection*

La base du protocole pour le lavage et la détection des membranes est celui fourni par la compagnie Boehringer Mannheim que nous avons modifié selon nos besoins. Les recettes des tampons utilisés pour la détection (Tampons n°1, 2 et 3) sont exactement ceux fournis par cette compagnie.

Toutes les étapes sauf la dernière sont effectuées avec agitation. Après l'hybridation la membrane est lavée dans 50 mL de 2XSSC, 0.1% SDS pendant 5 min, deux fois. On lave ensuite dans 50 mL d'une solution de 0.1 SSC, 0.1% SDS pendant 15 min, deux fois. La membrane est ensuite trempée dans 50 mL de tampon n°1 pendant 1 min puis dans 50 mL

de tampon n°2 pendant 30 min. La prochaine étape est celle de l'accrochage des anticorps anti-Digoxigénine aux molécules de Digoxigénine présentes sur la membrane. Ainsi on dilue 6 µL de la solution Anti-Digoxigenin-Ap Fab Fragments de la compagnie Boehringer-Mannheim dans 30 mL de tampon n°2 puis on y trempe la membrane pendant 30 minutes. La membrane est ensuite mise dans 50 mL de tampon n°1 pendant 15 min, deux fois. L'étape suivante consiste à tremper la membrane dans 50 mL de tampon n°3 pendant 2 minutes. La dernière étape est celle où se fera la détection proprement dite: on prépare une solution de 30 mL de tampon n°3 auquel on ajoute 45 µl de NBT à 75 mg/mL (sel de Nitroblue Tetrazolium) et 35µl de X-Phos à 50mg/mL (5-Bromo-4-Chloro-3-Indolyl phosphate). La membrane est déposée dans la solution, le bac est recouvert d'un papier d'aluminium afin de le rendre complètement opaque et on laisse reposer sans aucune agitation pendant toute une nuit. Le lendemain la membrane est trempée dans 50 mL de tampon n°1 pendant 5 min puis elle est séchée à l'air et est donc prête pour l'analyse finale des bandes.

### **Le séquençage**

Les séquençages de fragments de gènes des recombinaisons de nos différentes souches ont été faits manuellement dans notre laboratoire et également par le service de séquençage de l'Université du Maine avec un appareil ABI371. Le protocole et le matériel utilisé pour le séquençage manuel de nos échantillons était celui du T7 Sequencing Kit de Pharmacia. L'isotope utilisé pour marquer les nucléotides était le <sup>35</sup>S de la compagnie Dupont.

## Les analyses statistiques

Tous les tests statistiques ont été réalisés grâce au logiciel Genetics Computer Group du Wisconsin Package . Ce logiciel permet d'appliquer un grand nombre d'opérations sur une série de séquences d'acides aminés ou de nucléotides. Nous avons dû éditer les séquences pour que celles-ci aient des longueurs comparables. Ainsi, les séquences de *xerD* et *xerC* de *E.coli*, *xerC* et *xerD* de *H. influenzae*, *sss* de *P. aeruginosa*, *codV* de *B.subtilis* et celle de *L.leichmanii* ont été tronquées pour que leur début et leur fin coïncident avec la plus petite séquence que nous avons, soit celle de *L.caseii*. Les séquences de *L.plantarum* 8014 et *L.bulgaricus* 737 n'ont été tronquées qu'au début de leurs séquences puisque leurs fins coïncidaient déjà avec celles de *L.caseii*. Nous n'avons pas inclus la séquence de *L.plantarum* 14917 dans nos analyses car après édition elle était identique à celle de *L.plantarum* 8014 et ceci aurait pu introduire des biais dans nos analyses. Avant les analyses, nous avons préalablement traduit les séquences d'ADN en séquences d'acides aminés avec le programme "Map" et nous les avons éditées afin d'obtenir le bon cadre de lecture.

Nous avons utilisé le programme "Pileup" pour aligner les séquences de nucléotides et d'acides aminés des 10 espèces de recombinases (séquences de : *xerD* de *E. coli*, *xerC* de *E.coli*, *L. leichmanii*, *L. caseii*, *L. bulgaricus* 737, *L. plantarum* 8014, *xerD* d'*H. influenzae*, *xerC* d'*H. influenzae*, *codV* de *B. subtilis* et *sss* de *P. aeruginosa*). Ce programme permet de créer un alignement multiple des séquences à l'intérieur d'un groupe d'après leur ressemblance en utilisant des alignements pairés progressifs. Nous avons ensuite soumis ces alignement de séquences au programme "Pretty". Ce programme permet de comparer les différences et les ressemblances entre les séquences (par la mise en évidence

des positions consensus ou par la mise en évidence de positions différentes à l'intérieur des séquences).

Nous avons utilisé le programme "Distances" et "Growtree" pour générer des arbres phylogénétiques. Le programme "Distances" permet de créer une matrice des distances pairées à l'intérieur d'un groupe de séquences. Ces distances sont exprimées comme des substitutions par chaque 100 pb ou acides aminés. Nous avons choisi de calculer la matrice des distances entre les séquences de peptides par la méthode de Jukes-Cantor puisque l'autre méthode proposée (méthode de Kimura) générait des taux de substitutions exagérément élevés. La même méthode a été choisie pour calculer les distances entre les séquences de nucléotides puisque les deux autres méthodes (Jin-Nei et Tajima-Nei) ont généré trop de données aberrantes (Program manual for the Wisconsin Package, 1994).

Le programme "Growtree" permet, à partir de la matrice générée par le programme "Distances", de créer un arbre phylogénétique des distances entre les recombineses. Nous avons choisi l'option de calcul UPGMA qui procède par regroupement progressif des paires ou groupes de séquences contenant les plus petites distances. Ainsi, les positions phylogénétiques les plus rapprochées sont d'abord calculées à partir de la plus petite distance entre deux séquences dans la matrice. Ces deux séquences sont ensuite regroupées et ce nouveau groupe remplace les deux séquences originales pour le prochain calcul. La prochaine position phylogénétique est calculée avec la prochaine séquence qui a la plus petite distance matricielle avec ce nouveau groupe, puis elle est regroupée avec celui-ci et ainsi de suite jusqu'à ce qu'il ne reste qu'un seul groupe. A chaque tour de calcul la matrice perd donc une dimension. Ce test suppose que le taux de substitution génique est constant et que la mesure de distance est linéaire avec le temps.

## CHAPITRE 3

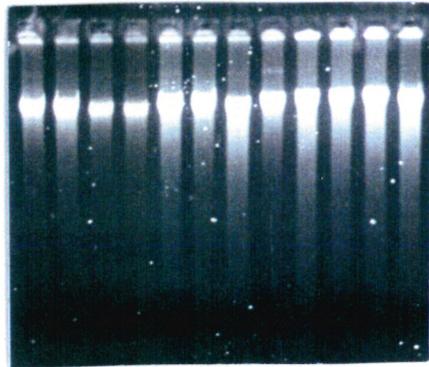
### RESULTATS

#### **Le site *dif* chez *Lactobacillus caseii***

La première étape de notre recherche était de trouver une séquence analogue au site *dif* de *E.coli* chez *Lactobacillus caseii* par hybridation Southern, avant tout il fallait extraire de l'ADN de *L. caseii*. Après plusieurs essais infructueux avec des protocoles d'extraction d'ADN de bactéries gram positives, nous avons mis au point un nouveau protocole (décrit au chapitre Matériel et Méthodes) très efficace. Les résultats de ces extractions sont présentés à la figure 7 (chaque puits contient 5  $\mu$ L).

Figure 7

#### Extraction d'ADN de *Lactobacillus caseii*



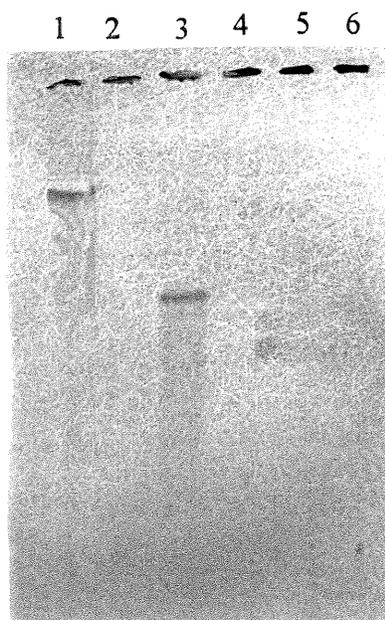
Nous avons fait une série d'hybridations Southern en utilisant l'ADN digéré de *L.caseii* comme cible et une séquence analogue à *dif* comme sonde, soit la séquence tronquée de *cer* de 80 pb du plasmide pSHON (voir figure 6 au chapitre Matériel et

méthodes). La sonde à été obtenue par amplification génique en utilisant des nucléotides marqués à la digoxigénine.

La figure 8 présente les résultats d'une hybridation Southern entre cette sonde et de l'ADN digéré avec l'enzyme *Eco* RI. Le puits 3 contient de l'ADN de *Lactobacillus caseii*. Le premier puits contient le témoin *E. coli*.

Figure 8

Hybridation Southern entre ADN de *L. caseii* digéré avec *Eco* RI, et la sonde *dif*



Puits 1 = ADN de *E. coli*  
Puits 3 = ADN de *L. caseii*

La figure 9 présente les résultats d'une hybridation Southern entre cette même sonde et de l'ADN de *L. caseii* digéré avec *Aat* II (puits 1), *Bam* HI (puits 2), *Dra* I (puits 3), *Eco* RI (puits 4), *Hinc* II (puits 5), *Hind* III (puits 6) et *Pst* I (puits 7). Le dernier puits contient le marqueur  $\lambda$  digéré avec *Bste* II.

Figure 9

Hybridation Southern entre ADN de *L. caseii* digéré avec plusieurs enzymes, et sonde *dif*



ADN de *L. caseii* digéré avec :  
 Puits 1 = *Aat* II  
 Puits 2 = *Bam* HI  
 Puits 3 = *Dra* I  
 Puits 4 = *Eco* RI  
 Puits 5 = *Hinc* II  
 Puits 6 = *Hind* III  
 Puits 7 = *Pst* I  
 Puits 8 = marqueur de poids  $\lambda$  *Bste* I

Les résultats de ces deux hybridations montrent que la sonde s'est hybridée à l'ADN de *L. caseii*. Certaines impuretés sur la membrane présentée à la figure 9 masquent les résultats obtenus pour le puits contenant l'ADN digéré avec l'enzyme *Dra*I (puits3). On note aussi la présence de trois bandes d'hybridation dans le puits contenant l'ADN digéré avec l'enzyme *Aat* II, reflétant probablement une digestion partielle. L'ensemble des résultats indiquent que le génome de *Lactobacillus caseii* comporte une séquence analogue à la séquence *dif* de *E. coli*. Nous avons également fait des essais d'hybridation de cette même sonde avec de l'ADN de *Lactobacillus bulgaricus* 737. Les résultats de cette expérience montrent de très faibles signaux d'hybridation suggérant que cette espèce a elle aussi une séquence analogue à *dif*. Néanmoins, avant de pouvoir se prononcer à cet effet, il faudrait

continuer les expériences afin d'ajuster adéquatement les paramètres de l'hybridation pour cette espèce particulière.

