



AVERTISSEMENT

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

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

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

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

http://www.cfcopies.com/V2/leg/leg_droi.php

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



Etude structurale et fonctionnelle du cluster *eps* de *Streptococcus thermophilus* IP6756 : spécificités et hypothèses nouvelles

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy 1

en Génétique Moléculaire

par **Guillaume TYVAERT**

Soutenue le 16 novembre 2005 devant la commission d'examen :

Membres du jury :

Président : P. Ritzenthaler, Professeur, Université Paul Sabatier, Toulouse

Rapporteurs : P. Lejeune, Professeur, INSA, Lyon
P. Ritzenthaler, Professeur, Université Paul Sabatier, Toulouse

Examineurs : P. Hols, Professeur, UCL, Louvain-La-Neuve
B. Decaris, Professeur, UHP, Nancy (Directeur de thèse)
F. Charron-Bourgoin, Maître de Conférence, UHP, Nancy

Je remercie Monsieur le Professeur Paul Ritzenthaler et Monsieur le Professeur Philippe Lejeune pour avoir accepté d'être rapporteurs de ce travail.

Merci également à Monsieur le Professeur Pascal Hols pour avoir accepté de juger ce travail.

REMERCIEMENTS

Je remercie le Professeur Decaris pour son accueil au sein de son laboratoire. Il m'a donné l'occasion de confronter connaissances théoriques et réalité du travail expérimental, en me proposant d'étudier les EPS de *Streptococcus thermophilus* et, par là, la variabilité génétique.

Je remercie le Docteur Charron-Bourgoin pour le co-encadrement de cette thèse, notamment pour la correction minutieuse du manuscrit, et Catherine Morel pour son assistance technique et nos conversations littéraires.

Je tiens également à remercier le Docteur Jolly, chimiste, avec qui j'ai pu vivre concrètement ce qu'est la pluridisciplinarité scientifique.

Ma gratitude va également aux membres du laboratoire dont j'ai apprécié la présence attentive tout au long de mon travail, notamment Brigitte Gintz, Céline Fourier, Alexandra Kleinclauss, Laurence Hotel, Katy Poiret et Paule Diebold.

Je n'aurais garde d'oublier les personnes qui ont contribué au cours de discussions informelles mais approfondies à ma réflexion, notamment les Docteurs Virginie Libante, Annabelle Thibessard et Gérard Guédon.

J'associe à ces remerciements Jessy et Frédéric, avec qui j'ai pu travailler à l'occasion de leurs stages ainsi que Jean-Daniel et Hervé qui m'ont apporté un éclairage extérieur toujours utile.

Ces remerciements seraient incomplets si je n'évoquais pas aussi Sibel, Kévin, Axelle, Jérôme, Jessy, Laeticia, Alexandre, Chloé, Céline, Philipe, Marc-Olivier, Jean-Yves, Muriel, Thomas, Sandra, Sébastien, Emilie, Marie-Amélie, Lars, Bertrand, Bruno, Brigitte, Alexandra, Laurence, Frédérique, Romain et Mickael, et, spécialement, Nicolas, Alexandre et Magali (pour leur patience, la Hot-Line informatique, le Nico's game, et nos moments de détente sur une musique de cirque). Merci pour tous les bons moments passés ensemble qui nous ont permis de tisser de véritables liens d'amitié au cours des trois années d'allocation-monitorat débouchant sur la présentation de ce mémoire de thèse.

J'adresse tous mes encouragements à ceux qui entament leur troisième cycle, et à tous, je souhaite « tout le bonheur du monde ».

à mes parents
à Anne et Claire

ABREVIATIONS

S.C.D. - B.P. 11
BIBLIOTHÈQUE D'AGRICULTURE
Rue du Jardin Botanique - BP 11
54601 VILLERS-LES-NANCY Cedex

aa : acide aminé
ADN : Acide DésoxyriboNucléique
ADNc : Acide DésoxyriboNucléique complémentaire
ARN : Acide RiboNucléique
ATCC : *American Type Culture Collection*
ATP : Adénosine TriPhosphate
BET : Bromure d'ETHidium
com. pers. : communication personnelle
CNRZ : Centre National de la Recherche Zootechnique
CPS : Capsular Polysaccharide (Polysaccharide de capsule)
CRE : *Catabolite Response Element* (Elément de réponse catabolique)
Da : Dalton
DO : Densité optique
 ΔG : Entropie
EPS : Exopolysaccharide
et al. : et collaborateurs
ICE : *Integrative and conjugative element* (élément conjugatif et intégratif)
GRAS : *Generally recognised as safe species* (Innocuité toxicologique)
IP : Institut Pasteur
IRL : *Left Inverted Repeat* (Répétition inversée gauche)
IRR : *Right Inverted Repeat* (Répétition inversée droite)
IS : *Insertion Sequence* (Séquence d'insertion)
kb : kilobase
LB : Luria-Bertani
LPS : Lipopolysaccharide
LMG : Collection de l'Université de Gand
LR-PCR : Long Range PCR
Mb : mégabase
NBT : NitroBlue Tétrazolium
NCFB : *National Collection of Food Bacteria*
ND : Non déterminé

NST : Nancy *Streptococcus thermophilus*

nt : nucléotide

ORF : *Open reading frame* (Phase ouverte de lecture)

pb : paire de base

PCR : *Polymerase chain reaction* (Réaction d'amplification en chaîne)

pH : potentiel Hydrogène

p/v : poids par volume

RBS : *Ribosome binding site* (Site de fixation du ribosome)

RM : Restriction-Modification

rpm : rotation par minute

RT-PCR : *Reverse transcriptase polymerase chain reaction*

SDS : Sodium Dodécyl Sulfate

SSC : Standard Sodium Citrate

subsp. : *subspecies*

TAE : Tris-Acétate-EDTA

TBE : Tris Borate EDTA

TLC : *Thin Layer Chromatography* (Chromatographie sur couche mince)

T_m : *melting Temperature* (Température de fusion)

TPPY : *Tryptose Proteose Peptone Yeast Extract*

UDP : UridineDiPhosphate

UE : Unité Enzymatique

UV : Ultra-Violet

V/cm : Volt par centimètre

v/v/v/v : volume par volume par volume par volume

W : Watt

X-Phosphate : 5-bromo-4-chloro-3-indolyl-phosphate

Cette étude a fait l'objet de la publication suivante :

The *eps* locus of *Streptococcus thermophilus* IP6756 is not involved in exopolysaccharide production

Guillaume TYVAERT, Catherine MOREL, Jean-Pierre JOLY, Bernard DECARIS and
Florence CHARRON-BOURGOIN

International Dairy Journal, sous presse

SOMMAIRE

<u>PREAMBULE</u>	7
<u>INTRODUCTION</u>	11
1 LES EXOPOLYSACCHARIDES (EPS)	11
1.1 DIVERSITE DES POLYSACCHARIDES DE L'ENVELOPPE CELLULAIRE CHEZ LES BACTERIES.....	11
1.1.1 <i>Le peptidoglycane : constituant principal de la paroi bactérienne</i>	11
1.1.2 <i>Polysaccharides spécifiques des bactéries à coloration de Gram positive</i>	11
1.1.3 <i>Les lipopolysaccharides sont spécifiques des bactéries à coloration de Gram négative</i>	13
1.1.4 <i>Les exopolysaccharides forment des capsules et/ ou des amas visqueux (« slime »)</i>	15
1.2 IMPORTANCE POUR LA BACTERIE DES INTERACTIONS ENTRE LES EPS ET L'ENVIRONNEMENT	16
1.2.1 <i>Importance des EPS de Rhizobium dans l'établissement d'une symbiose fonctionnelle</i>	16
1.2.2 <i>Rôle primordial des EPS lors de l'établissement d'un biofilm</i>	17
1.2.3 <i>Les EPS : barrière de protection face aux agressions du milieu extérieur</i>	18
1.3 TRES GRANDE DIVERSITE STRUCTURALE DES EPS.....	20
1.3.1 <i>Avantage sélectif conféré aux bactéries pathogènes par la variabilité de leurs EPS</i>	24
1.3.2 <i>Utilisation industrielle de la variabilité des EPS produits par les bactéries lactiques non pathogènes</i>	26
1.4 BIOSYNTHESE DES EPS	27
1.4.1 <i>Mécanisme ubiquitaire de biosynthèse des EPS</i>	27
1.4.2 <i>Le cas particulier de la biosynthèse des EPS simples</i>	30
1.4.3 <i>Interconnexion de la biosynthèse des EPS avec d'autres voies métaboliques</i>	31
1.5 VARIABILITE DES CLUSTERS EPS/CPS CHEZ LES BACTERIES LACTIQUES ET LES PROTEOBACTERIES	32
1.5.1 <i>Complexité du cluster eps de S. thermophilus</i>	33
1.5.1.1 <i>Organisation opéronique et modulaire de la première région du cluster eps (3' d'orf14.9)</i>	33
1.5.1.2 <i>Présence d'autres gènes dans le cluster : bordures du cluster</i>	36
1.5.2 <i>Organisation du cluster eps/cps des autres bactéries lactiques</i>	38
1.5.2.1 <i>Organisation opéronique et modulaire</i>	38
1.5.2.2 <i>Organisation différente des clusters des souches de S. pneumoniae type 3 et type 37</i>	42
1.5.3 <i>Organisation modulaire des clusters cps des Protéobactéries</i>	44
1.6 DEUX NIVEAUX DE REGULATION DE LA PRODUCTION D'EPS CHEZ LES BACTERIES LACTIQUES	46
1.6.1 <i>Régulation codée par le cluster eps/cps</i>	46
1.6.2 <i>Régulation codée à l'extérieur du cluster : la protéine RegM</i>	48
1.7 INTERVENTION DE GENES SITUES A L'EXTERIEUR DU CLUSTER DANS LA PRODUCTION DES EPS/CPS	50

1.7.1 Intervention de gènes du métabolisme cellulaire dans la biosynthèse de la capsule chez <i>S. pneumoniae</i> de type 3	50
1.7.1.1 Importance du gène <i>pgm</i>	50
1.7.1.2 Importance du gène <i>galU</i>	51
1.7.2 Intervention du gène <i>tts</i> dans la biosynthèse des CPS chez <i>S. pneumoniae</i> de type 37	51
2 POLYMORPHISME INTRASPECIFIQUE DE LA BACTERIE LACTIQUE <i>STREPTOCOCCUS THERMOPHILUS</i>.....	55
2.1 LE GROUPE POLYPHYLETIQUE DES BACTERIES LACTIQUES	55
2.2 <i>S. THERMOPHILUS</i> , ESPECE NON PATHOGENE DU GENRE <i>STREPTOCOCCUS</i>	56
2.3 POLYMORPHISME INTRASPECIFIQUE ET INTERVENTION D'EVENEMENTS DE TRANSFERTS HORIZONTAUX DANS L'EVOLUTION DE <i>S. THERMOPHILUS</i>	58
2.4 MECANISMES IMPLIQUES DANS LE TRANSFERT HORIZONTAL CHEZ <i>S. THERMOPHILUS</i>	61
2.5 IMPLICATION DU TRANSFERT HORIZONTAL DANS LA CONSTITUTION DES CLUSTERS EPS CHEZ <i>S. THERMOPHILUS</i>	63
<u>RESULTATS</u>	65
1 SPECIFICITES STRUCTURALES DU CLUSTER EPS DE LA SOUCHE IP6756	66
1.1 FRONTIERES DU CLUSTER EPS DE LA SOUCHE IP6756	66
1.2 ORGANISATION MODULAIRE PARTICULIERE DE LA PREMIERE REGION DU CLUSTER EPS DE LA SOUCHE IP6756	68
1.3 LA SECONDE REGION DU CLUSTER EPS DE LA SOUCHE IP6756 CONTIENT <i>ORF14.9</i> AINSI QUE D'AUTRES ORF SITUEES EN 5' D' <i>ORF14.9</i>	72
1.4 COMPARAISON DE L'ENSEMBLE DU CLUSTER EPS DE LA SOUCHE IP6756 AVEC LES AUTRES CLUSTERS EPS DE <i>S. THERMOPHILUS</i>	74
2 ETUDE DU LIEN ENTRE LE CARACTERE PRODUCTEUR D'EPS DE <i>S. THERMOPHILUS</i> IP6756 ET LE CLUSTER EPS IDENTIFIE	79
2.1 DETERMINATION DU CARACTERE PRODUCTEUR D'EPS DE LA SOUCHE IP6756	79
2.2 ETUDE DE L'ACTIVITE TRANSCRIPTIONNELLE DES GENES DE LA PREMIERE REGION DU CLUSTER EPS DE <i>S. THERMOPHILUS</i> IP6756	80
2.2.1 Analyse in silico : étude de la région promotrice et identification d'un terminateur de transcription potentiel	80
2.2.2 Les quatre premières ORF du cluster <i>eps</i> seraient organisées en opéron	87
2.3 ETUDE DE LA FONCTIONNALITE DU CLUSTER EPS DE LA SOUCHE IP6756	91
2.3.1 Délétion et interruption polaire du cluster <i>eps</i> de la souche IP6756	91
2.3.2 Recherche du ou des gènes codant la ou les glycosyltransférases nécessaires à la biosynthèse des EPS chez <i>S. thermophilus</i> IP6756	96
2.3.2.1 Recherche de séquences homologues aux gènes codant les glycosyltransférases de <i>S. thermophilus</i> CNRZ368, IP6757 et NST2280	96
2.3.2.2 Recherche de la présence du gène <i>tts</i> au sein du génome de la souche IP6756	98