### **Les recombinaisons chez les bactéries lactiques**

La deuxième étape de notre recherche était de trouver l'équivalent des recombinaisons XerC et/ou XerD chez plusieurs espèces de bactéries lactiques. Nous avons d'abord essayé de déceler la présence des gènes de recombinaisons chez plusieurs espèces de bactéries lactiques par PCR avec amorces dégénérées puis nous avons fait des PCR dits "inverses" dans le but de faire ressortir une séquence complète de recombinaison.

#### *A- Les amplifications géniques avec amorces dégénérées*

Dans un premier temps nous avons donc effectué une série d'amplifications géniques en utilisant les amorces A1 et C sur les matrices d'ADN bactériens de *L. casei*, *L. bulgaricus* souche 737, *L. plantarum* souches 8014 et 14917, *L. fermentii* souches BP83 et 1.16A, *Lactococcus lactis* spp *cremoris* souches H2, MG1363 et AM2, *Lactococcus lactis* spp *lactis* souches 763 , 1667 et IL 1403 et finalement *Streptococcus salivarius* souche ATCC 13419.

Les amorces A1 et C utilisées sont des oligonucléotides dégénérés faits à partir des séquences peptidiques XerC et XerD de *E. coli*, des séquences analogues chez *P. aeruginosa* et *H. influenzae* et de la séquence codV de *B.subtilis*. La recombinaison de *L. leichmanii* n'a été prise en considération puisqu'elle n'a été séquencée qu'en 1996. Les amorces tiennent compte des variations observées entre les gènes de ces différentes espèces ce qui permet l'hybridation à des régions pouvant ne comporter qu'une

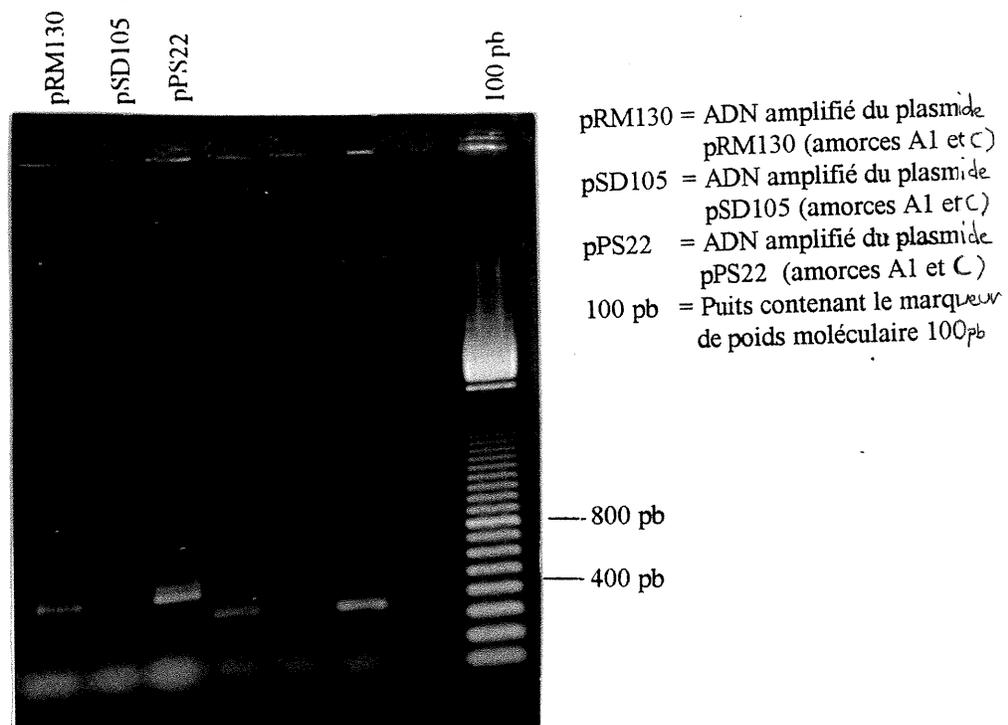
relativement faible homologie à l'une ou l'autre de ces espèces prises isolément. La région qu'elles devraient permettre d'amplifier est une séquence d'environ 450 pb (sur un total d'environ 1000 pb pour le gène complet).

Nous avons d'abord testé les amorces A1 et C avec un PCR sur trois différentes matrices, soit le plasmide pSD105 portant la séquence *xerC* de *E.coli*, le plasmide pRM130 portant la séquence *xerD* de *E.coli* et le plasmide pPS22 portant la séquence *codV* de *B.subtilis*. La température d'hybridation optimale de 37 C° est celle que nous avons utilisée pour le PCR.

Avec le témoin pPS22, nous nous attendions à une bande amplifiée d'environ 450 pb correspondant à la région comprise entre les séquences des amorces A1 et C. Pour les deux autres témoins nous nous attendions à retrouver des bandes amplifiées de poids moléculaires semblables. La figure 10 présente les résultats obtenus pour ce PCR témoin.

Figure 10

Amplification des témoins pSD105, pRM130 et pPS22 avec les amorces A1 et C



Comme on peut le constater, on obtient une bande amplifiée assez forte d'environ 450 pb pour le témoin pPS22. Le témoin pRM130 donne également une bande amplifiée entre 400 et 450 pb. Les signaux plus faibles sont probablement le fait d'hybridations non spécifiques. Ces résultats indiquent donc que les amorces A1 et C sont adéquates pour la détection de la séquence *codV* de *B.subtilis* et *xerD* de *E.coli*.

L'amplification génique du témoin pSD105 donne deux faibles bandes d'hybridations, de poids moléculaire moindre que celui auquel on s'attendait. Nous ne pouvons expliquer ce dernier résultat autrement que par une plus faible homologie entre le gène *xerC* de *E.coli* et les amorces A1 et C.

Puisqu'au moins les deux premiers témoins semblent indiquer que les amorces s'hybrident correctement, nous les avons utilisées dans une série d'amplifications géniques visant à détecter la présence de séquences homologues aux recombinaisons recherchées. Les éléments variables étaient la température d'hybridation (de 37 C° à 45 C°) et l'ADN de l'espèce examinée. Nous considérons un résultat de PCR comme non spécifique chaque fois que nous observions un nombre élevé de bandes faiblement amplifiées, ou des bandes de poids moléculaire plus faible ou plus élevé que le 400-450 pb attendu. Si le résultat d'un PCR était non spécifique ou s'il donnait des résultats mitigés (faible bande à 450 pb en plus d'une série de bandes non spécifiques) nous le reprenions avec une température d'hybridation plus élevée. Si les PCR ne produisaient plus de bandes, ou si les résultats étaient systématiquement des bandes non spécifiques, nous arrêtons les expériences pour cette espèce. Si le PCR produisait une bande amplifiée forte dont le poids était situé entre

400-450 pb, cette bande était clonée et séquencée (voir les séquences en annexe). Le tableau IV présente un résumé des résultats des PCR réalisés.

Tableau IV

Résultats des PCR avec les amorces A1 et C sur les ADN des bactéries lactiques  
Température d'hybridation

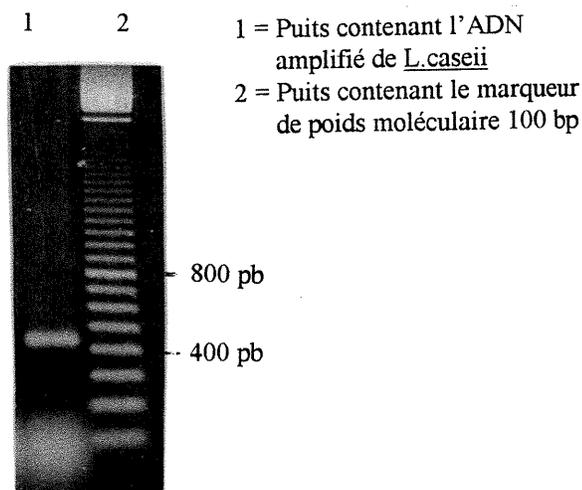
Espèce	37 C°	38 C°	39 C°	40 C°	41 C°	42 C°	43 C°	45 C°
<i>L.caseii</i>	Bande forte 450 pb		Bande forte 450 pb		Bande forte 450 pb		Bande forte 450 pb	Bande forte 450 pb
<i>L.plantarum</i> 8014	Bande forte 450 pb		Bande forte 450 pb					
<i>L.plantarum</i> 14917	Bande forte 450 pb		Bande forte 450 pb					
<i>L.bulgaricus</i> 737	Bandes non spécifiques		Bande forte 450 pb					
<i>L.fermentii</i> BP83	Bandes non spécifiques	Bandes non spécifiques	Bandes 500 pb et non spécifiques	Bandes 650 pb et non spécif		Bandes 650 pb et non spécif		
<i>L.fermentii</i> 1.16A	Bandes non spécifiques		Bandes non spécifiques	Bandes non spécif.	Bandes non spécif.	Bandes non spécif.		Bande 650 pb
<i>L.lactis cremoris</i> MG1363	Bandes non spécifiques		Bandes non spécifiques	Bandes non spécif.	Aucune bande	Aucune bande		Aucune bande
<i>L.lactis cremoris</i> AM2	Bandes non spécifiques	Bande 500 pb et non spécif.	Bande légère 500 pb			Aucune bande		
<i>L.lactis cremoris</i> H22	Aucune bande	Aucune bande	Aucune bande	Aucune bande				
<i>L.lactis lactis</i> H1403	Bandes non spécifiques		Bandes non spécifiques	Bandes non spécif.	Aucune bande			
<i>L.lactis lactis</i> 763	Bandes non spécifiques	Bandes 350 pb et non spécif.	Bandes 500 pb et non spécif	Aucune bande	Aucune bande			
<i>L.lactis lactis</i> 1667	Bandes non spécifiques	Bandes non spécifiques	Aucune bande					
<i>S.salivarius</i>	Bandes non spécifiques	Aucune bande	Aucune bande					

Les résultats du tableau IV montrent que relativement peu de PCR ont donné des résultats positifs. En effet, seules 4 souches sur les 13 testées ont produit une bande claire de 400-450 pb. Ces souches sont celles de *L. caseii*, *L. plantarum* 8014, *L. plantarum* 14917 et *L. bulgaricus* 737. L'absence de résultats positifs au PCR ne veut cependant pas dire que les recombinases sont absentes chez ces espèces puisqu'un écart d'homologie relativement petit peut se traduire par une absence d'hybridation ou une hybridation non spécifique même lorsque les conditions de l'amplification génique sont non astringentes (faible température d'hybridation et amorces dégénérées).

La figure 11 présente la bande de 450 pb obtenue pour l'ADN de *L. caseii*. Les bandes amplifiées positives ont été clonées dans le vecteur pT7-blue (*L. caseii*) et pTZ-PC (*L. plantarum* 8014 et 14917 et *L. bulgaricus* 737) puis elles ont été séquencées. Les produits de ces séquençages ainsi que les séquences déjà connues de recombinases sont présentés à l'annexe II.

Figure 11

Bande amplifiée de 450 pb de *L. caseii* (PCR avec A1 et C)



Une analyse sommaire des séquences obtenues montrent d'assez fortes homologies avec l'ADN codant pour la région du bout C-terminal des recombinaisons connues. Les souches *L.caseii*, *L.plantarum* 8014, *L.plantarum* 14917 et *L. bulgaricus* 737 posséderaient donc des recombinaisons comparables à XerC et/ou XerD.

Des analyses plus poussées et des comparaisons de ces séquences entre elles et avec des séquences connues de recombinaisons sont présentées dans la section portant sur les résultats de l'analyse statistique.

#### *B- Les amplifications géniques "inverses"*

Le but des PCR dits inverses était d'amplifier les régions de chaque côté du 450 pb qui avait déjà été séquencé afin d'éventuellement obtenir la séquence complète du gène. Nous avons utilisé les données de l'ADN séquencé de *L.caseii* dans l'élaboration des amorces. Cette espèce a été choisie pour la seule raison qu'elle a été la première à être séquencée était la seule disponible à ce moment-là. Les amorces ont été conçues de façon à s'hybrider à des régions limites de chaque côté de la séquence et orientées vers l'extérieur de celle-ci.

Malgré tous nos efforts, nous n'avons pas obtenu de résultat avec ce type de PCR et ce en dépit du fait que nous avons employé 6 paires d'amorces (deb-fin, deb-fin2, deb-fin3, XDRinv-XDFinv) dans 28 PCR différents. Cette absence de résultats s'explique difficilement puisque la séquence des amorces utilisées a été prise à même la région qui a été séquencée et devrait donc correspondre base pour base à celle-ci. Pour les amorces XDRinv et XDFinv nous avons même fait un contrôle en utilisant une paire d'amorces dont la séquence était identique à ces dernières mais dans l'orientation inverse. Ainsi, nous nous

attendions à amplifier une séquence témoin de 230 pb comprise entre les deux amorces. C'est en effet ce que nous avons obtenu de façon claire et non-équivoque. Les amorces complémentaires XDRinv et XDFinv étaient donc très spécifiques. De plus, certains de ces PCR ont été faits à des températures très basses afin de faciliter l'hybridation entre ADN et amorces. Nous ne pouvons pas non plus incriminer toutes les étapes préalables aux PCR (digestions et autoligations des fragments) puisque quelques PCR inverses ont quand même livré des bandes d'amplification. Certaines de ces bandes, dont le poids moléculaire était près de celui auquel on s'attendait pour le gène, ont été séquencées mais se sont finalement révélées ne pas correspondre au gène recherché.

Il est possible que chacun de ces échecs puisse s'expliquer individuellement soit par une mauvaise digestion, soit par une mauvaise autoligation, soit par une absence d'hybridation ou encore par une hybridation non spécifique entre la sonde et l'ADN. Il reste que cette série d'échecs malgré le fait que nous avons tenu compte de plusieurs variables qui auraient pu affecter les résultats, porte à penser que nous n'avons cependant pas réussi à contrôler tous les facteurs. Ainsi, en augmentant le nombre de témoins autant sur la façon de procéder que sur le matériel, nous aurions peut-être pu détecter le ou les problèmes.

### **Analyses statistiques des séquences des recombinaisons**

Les analyses statistiques comprennent deux volets. Le premier volet traite des comparaisons entre les différentes séquences tant au niveau des acides aminés qu'à celui des nucléotides. Le deuxième volet traite de phylogénies suggérées pour ces séquences. Nous avons édité les séquences de nucléotides pour qu'elles aient toutes des longueurs

comparables (maximum de 432 pb ou 144 acides aminés et minimum de 412 pb ou 138 a.a.). De plus, afin d'éviter d'introduire des biais, nous avons omis des analyses la séquence de *L. plantarum* 14917 puisqu'après édition elle était identique (à l'exception d'un nucléotide) à la séquence de *L. plantarum* 8014. Les analyses ont donc porté sur 10 séquences soit celles de : *L. caseii*, *L. plantarum* 8014, *L. bulgaricus* 737, XerC de *E.coli*, XerD de *E.coli*, Sss de *P.aeruginosa*, CodV de *B.subtilis*, XerC de *H.influenzae*, XerD de *H.influenzae* et XerC de *L.leichmannii*.

#### *A- Comparaison des séquences de recombinaisons*

Après édition des séquences, nous les avons alignées avec le programme "Pileup". Nous avons ensuite utilisé le programme "Pretty" qui permet de comparer les séquences entre elles. Nous avons ainsi obtenu une séquence de consensus absolu qui indique les nucléotides ou acides aminés complètement conservés pour les 10 séquences (voir l'annexe III).

Les résultats de cette analyse montrent que les séquences peptidiques sont relativement assez bien conservées. Ainsi le consensus absolu pour toutes les séquences est de 21.52 %. Le consensus est cependant beaucoup moins élevé au milieu qu'aux extrémités de la séquence, la fin présentant le consensus le plus élevé. Ainsi on retrouve neuf acides aminés conservés entre les positions 1 à 50, 1 entre 51 et 100 et 21 entre 101 et 144. Les acides aminés terminaux TQIYTH, contenant le site d'attaque initial de l'ADN de la protéine (tyrosine), sont conservés à 100%.

Le consensus au niveau des nucléotides est moins élevé (19.63%). Nous retrouvons

le même patron que pour les acides aminés, c'est-à-dire que le début et surtout la fin des séquences sont les mieux conservés. Ainsi, 30 nucléotides sont complètement conservés entre les positions 1 à 150, 9 entre les positions 151 à 300 et 46 entre les positions 301 à 433. Nous n'observons cependant pas un consensus absolu dans la région codant pour les acides aminés TQIYTH.

Par ailleurs, ce même programme a été utilisé pour comparer de façon dyadique les séquences de nucléotides et d'acides aminés de *L. plantarum* 8014, *L. caseii* et *L. bulgaricus* 737 aux séquences de XerC et XerD de *E. coli* afin de déterminer à quoi ces premières séquences ressemblent le plus. Pour chacune de ces trois séquences nous avons donc calculé le consensus existant entre elles et XerC, et entre elles et XerD. Le tableau V résume les pourcentages de consensus pour chacune des paires, dans le cas des acides aminés et dans celui des nucléotides.

Tableau V

Pourcentage de consensus entre XerC et XerD et les espèces séquencées

ESPECE	XerC		XerD	
	Acides aminés	Nucléotides	Acides aminés	Nucléotides
<i>L. caseii</i>	43.97 %	50.82 %	51.4 %	57.51 %
<i>L. bulgaricus</i> 737	43.66 %	48.59 %	53.14 %	55.24 %
<i>L. plantarum</i> 8014	46.47 %	48.36 %	49.65%	55.32 %

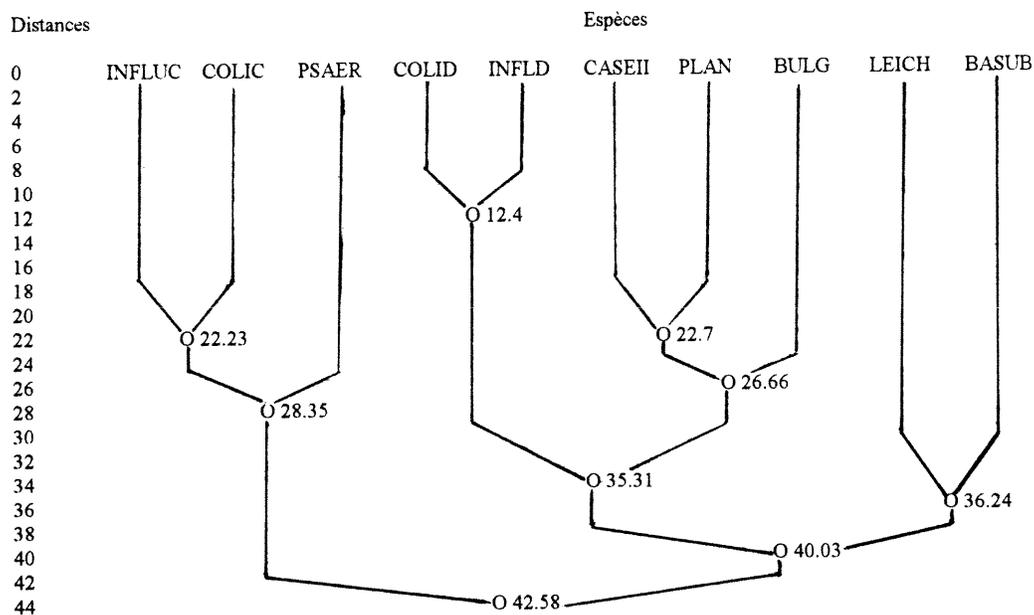
Les résultats indiquent que toutes les séquences tendent à ressembler plus à XerD qu'à XerC de *E. coli* (il est à noter que les consensus entre la paire témoin XerC et XerD de *E. coli* sont de 45.77 % pour les acides aminés et de 54.69 % pour les nucléotides).

### B- Phylogénie suggérée

Nous avons utilisé le programme "Growtree" pour exécuter deux arbres phylogénétiques, soit un arbre pour les séquences peptidiques et un autre pour les séquences d'ADN. La figure 12 présente l'arbre phylogénétique des dix espèces au niveau de la séquence de peptides et la figure 13 présente l'arbre phylogénétique de ces dix mêmes espèces pour les séquences d'ADN. Les cercles indiquent les points de divergence entre les espèces, et les chiffres représentent l'étendue de la distance entre deux espèces ou entre des groupes d'espèces, telle qu'exprimée par le logiciel "Growtree".