3 FONCTIONNALITE D'ORF14.9, LA SEULE ORF POTENTIELLEMENT FONCTIONNELLE DE LA SECONDE REGION DU CLUSTER *EPS* DE *S. THERMOPHILUS* IP6756..... 102

3.1	ANALYSE <i>IN SILICO</i> D'ORF14.9	102
3.2	ACTIVITE TRANSCRIPTIONNELLE D'ORF14.9	109
3.3	ETUDE DE LA FONCTION D'ORF14.9 CHEZ <i>S. THERMOPHILUS</i> A054 ET NST2280	109
3.3.1	<i>Etude de l'implication d'orf14.9 dans la biosynthèse des EPS</i>	111
3.3.2	<i>Implication d'orf14.9 dans la croissance cellulaire de S. thermophilus</i>	111

DISCUSSION-PERSPECTIVES 117

1 DES GENES EXTERIEURS AU CLUSTER *EPS* DE LA SOUCHE IP6756 SONT NECESSAIRES A LA BIOSYNTHESE DES EPS DE CETTE SOUCHE..... 120

1.1	LOCALISATION ET FRONTIERES DU CLUSTER <i>EPS</i> DE <i>S. THERMOPHILUS</i> IP6756	120
1.2	FONCTIONNALITE DE LA REGION OPERONIQUE DU CLUSTER <i>EPS</i> DE <i>S. THERMOPHILUS</i> IP6756	121
1.2.1	<i>Activité transcriptionnelle et traductionnelle</i>	121
1.2.2	<i>Régulation de la transcription de l'opéron du cluster eps de S. thermophilus</i> <i>IP6756</i>	123
1.3	QUELS SONT LES GENES IMPLIQUES DANS LA BIOSYNTHESE DES EPS CHEZ <i>S. THERMOPHILUS</i> IP6756 ?	126

2 FONCTIONNALITE D'ORF14.9, APPARTENANT A LA SECONDE REGION DU CLUSTER *EPS* DE LA SOUCHE IP6756 130

2.1	CETTE ORF EST TRES CONSERVEE CHEZ <i>S. THERMOPHILUS</i>	130
2.2	L'INACTIVATION D'ORF14.9 NE S'ACCOMPAGNE PAS D'UNE MODIFICATION QUANTITATIVE DE LA PRODUCTION D'EPS.....	131
2.3	LE CARACTERE FONCTIONNEL D'ORF14.9 EST NECESSAIRE A UNE CROISSANCE CELLULAIRE OPTIMALE.....	134

3 LE CLUSTER *EPS* DE *S. THERMOPHILUS* RESULTERAIT D'UN REGROUPEMENT DE GENES..... 136

3.1	L'ORGANISATION EN CLUSTER S'ACCOMPAGNE D'UNE POSSIBILITE DE TRANSFERT DE CELUI-CI A UNE SOUCHE RECEVEUSE PRESENTE DANS LA MEME NICHE ECOLOGIQUE	136
3.2	LA CONSTITUTION DU CLUSTER <i>EPS</i> DE LA SOUCHE IP6756 IMPLIQUE L'INTERVENTION DE TRANSFERTS HORIZONTAUX	137
3.2.1	<i>Un remplacement de séquence serait à l'origine de l'organisation du cluster eps</i> <i>de la souche IP6756</i>	137
3.2.2	<i>Origine de la région centrale du cluster eps chez la souche IP6756 et des gènes</i> <i>extérieurs au cluster eps nécessaires à la biosynthèse des EPS</i>	138

3.3	MODE D'ACQUISITION DE L'ADN EXOGENE : CAS DES SEQUENCES APPARTENANT AUX CLUSTERS <i>EPS</i>	139
3.4	HYPOTHESES CONCERNANT LE ROLE D' <i>ORF14.9</i> DANS L'ACQUISITION DE NOUVELLES SEQUENCES.....	141
3.5	QUELLES SONT LA OU LES PRESSION(S) DE SELECTION S'EXERÇANT SUR LE CLUSTER <i>EPS</i> CHEZ <i>S. THERMOPHILUS</i> ?	142
<u>CONCLUSION</u>		147
<u>MATERIEL ET METHODES</u>		150
1	SOUCHES BACTERIENNES ET CONDITIONS DE CULTURE....	150
2	PLASMIDE	152
3	PREPARATION D'ACIDES NUCLEIQUES.....	152
3.1	EXTRACTION D'ADN PLASMIDIQUE ET GENOMIQUE	152
3.2	EXTRACTION DES ARN TOTAUX	154
3.3	AMPLIFICATION D'ADN PAR POLYMERISATION EN CHAINE	154
3.3.1	<i>Amplification par PCR</i>	154
3.3.2	<i>Amplification d'ADN par LongRange PCR</i>	155
3.3.3	<i>Purification des fragments d'ADN</i>	157
3.4	DOSAGE DES ACIDES NUCLEIQUES.....	157
4	ANALYSE DE RESTRICTION ET ELECTROPHORESE	158
5	HYBRIDATION ADN-ADN (SOUTHERN BLOT).....	158
5.1	MARQUAGE DE SONDE.....	158
5.2	TRANSFERT D'ADN SUR MEMBRANE	159
5.3	HYBRIDATION ET DETECTION	159
6	TRANSCRIPTION INVERSE DES ARN EN ADNc	160
7	SEQUENÇAGE D'ADN ET ANALYSE DES SEQUENCES	161
7.1	REACTIONS DE SEQUENÇAGE.....	161
7.2	ANALYSE DES SEQUENCES.....	162
8	CLONAGE DE FRAGMENTS D'ADN.....	163
8.1	LIGATION	163
8.2	ELECTROPORATION	163
9	DELETION DU CLUSTER <i>EPS</i> CHEZ <i>S. THERMOPHILUS</i> IP6756 ET NST2280.....	165

10 INTERRUPTION D'ORF14.9 CHEZ <i>S. THERMOPHILUS</i> NST2280 ET A054.....	166
11 DETERMINATION DU CARACTERE PRODUCTEUR D'EPS.....	166
11.1 EXTRACTION DES EPS.....	166
11.2 DOSAGE DES EPS EXTRAITS.....	166
11.3 COMPOSITION DES EPS	167
12 DETERMINATION DES PARAMETRES DE CROISSANCE BACTERIENNE EN MILIEU LIQUIDE	167
 <u>REFERENCES BIBLIOGRAPHIQUES</u>	 168
 ANNEXE 1	 180
ANNEXE 2	202

PREAMBULE

Le cluster *eps/cps*, qui regroupe l'essentiel des gènes impliqués directement dans la biosynthèse des exopolysaccharides (EPS), a été identifié chez de nombreuses souches et espèces de bactéries lactiques, en particulier chez *Streptococcus thermophilus*. Les gènes qui le constituent présentent un polymorphisme très élevé en différant par leur nombre, leur séquence et la fonction des produits qu'ils codent. Le cluster *eps* constitue ainsi une des régions variables du chromosome de *S. thermophilus*. Cette variabilité est liée à des transferts horizontaux d'information génétique.