Figure 12

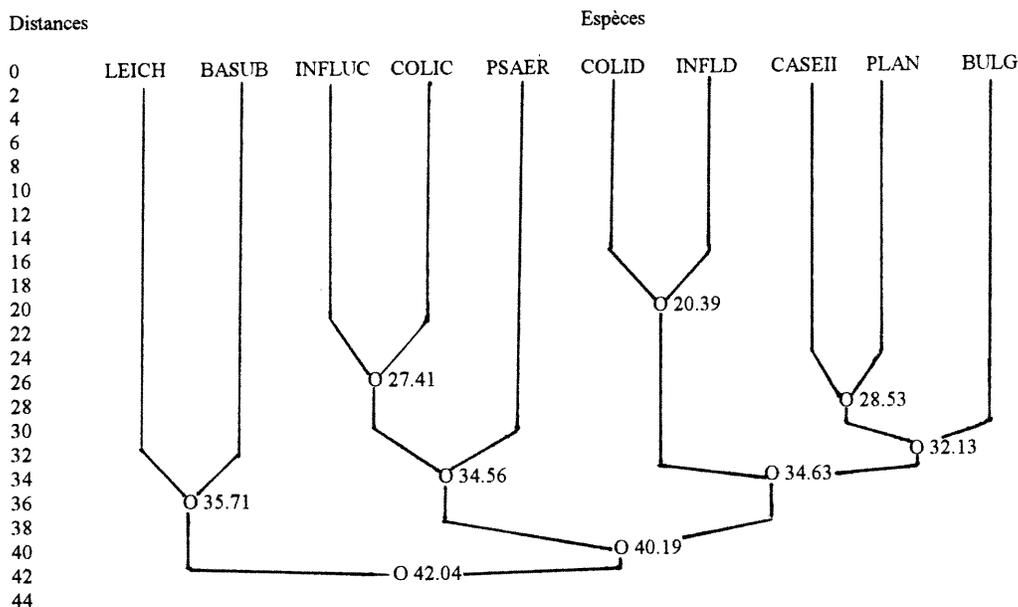
#### Arbre phylogénétique des séquences peptidiques



**Légende :**

INFLUC = XerC d'*H.influenzae*, COLIC = XerC de *E.coli*, PSAER = Sss de *P. aeruginosa*, COLID = XerD de *E.coli*, INFLD = XerD d'*H.influenzae*, CASEII = séquence de *L.caseii*, PLAN = séquence de *L.plantarum* 8014, BULG = séquence de *L.bulgaricus* 737, LEICH = XerC de *L.leichmannii*, BASUB = codV de *B.subtilis*

Figure 13

Arbre phylogénétique des séquences d'ADN

## Légende :

INFLUC = XerC d'*H.influenzae*, COLIC = XerC de *E.coli*, PSAER = SSS de *P. aeruginosa*, COLID = XerD de *E.coli*, INFLD = XerD d'*H.influenzae*, CASEII = séquence de *L.caseii*, PLAN = séquence de *L.plantarum* 8014, BULG = séquence de *L.bulgaricus* 737, LEICH = XerC de *L.leichmannii*, BASUB = CodV de *B.subtilis*

L'arbre phylogénétique des peptides montrent que les séquences les moins divergentes sont celles de XerD de *E.coli* et XerD d'*H.influenzae*, suivi de la paire formée par XerC de *E.coli* et XerC d'*H.influenzae* puis par celle formée de la séquence de *L.caseii* et de *L.plantarum*. La prochaine distance la plus petite se retrouve entre la paire (*L.caseii*-*L.plantarum*) et la séquence de *L.bulgaricus*. Celle-ci est suivie de très près par la distance entre la paire formée par (XerC de *E.coli*-XerC d'*H.influenzae*) et la séquence Sss de *P.aeruginosa*. La prochaine distance la plus faible est celle entre le groupe (*L.caseii*-*L.plantarum*-*L.bulgaricus*) et le groupe (XerD-*E.coli*-XerD d'*H.influenzae*). On retrouve ensuite la distance entre XerC de *L.leichmannii* et CodV de *B.subtilis*. Le prochain point de

divergence est celui entre le groupe (XerD de *E.coli*-XerD d'*H.influenzae*-*L.caseii*-*L.plantarum*-*L.bulgaricus*) et la paire (*L.leichmannii*-*B.subtilis*). Le dernier point de divergence est le point situé entre le groupe de (XerC de *E.coli*-XerC d'*H.influenzae*-*P.aeruginosa*) et le reste des espèces.

L'arbre phylogénétique des séquences d'ADN diffère en deux points de celui des peptides. La première différence, de moindre importance, est la distance entre la séquence de *L.caseii* et *L.plantarum*, qui passe du deuxième au troisième rang. La deuxième différence, celle-là plus importante, est celle de la position du groupe XerC de *L.leichmannii*-CodV de *B.subtilis*: ce groupe est maintenant celui qui se distingue le plus de toutes les autres espèces alors qu'au niveau des peptides on le retrouvait plus près du groupe (XerD-*E.coli*-XerD d'*H. influenzae*-*L.caseii*-*L.plantarum*-*L.bulgaricus*) que du groupe (XerC-*E.coli*-XerC d' *H. influenzae*-*P.aeruginosa*).

## CHAPITRE 4

### DISCUSSION

Les résultats des hybridations Southern montrent que le génome de *Lactobacillus caseii* comporte une séquence analogue à celle de *dif*. Puisqu'il est logique de penser que tout génome circulaire doit posséder un mécanisme de recombinaison permettant la résolution des dimères chromosomiques lors de la réplication, nous supposons que la séquence que nous avons détectée est celle qui permet la recombinaison entre les chromatides soeurs chez *L.caseii*. Les résultats préliminaires des hybridations Southern sur l'ADN de *Lactobacillus bulgaricus* 737 tendent à montrer que cette espèce comporterait également une séquence analogue à *dif*. Ces résultats restent cependant à être confirmés par d'autres expériences.

Les PCR utilisant les amorces dégénérées A1 et C et les séquençages des bandes obtenues montrent que les génomes de *L.caseii*, *L.plantarum* 14917, *L.plantarum* 8014 et *L.bulgaricus* 737 possèdent tous des séquences d'environ 450 pb analogues à l'extrémité de 450 pb des gènes *xerC* et /ou *xerD* codant pour la partie C-terminale des enzymes. Ces quatre espèces posséderaient donc les gènes codant pour des recombinases analogues à XerC et/ou XerD. Nous n'avons pas obtenu de résultats pour les autres espèces examinées (*Lactobacillus fermentii* BP83 et 1.16A, *Lactococcus lactis* spp *cremoris* MG1363, AM2 et H22, *Lactococcus lactis* spp *lactis* IL1403, 763 et 1667 et *Streptococcus salivarius* 13419). Cette absence de résultats n'implique pas automatiquement que ces espèces n'ont pas de recombinases mais pourrait tout simplement traduire un manque d'homologie entre

les amorces et les séquences géniques. Ainsi il est important de noter que la recombinaise de l'espèce *L. Leichmanii* séquencée par Becker et Brendel (1996) ne montre pas d'homologie avec l'amorce A1. Quant à l'absence de résultats pour les PCR inverses plusieurs explications sont possibles. Ainsi il pourrait s'agir soit de mauvaises autoligations, soit de digestions enzymatiques incomplètes, soit de la formation de boucles sur l'ADN simple brin produit lors des cycles de dénaturation de l'amplification empêchant par la suite la polymérisation du brin complémentaire (boucles G-C).

Bien que nous n'ayons pas de séquences complètes de gènes nous pensons être justifiée d'en analyser la composition puisqu'il s'agit de la moitié la plus importante et la plus hautement conservée des recombinaises, soit le bout C-terminal contenant le site actif de la protéine. Les résultats de ces analyses montrent que les similarités entre les séquences de *L. caseii*, *L. plantarum* (8014 et 14917) et *L. bulgaricus* 737 et les protéines déjà connues se situent surtout à chaque extrémité des séquences et en particulier à la fin. C'est à cette dernière position que l'on retrouve le site actif des recombinaises (soit les domaines I et II), contenant entre autres les acides aminés TQIYTH du domaine II impliqués dans le clivage initial de l'ADN. Les acides aminés TQIYTH sont présents chez ces quatre espèces et sont conservés à 100%. Cette constatation ne se maintient pas au niveau des codons mais ceci s'explique facilement en termes de dégénérescence du code génétique. Le site actif de la recombinaison, très conservé chez les espèces gram négatives, l'est donc également chez ces espèces gram positives, témoignant de l'importance évolutive de cette région puisque ce caractère serait apparu avant la divergence des deux groupes et aurait été maintenu jusqu'à présent.

Nous avons ensuite examiné le degré de ressemblance entre les séquences de *L.caseii*, *L.plantarum* (8014 et 14917) et *L.bulgaricus* 737 et les recombinaisons XerC et XerD de *E.coli*. Nous avons trouvé qu'elles ressemblaient plus à XerD qu'à XerC. Nous ne pouvons cependant conclure de façon non équivoque qu'il s'agit de séquences représentant l'équivalent de XerD, puisque nous ne possédons pas de séquences complètes et que nous n'avons pas soumis ces séquences à des tests statistiques. Néanmoins, puisqu'il s'agit des séquences comprenant la partie la plus importante et la plus hautement conservée des protéines, ces résultats peuvent constituer un bon indice quant à leur identité possible.

Les principes sous-jacents à la construction d'arbres phylogénétiques sont que le taux de substitution génique est constant et que la mesure de distance est linéaire avec le temps évolutif. Nous avons ici supposé que ces deux principes étaient vrais. Deux arbres ont été construits, un pour les séquences nucléiques et un autre pour les séquences peptidiques. Les phylogénies obtenues ne peuvent être interprétées qu'au niveau des recombinaisons et non au niveau des espèces, c'est-à-dire que le taux de divergence entre deux séquences ne reflète pas nécessairement celui qui existe réellement entre les deux espèces.

L'arbre obtenu pour les séquences peptidiques est intéressant sous plusieurs aspects. Tout d'abord nous remarquons que l'arbre se divise en deux, non pas selon le gram mais bien selon les recombinaisons (à l'exception de *B.subtilis*). Ainsi *L.caseii*, *L.plantarum* et *L.bulgaricus* se retrouvent plus près des séquences XerD de *E. coli* et d' *H. influenzae* que de celles de *L.leichmannii* et *B.subtilis*. Ceci s'explique facilement dans le cas de la séquence de *L. leichmannii* puisque cette dernière représente une séquence analogue à XerC alors que nos séquences auraient plutôt tendance à ressembler à XerD. Cependant, la

séquence CodV de *B. subtilis*, présentant une plus forte analogie à XerD qu'à XerC, devrait logiquement être plus rapprochée du reste des gram positives que ces dernières ne le sont des séquences XerD des gram négatives. Cette différence pourrait signifier que l'origine phylétique des séquences de *L. caseii*, *L. bulgaricus* et *L. plantarum* serait plus près de celle qui a donné les séquences XerD de *E. coli* et *H. influenzae* que de celle qui a donné la séquence CodV de *B. subtilis*. Il est également intéressant de noter que la séquence de *L. leichmannii* est la plus proche parente de la séquence de *B. subtilis* et ce, en dépit du fait qu'elles représentent deux recombinaisons différentes. De plus, nous observons que la séquence de *L. leichmannii*, bien qu'elle s'apparente plus à XerC qu'à XerD, ne se retrouve pas à proximité des séquences analogues à XerC des espèces gram négatives mais plutôt à proximité des séquences analogues à XerD contenant les espèces gram positives. La différence entre XerC des espèces gram négatives et XerC de *L. leichmannii* est donc plus élevée que la différence entre cette dernière et XerD des gram positives et négatives.

L'arbre phylogénétique des séquences d'ADN se distingue de façon marquée de celui des séquences peptidiques par la position des séquences de *L. leichmannii* et *B. subtilis* par rapport aux autres séquences. Ainsi celles-ci se retrouvent complètement isolées du reste des espèces alors qu'on les retrouvait plus près du groupe des séquences XerD que du groupe des XerC dans l'arbre des peptides. Ceci pourrait partiellement s'expliquer par la dégénérescence du code génétique produisant une accumulation de différences et augmentant ainsi la distance entre ces deux espèces et les autres. Néanmoins, s'il ne s'agissait que de différences liées à la dégénérescence génétique, on ne s'attendrait pas nécessairement à voir un changement de position simultané et identique pour les deux

espèces. Ce phénomène pourrait peut-être s'expliquer si l'on conçoit les deux séquences comme faisant partie d'un groupe qui, tout en partageant plusieurs caractères avec XerC et XerD, s'en distingue néanmoins par certaines caractéristiques propres à celui-ci.

Des expériences sur des séquences complètes de gènes permettraient d'étayer les résultats obtenus pour les arbres phylogénétiques et en raffiner ainsi l'interprétation. Pour ce faire il faudrait réussir à mettre au point des amorces et un protocole de PCR permettant d'aller chercher les séquences externes aux 450 pb déjà séquencés. Nous suggérons l'utilisation de plusieurs témoins négatifs et positifs pour chacune des étapes préalables au PCR et également pendant celui-ci, afin de détecter si le problème que nous avons rencontré est inhérent au matériel génétique ou s'il s'agit d'un problème lié à la manipulation du matériel. Enfin, la mise au point d'amorces de PCR pour les espèces du genre *Lactococcus* et *Streptococcus* permettrait d'augmenter le nombre de recombinases mises au jour. De plus il serait intéressant de mettre au point les paramètres des hybridations Southern afin de détecter l'éventuelle présence de sites analogues au site *dif* pour ces mêmes genres. Cette détection pourrait également se faire par PCR inverse en utilisant des amorces fabriquées à partir des séquences des bras droit et gauche du site *dif*.

## CHAPITRE 5

### CONCLUSION

Les résultats que nous avons obtenus montrent que le système de recombinaison spécifique de site Xer existe chez plusieurs espèces de *Lactobacillus*. Ainsi nous avons découvert que l'espèce *Lactobacillus caseii* possède un site analogue à *dif* sur son chromosome. De plus, la présence d'une séquence codant pour une recombinase probablement analogue à XerD de *E. coli* a été détectée chez cette dernière espèce et chez les espèces *L. plantarum* 8014, *L. plantarum* 14917 et *L. bulgaricus* 737. Nous n'avons pas détecté la présence de recombinases chez les différentes espèces de *L. fermentii*, *Lactococcus lactis* spp *lactis*, *Lactococcus lactis* spp *cremoris* et *Streptococcus salivarius* 13419. Ces derniers résultats n'impliquent cependant pas que les recombinases sont absentes chez ces espèces.

Les régions séquencées des *Lactobacillus* partagent de nombreux points communs entre elles et avec les séquences déjà connues de recombinases (XerC et XerD de *E. coli*, *H. influenzae*, *P. aeruginosa*, *B. subtilis*, *L. leichmannii*) dont les plus nombreux se retrouvent dans le site actif situé au bout C-terminal de la recombinase. L'analyse phylogénétique des séquences de *Lactobacillus caseii*, *plantarum* et *bulgaricus* montrent que ceux-ci s'apparentent plus aux séquences XerD des espèces gram négatives *E. coli* et *H. influenzae* que de la séquence CodV (analogue à XerD) de la bactérie gram positive *B. subtilis*. Ces derniers résultats sont surprenants et pourraient impliquer une évolution différente pour les recombinases analogues à XerD chez les gram positives. Les résultats

obtenus pour l'arbre phylogénétique des séquences d'ADN montrent que les séquences de *L. leichmannii* et *B.subtilis* sont complètement isolées du reste des séquences alors qu'elles sont plus près du groupe des recombinaases XerD dans l'arbre des séquences peptidiques. Nous avons interprété cette différence en terme de dégénérescence du code génétique et également comme l'expression de caractères particuliers uniques aux séquences de *L.leichmannii* et *B.subtilis*.

## BIBLIOGRAPHIE

**Arciszewska, L., Grainge, I. et D.J. Sherratt** (1995). *Effects of Holliday Junction Position on Xer-Mediated Recombination in Vitro*. The EMBO Journal, Vol 14, no 11:2651-2660

**Arciszewska, L. et D.J. Sherratt**. (1995). *Xer Site-Specific Recombination in vitro*. The EMBO Journal, Vol 14, No 9:2112-2120

**Becker, J. et M. Brendel** (1996). *Molecular Characterization of the xerC Gene of Lactobacillus Leichmannii Encoding a Site-Specific Recombinase and Two Adjacent Heat Shock Genes*. Current Microbiology, Vol 32:232-236

**Blakely, G., Colloms, S., May, G., Burke, M. et D.J. Sherratt** (1991). *Escherichia coli XerC Recombinase is Required for Chromosomal Segregation at Cell Division*. The New Biologist, Vol 3, No 8: 789-798

**Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T. et D.J. Sherratt** (1993). *Two Related Recombinases are Required for Site-Specific Recombination at dif and cer in E. coli K12*. Cell, Vol 75:351-361

**Blakely, G.W. et D.J. Sherratt** (1994). *Interactions of the Site-Specific Recombinases XerC and XerD with the Recombination Site dif*. Nucleic Acids Research, Vol 22, No 25:5613-5620

**Clerget, M.** (1991). *Site-Specific Recombination Promoted by a Short DNA Segment of Plasmid R1 and by a Homologous Segment in the Terminus Region of the Escherichia coli Chromosome*. The New Biologist, Vol 3, No 8:780-788

**Colloms, S.D., Sykora, P., Szatmari, G. et D.J. Sherratt** (1990). *Recombination at ColE1 cer Requires the E. Coli xerC Gene Product, A Member of the Lambda Integrase Family of Site-Specific Recombinases*. Journal of Bacteriology, Vol.172, No12 pp 6973-6980

**Colloms, S.D., McCulloch, R., Grant, K., Neilson, L. et D.J. Sherratt** (1996). *Xer Mediated Site-Specific Recombination in vitro*. The EMBO Journal, Vol 15, No 5:1172-1181

**Cornet, F., Mortier, I., Patte, J. et J.M. Louarn** (1994). *Plasmid pSC101 Harbors a Recombination Site psi, Which is Able to Resolve Plasmid Multimers and to Substitute for the Analogous Chromosomal E. coli site dif*. Journal of Bacteriology, Vol 176, No 11, pp 3188-3195

**Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R. et al** (1995). *Whole genome Random Sequencing and Assembly of Haemophilus influenzae*. Science, 269:496-512

**Hill, T.M.** (1996). *Features of the Chromosomal Terminus Replication in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. Vol II, pp 1602-1613.

**Höfte, M., Doag, Q.G., Kourambas, S., Krishnapillai, D., Sherratt, D.J. et M. Mergeay** (1994). *The sss Gene Product, Which Affects Pyoverdinin Production in*

*Pseudomonas aeruginosa* 7NSK2 is a Site-Specific Recombinase. *Molecular Microbiology*, Vol 14(5):1011-1020

**Holo, H. et I.F. Nes** (1989). *High-Frequency Transformation, by Electroporation, of Lactococcus lactis subsp. cremoris Grown with Glycine in Osmotically Stabilized Media*. *Applied and Environmental Microbiology*, 55, Dec. 1989, p.3119-3123

**Kadner, R.J.**, (1996). *The Cytoplasmic Membrane in E.coli and Salmonella*. *Cellular and Molecular Biology*, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. Vol.I. pp. 58-97

**Kilby, N.J., Snaith M.R. et J.A.H. Murray** (1993). *Site-Specific Recombinases: Tools for Genomic Engineering*. *Trends in Genetics*, Vol 9, No 12:413-421

**Kobayashi, T., Midaka, M., Horiuchi, T.**(1989). *Evidence of a ter specific binding protein essential for the termination reaction of DNA replication in Escherichia coli*. *The EMBO Journal*, vol 8, no 8, pp 2435-2441

**Kuempel, P.L., Henson J.M., Dirks, L., Tecklenburg M. et D.F. Lim** (1991). *dif A RecA-Independent Recombination Site in the Terminus Region of the Chromosome of E.coli*. *The New Biologist*, Vol 3, No 8: 799-811

**Kuempel, P.L., Pelletier A.J. et T.M. Hill** (1989). *Tus and the Terminators: the Arrest of Replication in Procaryotes*. *Cell*, Vol 59: 581-583

**Leslie, N. et D.J. Sherratt** (1995). *Site Specific Recombination in the Replication Terminus Region of E.coli : Functional Replacement of dif*. *The EMBO Journal*, Vol 14, no 7:1561-1570

**Lloyd, R.G. et K. Brooks Low** (1996). *Homologous Recombination in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. Vol II.