Une illustration de cette variabilité est fournie par la comparaison des clusters *eps* très différents des souches IP6757 et NST2280 de *S. thermophilus*. Cependant, l'examen de l'arbre phylogénétique de différentes souches de *S. thermophilus* montre que ces deux souches sont très proches phylogénétiquement (Figure 1). Ces résultats indiquent une évolution rapide du cluster *eps* de la souche IP6757 qui pourrait résulter d'un remplacement de séquence (Pluvinet *et al.*, 2004). Des événements de transferts horizontaux sont très vraisemblablement impliqués dans l'acquisition de ces nouvelles séquences. Des signatures potentielles de tels événements ont été détectées au sein des clusters *eps* de *S. thermophilus*.

La seule niche écologique connue pour *S. thermophilus* est constituée par les produits laitiers fermentés, dans lesquels l'espèce est en présence d'une flore lactique complexe incluant *Lactococcus lactis*. Des travaux antérieurs menés au laboratoire ont montré le transfert horizontal de séquences d'insertion (IS) entre ces deux espèces bactériennes, probablement lors des fermentations alimentaires (Guedon *et al.*, 1995) (Bourgoin *et al.*, 1996). Ces résultats amènent à considérer l'ensemble des bactéries de la flore lactique présentes dans les produits laitiers fermentés comme constituant un pool génétique ou un

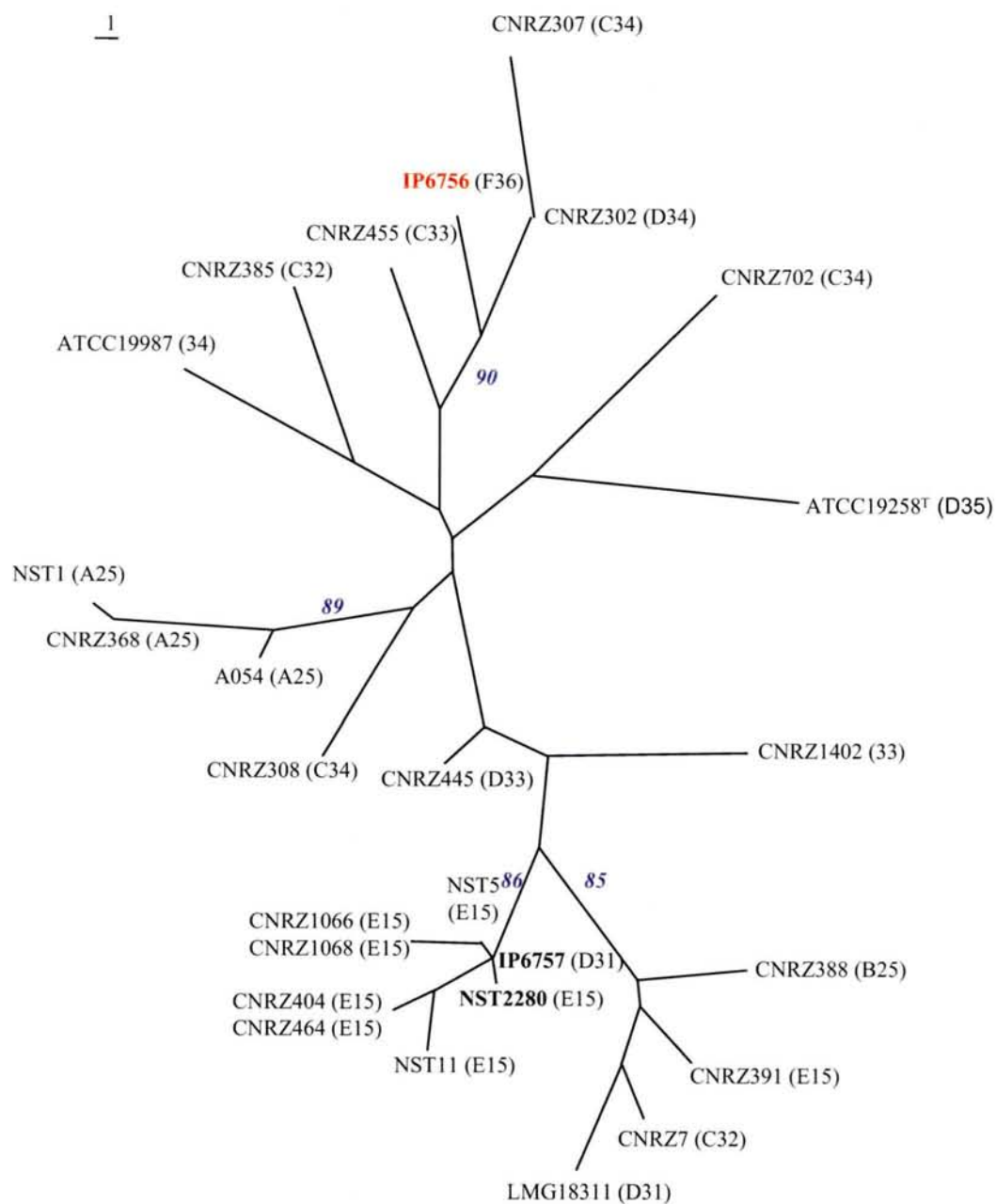


Figure 1 : Arbre phylogénétique de 26 souches de *S. thermophilus* obtenu par Pluvinet *et al.* (2004). La barre d'échelle correspond à une différence dans les profils d'hybridations. Seules les valeurs de bootstrap $\geq 50\%$ sont indiquées. Le chiffre indiqué pour toutes les souches correspond au type de cluster *eps*.

métagénome. Un ensemble complexe d'interactions peut être observé au sein de cette flore, en particulier des relations de protocoopération entre les espèces.

Au sein du genre *Streptococcus*, *S. thermophilus* est la seule espèce considérée comme non pathogène. L'analyse de la séquence du génome des souches LMG18311 et CNRZ1066 révèle l'absence de gènes de virulence (Bolotin *et al.*, 2004). De nombreux clusters *cps* présents chez *Streptococcus pneumoniae* ont été étudiés. Ils présentent, comme chez *S. thermophilus*, une très grande variabilité liée à des transferts horizontaux d'information génétique (Lopez & Garcia, 2004). Plus généralement, les loci *cps* de nombreuses espèces pathogènes présentent cette très grande variabilité liée au transfert horizontal. Chez les bactéries pathogènes, les EPS sont organisés sous forme de capsule qui protège, en particulier, les cellules bactériennes de la phagocytose par le système immunitaire de l'hôte. La variabilité des EPS produits par les bactéries pathogènes serait corrélée à la variabilité des réactions de défense immunitaire lors d'une infection. La variabilité du système immunitaire pourrait ainsi être considérée comme la pression de sélection s'exerçant au niveau des clusters *cps* des bactéries pathogènes. Dans le cas de l'espèce non pathogène *S. thermophilus*, l'identification d'une pression de sélection s'exerçant sur le cluster *eps/cps*, qui se traduit également par une très grande variabilité, est plus incertaine. Ceci amène à s'interroger sur les interactions de cette bactérie avec l'environnement et sur le ou les rôle(s) des EPS chez cette bactérie.

La comparaison de la séquence d'une vingtaine de clusters *eps* de *S. thermophilus* suggère qu'ils présentent une organisation modulaire de la première région qui est conservée entre les souches (Jolly & Stingle, 2001). Deux modules peuvent ainsi y être définies : un module fortement conservé qui contient des gènes impliqués dans la régulation transcriptionnelle (*epsA*) et post-transcriptionnelle (*epsB*, *epsC* et *epsD*) de la production d'EPS et en aval un module très variable qui contient les gènes codant les

glycosyltransférases (enzymes catalysant la formation des unités répétées constituant les molécules d'EPS) et les gènes impliqués dans la polymérisation et l'export des EPS (Jolly & Stinglele, 2001).

Ce schéma général est observé pour seize des dix-sept clusters *eps* étudiés par hybridation, à l'exception de celui de la souche IP6756 (Charron-Bourgoin *et al.*, 2001). Le cluster *eps* de la souche IP6756, dont la position phylogénétique est précisée figure 1, présenterait ainsi une constitution particulière. En effet, aucun signal d'hybridation spécifique à *epsC* et *epsD* de la région conservée n'est obtenu. Ces deux ORF sont donc probablement absents du génome de la souche IP6756. Par contre, le cluster *eps* de cette souche contient *orf14.9*, une ORF distale de fonction inconnue qui est commune à tous les clusters *eps* de *S. thermophilus*. Le cluster *eps* de la souche IP6756 présente une autre particularité : alors que cette souche est productrice d'EPS, aucune ORF codant une glycosyltransférase potentielle n'est identifiée au sein du cluster. L'étude structurale et fonctionnelle du cluster *eps* de la souche IP6756 amène à s'interroger sur l'autonomie fonctionnelle de ce cluster et sur le rôle d'*orf14.9* dans la biosynthèse des EPS. Par ailleurs, ce cluster résulterait d'un remplacement de séquence qui n'interrompt pas la transcription du cluster et qui serait à l'origine de la constitution d'une ORF mosaïque potentiellement fonctionnelle.

Cette introduction traitera donc des exopolysaccharides en se focalisant sur le déterminisme génétique de la production de telles molécules chez les bactéries. Ainsi, elle abordera l'organisation et la régulation de l'activité des clusters *eps*, plus particulièrement chez les bactéries lactiques. Puis, elle traitera des bactéries lactiques et plus précisément de la bactérie *S. thermophilus*, dont elle présentera les mécanismes d'acquisition de nouvelles séquences.

INTRODUCTION

1 LES EXOPOLYSACCHARIDES (EPS)

1.1 Diversité des polysaccharides de l'enveloppe cellulaire chez les bactéries

1.1.1 Le peptidoglycane : constituant principal de la paroi bactérienne

La paroi des bactéries est principalement constituée de peptidoglycane (Figure 2). Cet hétéropolymère complexe résulte de la polymérisation d'un motif de base composé d'acide N-acétylmuramique et de N-acétylglucosamine pour constituer les chaînes principales. Sur les molécules d'acide N-acétylmuramique sont attachés des pentapeptides qui constituent les chaînes latérales. Ces chaînes latérales peuvent subir une réaction de transpeptidation, formant une liaison interpeptidique qui relie deux chaînes principales du peptidoglycane. Le peptidoglycane forme ainsi un réseau tridimensionnel rigide, qui confère à la cellule sa forme et sa capacité à résister à la pression osmotique (Delcour *et al.*, 1999). Chez les bactéries à coloration de Gram positive, plusieurs couches concentriques de peptidoglycane sont reliées de façon covalente par des liaisons additionnelles, renforçant l'ensemble de la structure (Delcour *et al.*, 1999).

1.1.2 Polysaccharides spécifiques des bactéries à coloration de Gram positive

Des molécules d'acide téichoïque, composées en général de plusieurs molécules de glycérol et de ribitol, sont présentes au niveau de la paroi des bactéries à coloration de Gram positive (Figure 2). Ces molécules peuvent représenter jusqu'à la moitié de la masse de la paroi et sont synthétisées au niveau de la face externe de la membrane cytoplasmique. Une très grande diversité de leur structure et de leur abondance est observée en fonction de nombreux paramètres tels que l'espèce, la souche, la phase de croissance, le pH du milieu ou