**Lovett, S.T. et R.D.Kolodner** (1991). *Nucleotide Sequence of the Escherichia coli recJ Chromosomal Region and Construction of RecJ-Overexpression Plasmids*. Journal of Bacteriology , vol 173, no 1, p. 353-364

**Maniatis, T., Fritsch, E.F. et J. Sambrook, Eds** (1982). *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory

**Marians, K.J.** (1996). *Replication Fork Propagation in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. VolII.  
pp. 749-763

**Mathews, R.G.** (1996). *One-Carbon Metabolism in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. VolII.  
pp. 600-666

**McCulloch, R., Burke, M.L. et D.J. Sherratt** (1994a). *Peptidase Activity of Escherichia coli Aminopeptidase A is not Required for its Role in Xer Site-Specific Recombination*. Molecular Microbiology, Vol 12(2):241-251

**McCulloch, R., Coggins, L.W., Colloms, S.D. et D.J. Sherratt** (1994b). *Xer Mediated Site-Specific Recombination at cer Generates Holliday Junctions in vivo*. The EMBO Journal, Vol 13, No 8:1844-1855

**Messer, A. et C. Weigel** (1996). *Initiation of Cell Replication in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. Vol II, pp 1509-1661

**Moyer, C.L., Tiedje, J.M., Dobbs, F.C. et D.M. Karl** (1996). *A Computer-Simulated Restriction Fragment Length Polymorphism Analysis of Bacterial Small-Subunit rRNA Genes: Efficacy of Selected Tetrameric Restriction Enzymes for Studies of Microbial Diversity in Nature*. *Applied and Environmental Microbiology*, Vol. 62, No 7, p.2501-2507

**McMacken, R., Silver, L. et C. Georgopoulos** (1987). *DNA Replication in Escherichia coli and Salmonella typhimurium*, Cellular and Molecular Biology. Neidhardt, F.C., Ingraham, J.L., Brooks Low, Magasanik, B., Schaechter, M. et H. Edwin Umbarger Eds, Vol 1, Washington , pp. 564-612

**Pelletier, A.J., Hill, T.M. et P.L.Kuempel** (1988). *Location of Sites that Inhibit Progression of Replication Forks in the Terminus Region of Escherichia coli*. *Journal of Bacteriology*, Vol.170, No 9, pp 4293-4298

**Rothfield, L.I.** (1994). *Bacterial Chromosome Segregation*. *Cell*, Vol 77:963-966

**Sadowski, P.** (1986). *Site-Specific Recombinases: Changing Partners and Doing the Twist*. *Journal of Bacteriology*, Vol.165, No.2, pp 341-347

**Schmid, M.B. et U. von Freiesleben** (1996). *Nucleoid Segregation in E.coli and Salmonella*, Cellular and Molecular Biology, Frederick C. Neidhardt ed. In chief, ASM Press, Washington D.C. Vol II, pp 1662-1669

**Sherratt, D.J., Blakely, G., Burke, M., Colloms, S., Leslie, N., McCulloch, R., May, G. et J. Roberts** (1993). *Site-Specific Recombination and the Partition of Bacterial chromosomes In The Chromosome*. Heslop-Harrison, J., and Flavell R.B. Eds. Bios, Oxford, pp 25-41

**Sirois, S. et G. Szatmari**. (1995). *Detection of XerC and XerD Recombinases in Gram-Negative Bacteria of the Family Enterobacteriaceae*. Journal of Bacteriology, Vol 177, No 14, pp 4183-4186

**Stark, W., Boocock, M.R. et D.J. Sherratt** (1989). *Site-Specific Recombination by Tn3 Resolvase*. Trends in Genetics, Vol 5, No 9:304-309

**Stark, W., Boocock, M.R. et D.J. Sherratt** (1992). *Catalysis by Site-Specific Recombinases*. Trends in Genetics, Vol 8, No 12:432-439

**Stirling, C.G., Szatmari, G., Stewart, G., Smith M.C.M. et D.J. Sherratt** (1988). *The Arginine Repressor is Essential for Plasmid-Stabilizing Site-Specific Recombination at the Col E1 cer locus*. The EMBO Journal, Vol 7, No 13:4389-4395

**Summers, D.K.** (1989). *Derivatives of ColE1 cer Show Altered Topological Specificity in Site-Specific Recombination*. The EMBO Journal, Vol 8, No 1: 309-315

**Summers, D.K. et D.J. Sherratt** (1988). *Resolution of ColE1 Dimers Requires a DNA Sequence Implicated in the Three-Dimensional Organization of the cer Site*. The EMBO Journal, Vol 7, No3:851-858

**Summers, D.K.** (1991). *The kinetics of plasmid loss*. Trends in biotechnology, 9:273-278

**Summers, D.K., Beton, C.W., Withers, H.L.** (1993). *Multicopy plasmid instability: the dimer catastrophe hypothesis*. Molecular microbiology 8(6). 1031-1038

## ANNEXE I

Séquences des espèces utilisées pour fabriquer l'amorce A1 (du nucléotide N° 404 au nucléotide N° 429) :

XerC d' <i>H.influenzae</i>	GTG ATC GTG CTA TTT TGG AAT T
XerC de <i>E.coli</i>	GCG ACC GTG CAA TGC TGG AAG T
SSS de <i>P.aeruginosa</i>	GCG ACC AGG CCC TGC TGG AGC T
XerD de <i>E.coli</i>	GCG ATA AAG CCA TGC TTG AAG T
XerD d' <i>H.influenzae</i>	GTG ATA AAG CAA TGT TAG AGT T
CodV de <i>B.subtilis</i>	GGG ATC AAG CGC TGT TAG AGC T

Séquences des espèces utilisées pour fabriquer l'amorce C (du nucléotide N °834 jusqu'au nucléotide N °852):

XerC d' <i>H.influenzae</i>	CAC ACA AAT TTA TAC TCA C
XerC de <i>E.coli</i>	CAC GCA AAT CTA TAC TCA T
SSS de <i>P.aeruginosa</i>	GAC CCA GAT CTA TAC CCA C
XerD de <i>E.coli</i>	CAC GCA AAT TTA TAC GCA T
XerD d' <i>H.influenzae</i>	AAC ACA GAT TTA TAC TCA C
CodV de <i>B.subtilis</i>	TAC ACA GAT ATA CAC GCA C

## ANNEXE II

Séquences (éditées) des gènes des recombinases de *xerC* d'*H. influenzae*, *xerC* de *E. coli*, *sss* de *P.aeruginosa*, *xerD* de *E.coli*, *xerD* d'*H.influenzae*, *L. plantarum* 8014, *L. caseii*, *L. bulgaricus* 737, *xerC* de *L. leichmannii* et *codV* de *B.subtilis*

	1				50
leichseqm	GCGCTTTTT	ACGCGACCGG	GATGCGGCTG	AGCGAACTGA	CTGACTTGAA
basubseqm	GAGCTGCTCT	ATGCCACTGG	AATGAGGGTC	AGTGAATGCT	GTTCCATAAC
influseqm	GAATTAATGT	ATAGTTCAGG	TTTGCGTTTA	TCTGAATTAC	AAGGCTTAGA
colicseqm	GAAGTGATGT	ACGGCGCGGG	TCTGCGTCTT	TCTGAGCTGG	TGGGGCTGGA
psaerseqm	GAGCTGTTCT	ATTCCTCCGG	GCTGCGCCTC	TCGGAAC TGG	TCGGCCTCGA
colidseqm	GAAGTGTTGT	ATGCTACCGG	GCTGCGTGTC	TCTGAACTGG	TCGGACTGAC
infldseqm	GAGTTGCTTT	ATGCAACGGG	GCTTCGTGTA	ACGGAGTTAG	TTTCGCTTAC
caseiseqm	GAGGTGATGT	ATGCGACTGG	CTTGCGGGTA	TCGGAAGTTG	TCGGATTGCG
pla80seqm	GAGGTGATGT	ATGCGACCGG	TCTGCGAGTC	AGCGAGACAA	TTAATTTAAC
bul73seqm	GAACTGCTTT	ACGCGACCGG	GATGCGGGTC	AGCGAAGCGG	TCAACTTGCA
	51				100
leichseqm	AATTAAGCAG	CTTGACCTGG	AAAACGGCAT	GATTTTGGTG	CACGGCAAGG
basubseqm	TATTAACGAC	GTTGATTTAT	TTATGGACAC	TGTGCTTGTT	CACGGTAAAG
influseqm	TTTAAACAGC	ATTAATACCC	GTGTGCGGGA	AGTACGCGTT	ATTGGA AAAAG
colicseqm	TATTAACAC	CTCGACCTGG	AGTCTGGTGA	AGTGTGGGTT	ATGGGGAAAAG
psaerseqm	TCTCGAATGG	CTGGACCTCA	AGGAAGGCCT	GGTGCAGAGT	CGCGGCAAGG
colidseqm	AATGAGTGAT	ATCAGCCTGC	GTCAGGGCGT	GGTACGGGTC	ATTGGTAAAG
infldseqm	GATTGAAAAC	ATGAGCGTGC	AACAAGGCGT	GGTGCAGAGT	ATTGGTAAAG
caseiseqm	ATTGGATCAG	TTACATTTAG	CAATGAACTT	GCTTCAGGTT	ACCGGAAAAG
pla80seqm	GATGGATGAT	CTGCACTTAG	ACTTGGGATT	GATTCAAACA	ATCGGTAAAG
bul73seqm	ATTAGCCGAT	GTCCATGCCA	ACCTGCAGCT	GATCAAAGTC	CTGGGCAAGG
	101				150
leichseqm	GCAATAAGGA	CCGTTATGTC	TTTTTTGACC	AAGAGAGCAA	AAAGTACTTG
basubseqm	GCAAGAAGCA	GCGCTATATC	CCCTTTGGGT	CTTATGCCCG	CGAAGCGTTG
influseqm	GTAACAAAGA	ACGAGTTGTG	CCGTTTGGGC	GTTACGCTTC	GCAATGCCATT
colicseqm	GCAGCAAAGA	GCGCCGCTG	CCGATTGGTC	GCAACGCTGT	GGCGTGGATT
psaerseqm	GCAACAAGGT	CCGCGAGCTG	CCGGTCGGCA	AGGCCGCGCG	CCAGGCGCTC
colidseqm	GCAACAAGA	GCGTCTGGTG	CCGTTAGGTG	AAGAGGCGGT	TTACTGGCTG
infldseqm	GGAACAAGGA	ACGTATTGTG	CCAATGGGGG	AAGAAGCTGC	CTATTGGGTT
caseiseqm	GCGATAAGGA	GCGACTGGTT	CCAATCAGTC	CCCAAGCAGC	CGATTGGGTT
pla80seqm	GTGATAAAGA	GCGGATCATT	CCTATCGGTG	ATGTTGCCAT	TGATTTTATT
bul73seqm	GGAGCAAGGA	ACGGCTTATT	CCGGTGAGCG	CCCTGGCCTT	AGACTGGATT
	151				200
leichseqm	GAGGAATATC	TGCAGGTAGC	CCGGCCAAGT	TTATTGAAAA	ATGAGCCAGA
basubseqm	AAGGTATACA	TGAATAGCGG	AAGACAGTGC	TTGCTGATGA	AGGCAAAAAG
influseqm	CAAGAATGGT	TAAAAGTGCG	GGCATTATT.	.....	.TAACCCAAA
colicseqm	GAGCACTGGC	TTGATTTGCG	CGACCTGTT.	.....	.TGGTAGCGA
psaerseqm	GAAGCCTGGC	TGCCGCTGCG	CACCCAGGC.	.....	.CGCTCCGGA
colidseqm	GAAACCTATC	TGGAACATGG	GCGTCCGTGG	CTGCTGAAT.	..GGTGTGTC
infldseqm	CGTCAATTTA	TGCTTTATGG	TCGCCCTGTG	CTATTAAAT.	..GGGCAGAG
caseiseqm	AACCGGTATT	TGCAGGAATC	CCGTCCGCGT	TTGATCAAAC	ACCAGCAACC
pla80seqm	ACCCGTCATT	TAAAAACGGC	TCGGCCCAA	CTCGTCTCTG	CCAAACGCCG
bul73seqm	AAGCGCTACC	TGCCTGTGAG	GGACCAGCAG	CTTTTAAAGC	GGGGCAAGAG