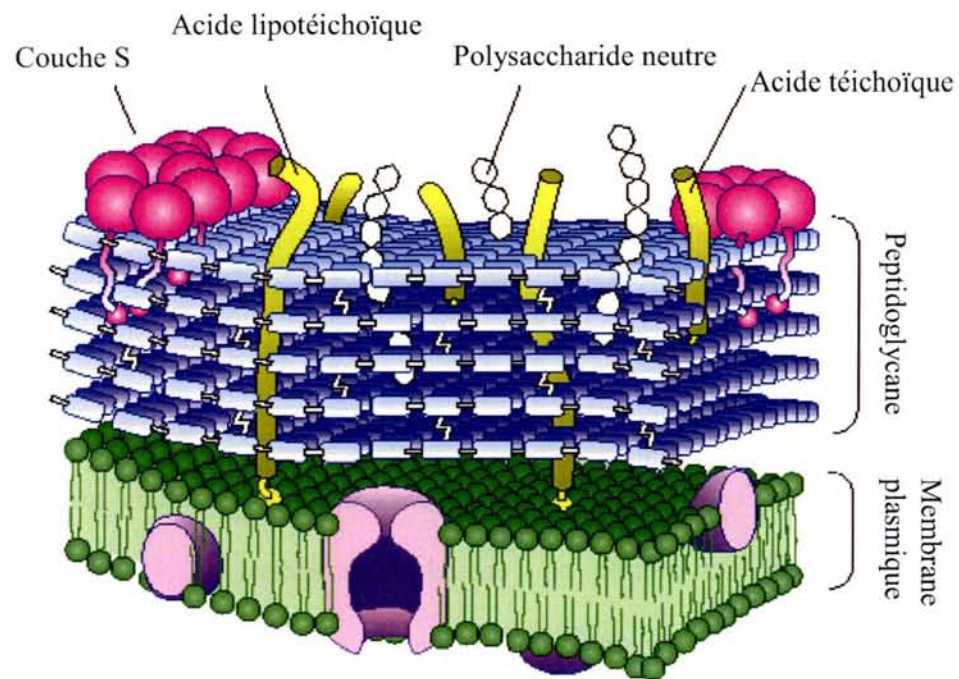


Figure 2 : Représentation de l'enveloppe d'une bactérie à coloration de Gram positive d'après Delcour *et al.* (1999).

la source de carbone (Delcour *et al.*, 1999). Il est possible de distinguer, d'une part les acides lipotéichoïques et les lipoglycanes qui restent majoritairement attachés à la membrane cytoplasmique et d'autre part, les acides téichoïques (*sensu stricto*) et téichuroniques qui sont liés de façon covalente au peptidoglycane.

Les molécules d'acides téichoïques jouent un rôle important sur le plan fonctionnel. Ainsi, leur nature anionique leur permettrait de capter des cations et d'établir un gradient de pH au niveau de la paroi. De plus, ils peuvent constituer une réserve de phosphate et influencer l'activité d'autolysine. Ils présentent des propriétés immunogéniques importantes et peuvent être considérés comme les équivalents des lipopolysaccharides présents chez les bactéries à coloration de Gram négative (Delcour *et al.*, 1999).

Des molécules de polysaccharides neutres de structure très variée peuvent également être associées de façon covalente ou non au peptidoglycane (Figure 2) (Delcour *et al.*, 1999). De même, l'ancrage au peptidoglycane des protéines constituant la couche S, qui peuvent elles-mêmes être glycosylées, implique une classe de polysaccharides attachés de façon covalente au peptidoglycane (Schaffer & Messner, 2005). Cet ancrage se ferait de façon non-covalente, grâce à une interaction de type lectine entre ces polysaccharides et les protéines constituant la couche S.

1.1.3 Les lipopolysaccharides sont spécifiques des bactéries à coloration de Gram négative

Les lipopolysaccharides (LPS) sont retrouvés chez les bactéries à coloration de Gram négative. Ils présentent une architecture organisée en deux ou trois régions : une ancre membranaire hydrophobe (le lipide A), un polysaccharide, qui peut être prolongé chez certaines bactéries par un oligosaccharide constitué d'unités répétées, encore appelé antigène spécifique O (Figure 3). Chez les bactéries pathogènes, ces chaînes spécifiques O sont en

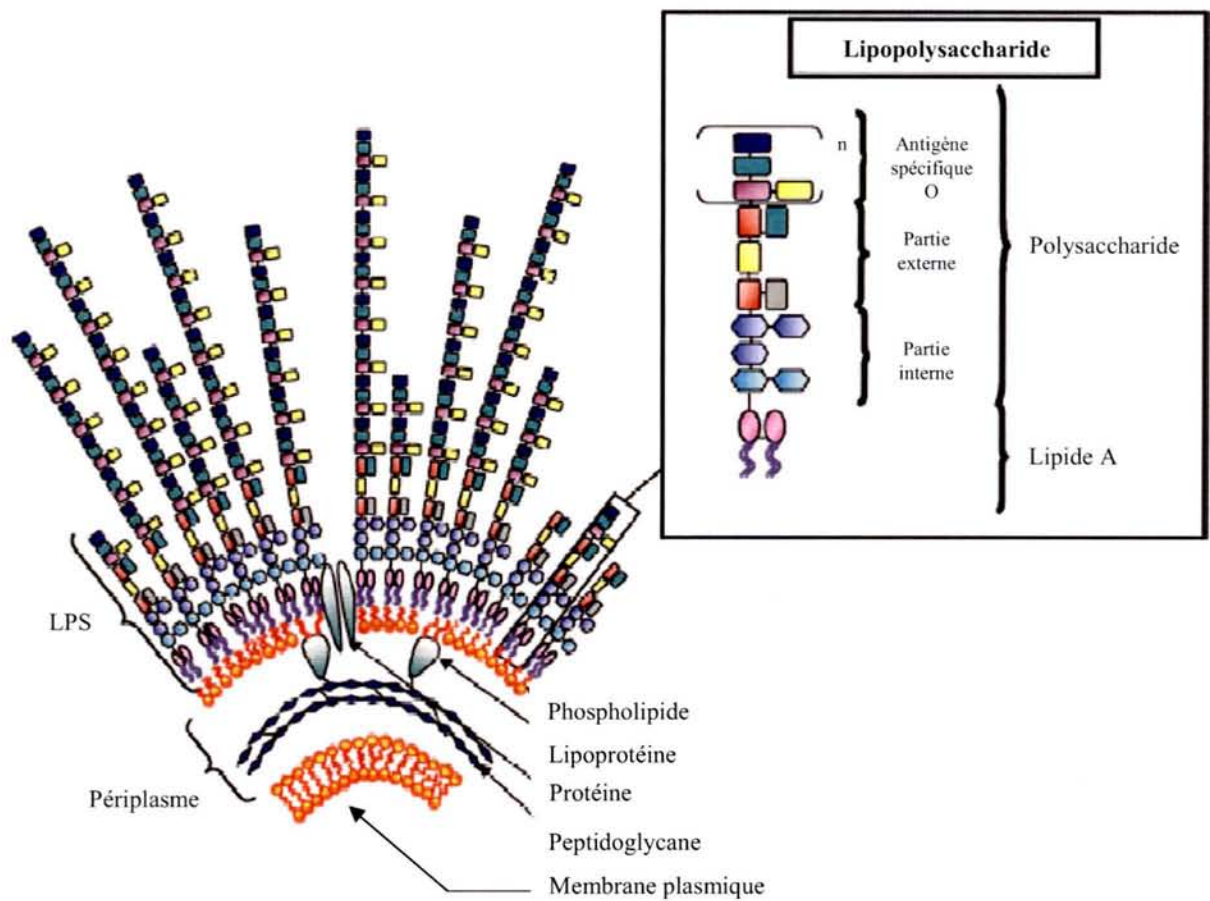


Figure 3 : Représentation schématique de l'enveloppe cellulaire des Entérobactéries (gauche) et de la structure d'une molécule de lipopolysaccharide (droite) d'après Caroff *et al.* (2003).

contact direct avec l'hôte durant l'infection et peuvent protéger la bactérie de l'action d'antibiotiques ou du complément (Caroff & Karibian, 2003).

1.1.4 Les exopolysaccharides forment des capsules et/ ou des amas visqueux (« slime »)

Le terme « exopolysaccharide » (EPS) est un terme général qui se réfère à deux types de polysaccharides sécrétés. Ainsi, les EPS peuvent être liés de façon covalente à la paroi et former une capsule. Ils sont alors désignés sous le terme de CPS (Capsular PolySaccharide). Ils peuvent également être sécrétés dans l'environnement cellulaire, constituant un amas visqueux appelé *slime* (Degeest *et al.*, 2001).

Les EPS sont très répandus dans le monde bactérien. Certaines bactéries sont capables de produire les deux types d'EPS, alors que d'autres n'en produisent qu'un seul. Ainsi, *S. thermophilus* produit généralement des EPS sous forme de *slime* (Broadbent *et al.*, 2003). Mais certaines souches, telle la souche S1, sont capables de synthétiser une capsule, dont la taille varie en fonction des conditions de culture (Hassan *et al.*, 2001). Par exemple, quelle que soit la souche productrice de capsule étudiée, le diamètre de la capsule produite est maximal lorsque les cultures sont effectuées en milieu lait. L'étude de la capacité à produire une capsule chez cinq souches de *S. thermophilus* a permis de mettre en évidence que les oses nécessaires à la production d'une capsule sont souches-spécifiques. Ainsi les souches S1 et S3855 sont les seules parmi les souches étudiées à pouvoir synthétiser une capsule en milieu Elliker dont la seule source glucidique est le galactose (Hassan *et al.*, 2001).

La synthèse des EPS est généralement réalisée par des protéines codées par des gènes organisés en *cluster*, appelés cluster *cap*, *cps* ou *eps* selon les espèces. Chez les bactéries lactiques, ces clusters présentent très souvent une même organisation fonctionnelle mais une très forte variabilité génétique conduit à des polysaccharides de nature très différente.

1.2 Importance pour la bactérie des interactions entre les EPS et l'environnement

Les EPS ne semblent pas correspondre à une réserve d'énergie externalisée directement utilisable puisque la plupart des bactéries productrices d'EPS ne sont pas capables de cataboliser les EPS qu'elles produisent. Cependant, *Streptococcus mutans* et *Streptococcus sobrinus* peuvent dégrader le dextrane et *S. mutans* peut utiliser les oligosaccharides résultants (De Vuyst *et al.*, 2001). Les EPS exerceraient ainsi une ou plusieurs fonctions plus complexes, d'autant que l'énergie cellulaire affectée à la production d'EPS peut atteindre 70% de l'énergie cellulaire disponible chez certaines bactéries. Ainsi, de nombreuses espèces croissent plus vite en laboratoire après la perte de leur caractère producteur d'EPS (Weiner *et al.*, 1995). Ceci suggère que des facteurs environnementaux interviennent dans la sélection et le maintien du phénotype producteur d'EPS.

Les EPS sont en contact avec l'environnement immédiat de la cellule. Ces molécules sont impliquées dans de multiples phénomènes d'interaction avec l'environnement telles les interactions à bénéfice réciproque avec d'autres organismes présents dans la même niche écologique, les phénomènes d'adhésion cellulaire pouvant conduire à l'établissement de biofilms, ou la protection de la cellule face à la variation de paramètres de l'environnement.

1.2.1 Importance des EPS de *Rhizobium* dans l'établissement d'une symbiose fonctionnelle

L'implication des EPS dans l'établissement d'interactions à bénéfice réciproque a été décrite chez le genre *Rhizobium*. Ainsi, *Rhizobium meliloti* 2011 produit deux types de molécules EPS I : un de haut poids moléculaire et l'autre de faible poids moléculaire (Becker *et al.*, 1995). Il a été montré que la molécule qui promeut l'invasion des nodules est le tétramère le plus chargé des unités répétées octasaccharidiques des molécules d'EPS I de faible poids moléculaire (Battisti *et al.*, 1992). Des observations comparables ont été faites chez la souche Rm1021 de cette espèce (Gonzalez *et al.*, 1996a). Cette souche peut produire deux types d'EPS différents selon les conditions de croissance: EPS I (molécule de

succinoglycane) en conditions normales et EPS II (galactoglucane) en conditions limitantes en phosphate (Lloret *et al.*, 1998). Les mutants déficients pour la production d'EPS I n'induisent que des nodules vides et non fonctionnels. La formation de nodules fonctionnels peut être restaurée par une quantité infime -de l'ordre du picogramme- d'EPS II (Gonzalez *et al.*, 1996b). Les molécules d'EPS pourraient ainsi agir comme signal pour l'invasion des nodules.

1.2.2 Rôle primordial des EPS lors de l'établissement d'un biofilm

L'implication des EPS dans les phénomènes d'adhésion cellulaire et dans la constitution de la matrice des biofilms ont été décrites chez de nombreuses espèces bactériennes. L'établissement d'un biofilm peut être divisé en deux phases. La première fait intervenir un mécanisme complexe impliquant de nombreux facteurs physico-chimiques, des protéines ainsi que les molécules d'EPS et conduit à l'attachement primaire de la bactérie à la surface du polymère. Dans la seconde phase, les bactéries attachées prolifèrent et forment un biofilm multicouche constitué de cellules englobées par un manchon polysaccharidique (Mack, 1999) ; (Mack *et al.*, 2000).

La formation de biofilm est un domaine d'étude important du fait des conséquences que peut avoir leur présence sur la santé humaine. Ainsi, la formation de biofilm, par la présence de molécules d'EPS PIA à la surface cellulaire, est essentielle à la pathogénicité de *Staphylococcus epidermidis*. L'organisation des bactéries en biofilm s'accompagne d'une réduction de leur exposition aux traitements antibiotiques et aux attaques du système de défense de l'hôte (Vuong *et al.*, 2004).