	201				250
leichseqm	...TACCGAG	GCAGTCTTTT	TAAACAAGCT	CGGTAGGCCA	ATTTCCAGCC
basubseqm	ACCTCATGAT	CTATTATTCTG	TAAATCAAAG	AGGCGGACCG	CTTACAGCCC
influseqm	...AGATGAG	GCATTATTTG	TTAGCCAGCT	TGGAAATCGC	ATTTCTCATC
colicseqm	...AGACGAC	GCGCTTTTTT	TGTCGAAACT	GGGCAAGCGT	ATCTCCGCGC
psaerseqm	...AGACGGT	GCGGTGTTCA	TCGGCCGCG	CGGCAAGCGC	CTGACGCCAC
colidseqm	...AATTGAC	GTGTTGTTTC	CCAGCCAGCG	TGCGCAGCAG	ATGACGCGAC
infldseqm	...TTCTGAT	GTTGTGTTTC	CAAGTCAGCG	CGCACAGCAA	ATGACTCGTC
caseiseqm	...CAA...G	GCAGTGTTCG	TCAACTTTCA	TGGTCATGCG	CTCACCCGTC
pla80seqm	...GAATAGC	TATTTGTTTA	TTAATAATCA	CGGTGGTAAA	TTAAGTCGGC
bul73seqm	...TAGCGAT	TACCTGTTCT	TGAACAGCCG	GGGCGGGCAG	TTGACCCGGC
	251				300
leichseqm	GGGGGATTGC	CAAAGCCGTC	CAGCAAATTT	TTCAAAAAGC	AGGCTTGACG
basubseqm	GTGGCATCAG	ACATATTTTA	AGCGGGCTTG	TTCAAAAAGC	GTCAAGCACT
influseqm	GAGCTATTCA	AAAACGTTTA	GAAACTTGGG	GTATTCGTCA	AGGCTTAAAT
colicseqm	GTAATGTGCA	GAAACGCTTT	GCCGAATGGG	GCATAAAACA	AGGGCTGAAT
psaerseqm	GGGCGATCCA	GCTGCGGGTG	CGCCAGGCCG	GCGT...CCG	CGAGCTGGGC
colidseqm	AGACCTTCTG	GCACCGTATT	AAACATTATG	CTGTGCTGGC	GGGTATCGAC
infldseqm	AAACTTTCTG	GCATCGTGTA	AAACATTATG	CCATTCTCGC	GGATATTGAT
caseiseqm	AAGCAATTTG	GAAGAATCTG	AAAGCCTATA	TTGCAAGTGT	TGGCATTGAA
pla80seqm	AAGGGGTTTG	GAAAAATCTG	AAGGCAGAAG	TCAAGGCGGC	TGGCATCGAA
bul73seqm	AGGCCGTCTG	GCAGAAGATC	AAGAAGTACT	GCCAGCAGGC	CGGCATTGCC
	301				350
leichseqm	GCTGGTG...	CCCACCCCCA	TGAACTCCGC	CATAGCTTCG	CCACCGCCAT
basubseqm	TTACATA...	TCCATCCGCA	TATGCTTCGA	CATACGTTCC	CCACGCATCT
influseqm	AGCCAT...C	TTAATCCGCA	CAAATTACGT	CATTCTTTTG	CGACGCATAT
colicseqm	AATCAC...G	TTCATCCGCA	TAAATTACGT	CACTCGTTCC	CCACGCATAT
psaerseqm	CAGCAC...C	TGCACCCGCA	CATGCTGCGG	CACTCGTTCC	CCAGCCATCT
colidseqm	AGCGAAAAGC	TGTCACCGCA	TGTGTTGCGT	CACGCTTTTG	CCACTCATTT
infldseqm	GCTGATGCAC	TTTCTCCTCA	CGTTCTTCGC	CATGCTTTTG	CCACCCATTT
caseiseqm	AAGGATG...	TAACACCGCA	TACGTTACGG	CACAGTTTTG	CCACGCGTCT
pla80seqm	AAAAACG...	TGACGCCGCA	TACGTTGCGC	CACTCGTTTG	CGACTCACAT
bul73seqm	AAGGATG...	TGACTCCCCA	CACCCTCCGC	CACAGCTTCG	CGACCCATTT
	351				400
leichseqm	GCTCAACAAC	GGGGCAGACC	TCAGGAGCGT	GCAGGAGCTC	TTGGGCCATG
basubseqm	GTTAAATGAA	GGAGCGGATT	TGAGAAGCGT	TCAAGAACTG	CTCGGGCATT
influseqm	GTTGGAAGCC	AGTTCTGATT	TACGCGCCGT	GCAAGAATTG	CTTGGGCATA
colicseqm	GCTGGAGTCG	AGCGGCGATC	TTCGTGGTGT	GCAGGAGCTG	CTGGGTCATG
psaerseqm	GCTGGAGTCG	TCCGGCGACC	TGCGCGCCGT	GCAGGAACTG	CTCGGCCACG
colidseqm	ATTAAATCAT	GGTGCGGATT	TACGCGTGGT	GCAGATGCTA	CTGGGCCACA
infldseqm	AGTAAATCAC	GGTGCGGATT	TACGCGTGGT	GCAAATGTTG	TTGGGTCATA
caseiseqm	GTTGGAATAA	GGCGCCGATT	TACGGGTGGT	TCAGGAGTTA	CTGGGGCATA
pla80seqm	CTTAGAAAAT	GGTGCGGACT	TGCGTGTCGT	TCAGGAGCTT	TTAGGGCATG
bul73seqm	GCTGGAGCAT	GGGGCGGATT	TGCGGGTGGT	CCAGGAAATT	CTGGGCCATA

	401		432
leichseqm	AGGACTTGTC	TACGACCCAG	ATCTATACTC AT
basubseqm	CCAATCTGTC	TTCTACACAG	ATATACACGC AC
influseqm	GCAACCTTTC	CACCACACAA	ATTTATACTC AC
colicseqm	CCAACCTCTC	CACCACGCAA	ATCTATACTC AT
psaerseqm	CCGACATCGC	CACGACCCAG	ATCTATACCC AC
colidseqm	GCGATCTCTC	CACCACGCAA	ATTTATACGC AT
infldseqm	CAGATTTATC	TACAACACAG	ATTTATACTC AC
caseiseqm	GTGATATTAG	TACCACCCAA	ATCTACACGC AT
pla80seqm	CCGATATTTTC	GACGACCCAA	ATTTATACGC AC
bul73seqm	CTGACATTTTC	CACGACCCAA	ATATATACAC AC

Légende :Influseqm = xerC d'*H.influenzae*, colicseqm = xerC de *E.coli*,  
 psaerseqm = sss de *P.aeruginosa*, colidseqm = xerD de *E.coli*, infldseqm  
 = xerD d'*H.influenzae*, pla80seqm = *L.plantarum* 8014, caseiseqm =  
*L.caseii*, bul73seqm = *L.bulgaricus* 737, leichseqm = xerC  
*L.leichmannii*, basubseqm = codV de *B.subtilis*.

## ANNEXE III

A- Consensus absolu entre les séquences peptidiques de XerC d'*H. influenzae*, XerC de *E. coli*, Sss de *P.aeruginosa*, XerD de *E.coli*, XerD d'*H.influenzae*, *L. plantarum* 8014, *L. caseii*, *L. bulgaricus* 737, XerC de *L. leichmannii* et CodV de *B.subtilis*

(\* = consensus, - = pas de consensus)

	1	DOMAINE I				50
{influpepm}	ELMYS	[SGLRL	SELQGLDLNS	INTRVREVRV	IGKGNKERVV	PFGRYASHAI
{colicpepm}	EVMYG	[AGLRL	SELVGLDIKH	LDLESGEVWV	MKGKSKERRL	PIGRNAVAWI
{psaerpepm}	ELFYS	[SGLRL	SELVGLDLEW	LDLKEGLVRV	RGKGNKVREL	PVGKAARQAL
{colidpepm}	EVLYA	[TGLRV	SELVGLTMSD	ISLRQGVVRV	IGKGNKERLV	PLGEEAVYWL
{infldpepm}	ELLYA	[TGLRV	TELVSLTIEN	MSVQQGVVRV	IGKGNKERIV	PMGEEAAYWV
{pla80pepm}	EVMYA	[TGLRV	SETINLTMD	LHLDLGLIQT	IGKGDKERII	PIGDVAIDFI
{caseipepm}	EVMYA	[TGLRV	SEVVGRLDQ	LHLAMNLLQV	TGKGDKERLV	PISPQAADWV
{bul73pepm}	ELLYA	[TGMRV	SEAVNLQLAD	VHANLQLIKV	LGKGSKERLI	PVSALALDWI
{leichpepm}	ALFYA	[TGMRL	SELTDLKIKQ	LDLENGMILV	HGKGNKDRYV	FFDQESKKYL
{basubpepm}	ELLYA	[TGMRV	SECCSITIND	VDFMMDTVLV	HGKGGKQRYI	PFGSYAREAL
Consensus	----	*-*	-*	-----	-----	-----
	51					100
{influpepm}	QEWLK	.....	VRALFNPKE	ALFVSQNGNR	ISHRAIQKRL	ETWGIROGLN
{colicpepm}	EHWLD	.....	LRDLFGSEDD	ALFLSKLGKR	ISARNVQKRF	AEWGIKQGLN
{psaerpepm}	EAWLP	.....	LRTQAAPEDG	AVFIGRSGKR	LTPRAIQLRV	RQAGVRE.LG
{colidpepm}	ETYLEHGRPW	LL.N.GV	SID	VLFPSQRAQQ	MTRQTFWHRI	KHYAVLAGID
{infldpepm}	RQFMLYGRPV	LL.N.GQ	SSD	VVFPQRAQQ	MTRQTFWHRV	KHYAILADID
{pla80pepm}	TRHLKTARPK	LV.SAK	RRNS	YLFINNHGGK	LSRQGVWKNL	KAEVKAAGIE
{caseipepm}	NRYLQESRPR	LI.KHQ	QPKA	.VFVNFHGH	LTRQAIWKNL	KAYIASVGIE
{bul73pepm}	KRYLPVRDQQ	LL.KRG	KSSD	YLFNLSRGGQ	LTRQAVWQKI	KKYCCQAGIA
{leichpepm}	EEYLQVARPS	LLKNE	PD.T.E	AVFLNKLGRP	ISSRGIKAV	QQIFQKAGLT
{basubpepm}	KVYMNSGRQC	LLMKAKE	PHD	LLFVNQRGGP	LTARGIRHIL	SGLVQKASST
Consensus	-----	-----	-----	-----	-----	-----
	101					144
{influpepm}	SH.LNPHKLR	HS]	[FATHMLEA	SSDLRAVQEL	LGHSNLSTTQ	IYTH]
{colicpepm}	NH.VHPHKLR	HS]	[FATHMLES	SGDLRGVQEL	LGHANLSTTQ	IYTH]
{psaerpepm}	QH.LHPHMLR	HS]	[FASHLLES	SGDLRAVQEL	LGHADIATTQ	IYTH]
{colidpepm}	SEKLSPHVLR	HA]	[FATHLLNH	GADLRVVQML	LGHSDDLSTTQ	IYTH]
{infldpepm}	ADALSPHVLR	HA]	[FATHLVNH	GADLRVVQML	LGHTDLSTTQ	IYTH]
{pla80pepm}	KN.VTPHTLR	HS]	[FATHILEN	GADLRVVQEL	LGHADISTTQ	IYTH]
{caseipepm}	KD.VTPHTLR	HS]	[FATRILEN	GADLRVVQEL	LGHSDISTTQ	IYTH]
{bul73pepm}	KD.VTPHTLR	HS]	[FATHLLEH	GADLRVVQEI	LGHTDISTTQ	IYTH]
{leichpepm}	AG.AHPHELRL	HS]	[FATAMLNN	GADLRVQEL	LGHEDLSTTQ	IYTH]
{basubpepm}	LH.IHPHMLR	HT]	[FATHLLNE	GADLRVQEL	LGHSNLSSTQ	IYTH]
Consensus	-----*	*-]	[**-----	-----*	-----*	****]

Légende :Influpepm = XerC d'*H.influenzae*, colicpepm = XerC de *E.coli*, psaerpepm = Sss de *P.aeruginosa*, colidpepm = XerD de *E.coli*, infldpepm = XerD d'*H.influenzae*, pla80pepm = *L.plantarum* 8014, caseipepm = *L.caseii*, bul73pepm = *L.bulgaricus* 737, leichpepm = XerC *L.leichmannii*, basubpepm = CodV de *B.subtilis*.

ACIDES AMINES EN CARACTERES GRAS = RESIDUS COMPLETEMENT CONSERVES PARMIS LES A-  
INTEGRASES

B-Consensus absolu entre les séquences d'ADN de *xerC* d'*H. influenzae*, *xerC* de *E. coli*, *sss* de *P.aeruginosa*, *xerD* de *E.coli*, *xerD* d'*H.influenzae*, *L. plantarum* 8014, *L. caseii*, *L. bulgaricus* 737, *xerC* de *L. leichmannii* et *codV* de *B.subtilis*

(\* = consensus, - = pas de consensus)

	1				50
{leichseqm}	GCGCTTTTTT	ACGCGACCGG	GATGCGGCTG	AGCGAACTGA	CTGACTTGAA
{basubseqm}	GAGCTGCTCT	ATGCCACTGG	AATGAGGGTC	AGTGAATGCT	GTTCCATAAC
{influseqm}	GAATTAATGT	ATAGTTCAGG	TTTGCCTTTA	TCTGAATTAC	AAGGCTTAGA
{colicseqm}	GAAGTGATGT	ACGGCGCGGG	TCTGCGTCTT	TCTGAGCTGG	TGGGGCTGGA
{psaerseqm}	GAGCTGTCT	ATTCCTCCGG	GCTGCGCCTC	TCGGAACTGG	TCGGCTCGA
{colidseqm}	GAAGTGTGT	ATGCTACCGG	GCTGCGTGTG	TCTGAACTGG	TCGGACTGAC
{infldseqm}	GAGTTGCTTT	ATGCAACGGG	GCTTCGTGTA	ACGGAGTTAG	TTTCGCTTAC
{caseiseqm}	GAGGTGATGT	ATGCGACTGG	CTTGCGGGTA	TCGGAAGTTG	TCGGATTGCG
{pla80seqm}	GAGGTGATGT	ATGCGACCGG	TCTGCGAGTC	AGCGAGACAA	TTAATTTAAC
{bul73seqm}	GAACTGCTTT	ACGCGACCGG	GATGCGGGTC	AGCGAAGCGG	TCAACTTGCA
Consensus	*-----*	*-----*	---*-----	-----*	-----*

	51				100
{leichseqm}	AATTAAGCAG	CTTGACCTGG	AAAACGGCAT	GATTTTGGTG	CACGGCAAGG
{basubseqm}	TATTAACGAC	GTTGATTTAT	TTATGGACAC	TGTGCTTGTT	CACGGTAAAG
{influseqm}	TTTAAACAGC	ATTAATACCC	GTGTGCGGGA	AGTACGGGTT	ATTGGAAAAG
{colicseqm}	TATTAACAC	CTCGACCTGG	AGTCTGGTGA	AGTGTGGGTT	ATGGGGAAAG
{psaerseqm}	TCTCGAATGG	CTGGACCTCA	AGGAAGGCCT	GGTGCAGATG	CGCGGCAAGG
{colidseqm}	AATGAGTGAT	ATCAGCCTGC	GTCAGGGCGT	GGTACGGGTC	ATTGGTAAAG
{infldseqm}	GATTGAAAAC	ATGAGCGTGC	AACAAGGCGT	GGTGCAGATG	ATTGGTAAAG
{caseiseqm}	ATTGGATCAG	TTACATTTAG	CAATGAACCT	GCTTCAGGTT	ACCGGAAAAG
{pla80seqm}	GATGGATGAT	CTGCACCTAG	ACTTGGGATT	GATTCAAACA	ATCGGTAAAG
{bul73seqm}	ATTAGCCGAT	GTCCATGCCA	ACCTGCAGCT	GATCAAAGTC	CTGGGCAAGG
Consensus	---*-----	---*-----	---*-----	---*-----	---*-----

	101				150
{leichseqm}	GCAATAAGGA	CCGTTATGTC	TTTTTTGACC	AAGAGAGCAA	AAAGTACTTG
{basubseqm}	GCAAGAAGCA	GCGCTATATC	CCCTTTGGGT	CTTATGCCCG	CGAAGCGTTG
{influseqm}	GTAACAAAGA	ACGAGTTGTG	CCGTTTGGGC	GTTACGCTTC	GCATGCCATT
{colicseqm}	GCAGCAAAGA	GCGCCGCTTG	CCGATTGGTC	GCAACGCTGT	GGCGTGGATT
{psaerseqm}	GCAACAAGGT	CCGCGAGCTG	CCGGTCGGCA	AGGCCGCGCG	CCAGGCCTC
{colidseqm}	GCAACAAAGA	GCGTCTGGTG	CCGTTAGGTG	AAGAGGCGGT	TTACTGGCTG
{infldseqm}	GGAACAAGGA	ACGTATTGTG	CCAATGGGGG	AAGAAGCTGC	CTATTGGGTT
{caseiseqm}	GCGATAAGGA	GCGACTGGTT	CCAATCAGTC	CCCAAGCAGC	CGATTGGGTT
{pla80seqm}	GTGATAAAGA	GCGGATCATT	CCTATCGGTG	ATGTTGCCAT	TGATTTTATT
{bul73seqm}	GGAGCAAGGA	ACGGCTTATT	CCGGTGAGCG	CCCTGGCCTT	AGACTGGATT
Consensus	*-----*	---*-----	---*-----	---*-----	---*-----

	151				200
{leichseqm}	GAGGAATATC	TGCAGGTAGC	CCGGCCAAGT	TTATTGAAAA	ATGAGCCAGA
{basubseqm}	AAGGTATACA	TGAATAGCGG	AAGACAGTGC	TTGCTGATGA	AGGCAAAAGA
{influseqm}	CAAGAATGGT	TAAAAGTGCG	GGCATTATT.	.....	.TAACCCAAA
{colicseqm}	GAGCACTGGC	TTGATTTGCG	CGACCTGTT.	.....	.TGGTAGCGA
{psaerseqm}	GAAGCCTGGC	TGCCGCTGCG	CACCCAGGC.	.....	.CGCTCCGGA
{colidseqm}	GAAACCTATC	TGGAACATGG	GCGTCCGTGG	CTGCTGAAT.	..GGTGTGTC
{infldseqm}	CGTCAATTTA	TGCTTTATGG	TCGCCCTGTG	CTATTAAT.	..GGGCAGAG
{caseiseqm}	AACCGGTATT	TGCAGGAATC	CCGTCCGCGT	TTGATCAAAC	ACCAGCAACC
{pla80seqm}	ACCCGTCATT	TAAAACGGC	TCGGCCCAA	CTCGTCTCTG	CCAAACGCCG
{bul73seqm}	AAGCGCTACC	TGCCTGTCAG	GGACCAGCAG	CTTTTAAAGC	GGGGCAAGAG
Consensus	-----	*-----	-----	-----	-----



## REMERCIEMENTS

Je remercie mon directeur de recherche Dr. George Szatmari pour la patience et la générosité dont il a fait preuve à mon égard tout au long de ma maîtrise.

Je tiens également à remercier Dr. Nicolina Zakova Pavlova pour m'avoir donné la piqure de la génétique et pour toute l'aide qu'elle m'a procurée au début de ma maîtrise. Je remercie aussi Dr Barbara Karska-Wysoki pour m'avoir fourni plusieurs souches de bactéries lactiques et pour ses conseils à propos de la manipulation de celles-ci .

Je voudrais aussi remercier Bernard Chapais pour son soutien et ses encouragements constants; Josiane Courteau pour avoir patiemment enduré mes longs retards à l'heure du dîner; et ma famille pour toute la confiance qu'elle m'a témoignée.

Enfin, je remercie mes compagnons de travail : Nicolas Réhel, alias Ronald, et Stéphane Sirois, alias Stéfoz, pour toute l'aide et le soutien qu'ils m'ont donnés et aussi pour les heures hilarantes que nous avons passées au laboratoire.

Finalement, je voudrais remercier les fonds FCAR pour m'avoir octroyé une bourse sans laquelle je n'aurais pas pu faire ces études.