De même, la virulence de *S. mutans*, agent de la carie dentaire, est liée à sa capacité à produire un polymère de glucane insoluble à partir du saccharose alimentaire. Ce polymère

est un des constituants principaux du biofilm de la plaque dentaire. Des molécules de fructane peuvent également faciliter l'adhésion bactérienne (Browngardt *et al.*, 2004).

1.2.3 Les EPS : barrière de protection face aux agressions du milieu extérieur

Les agressions du milieu extérieur peuvent être de nature physico-chimique, comme la dessiccation, la présence de composés toxiques et les variations de température. Ainsi, chez *Pseudomonas syringae*, une souche mutante déficiente pour la production d'EPS (ici, l'alginate) présente un taux de mortalité supérieur à la souche sauvage lorsqu'elle est exposée à une température de 43°C (Figure 4), ainsi qu'une sensibilité supérieure au paraquat et au peroxyde d'hydrogène. De plus, la production d'alginate est activée lors d'un stress osmotique, d'un stress oxydant, en présence de cuivre ainsi que lors d'une augmentation de température (Keith & Bender, 1999).

Le spectre de protection contre ces différents facteurs physico-chimiques est plus ou moins étendu selon les bactéries. Ainsi, l'étude de la souche MG1614 de *Lactococcus lactis* suggère que les EPS ne sont pas capables de protéger la bactérie face à une hausse de température, à la présence de pénicilline et de vancomycine. Par contre, les EPS de cette bactérie confèrent une protection face aux ions métalliques, à la nisine (une bactériocine) et au lysozyme. Pour certains de ces facteurs, la présence d'EPS associés à la cellule est nécessaire à la protection alors que pour d'autres, l'ajout d'EPS en suspension dans le milieu de culture protège également la bactérie (Looijesteijn *et al.*, 2001).

La capsule polysaccharidique peut également protéger la cellule bactérienne de la prédation par des protozoaires et de la phagocytose (Roberts, 1996). Ainsi, les souches encapsulées de *Streptococcus suis*, un des agents de la méningite, sont résistantes à la phagocytose. Lors de l'infection par une souche capsulée, une réduction de l'activité de la voie métabolique du processus de phagocytose impliquant la protéine Akt et la protéine kinase C est observée au niveau des macrophages. De même, lorsque des EPS purifiés de

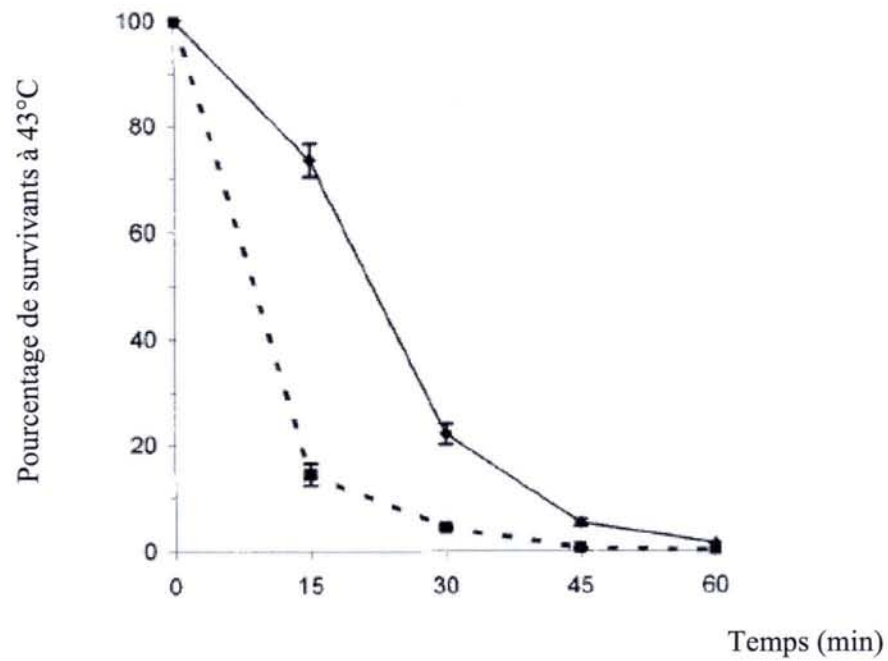


Figure 4 : Suivi du pourcentage de survie de *Pseudomonas syringae* à une température de 43°C pour une souche sauvage (trait plein) et pour une souche déficiente dans la production d'EPS (trait pointillé), d'après Keith *et al.* (1999).

Les écarts-types de chaque mesure sont indiqués par une barre.

capsule sont ajoutés aux macrophages, aucune réaction de phagocytose n'est observée (Segura *et al.*, 2004).

La protection conférée par la présence d'EPS autour de la cellule peut aussi s'appliquer face aux bactériophages. Ainsi, chez *R. meliloti*, il a été montré que les EPS inhibent l'adsorption des phages (Defives *et al.*, 1996). De même, le plasmide pCI658, identifié chez *Lactococcus lactis* subsp *cremoris* HO2, code la biosynthèse d'un EPS qui diminue l'adsorption des phages $\phi 712$ et $\phi c2$ sur leurs récepteurs cellulaires (Forde *et al.*, 1999). La présence d'EPS associés à la cellule confère également une faible protection à la souche MG1614 de *L. lactis* (Looijesteijn *et al.*, 2001). Cependant, des phages spécifiques de souches capsulées de *Vibrio cholerae* (Albert *et al.*, 1996) ou d'*E. coli* (Clarke *et al.*, 2000) ont été isolés. De même, des travaux menés sur de nombreuses souches de *L. lactis* ne permettent pas d'observer de différence de sensibilité aux phages entre les souches productrices d'EPS et les variants non producteurs (Deveau *et al.*, 2002). Des résultats obtenus chez *S. thermophilus* MR-1C suggèrent que la présence d'EPS n'est pas suffisante pour protéger les cellules contre l'attaque des phages de types Q1, Q3 et Sfi11 (Broadbent *et al.*, 2003).

1.3 Très grande diversité structurale des EPS

Cette propriété est liée au fait que deux molécules d'ose identiques peuvent former onze dissacharides différents, alors que deux acides aminés identiques ne peuvent constituer qu'un seul dipeptide. Une grande diversité des molécules glucidiques composant les EPS est également retrouvée : glucose, mannose, acide glucuronique, rhamnose, galactose, xylose par exemple. Les molécules d'EPS peuvent également adopter des structures d'ordre supérieur. Par exemple, les molécules de xanthane seraient organisées en hélices double-brin antiparallèles (Weiner *et al.*, 1995).

Classiquement, deux catégories d'EPS peuvent être distinguées chez les bactéries lactiques: les homopolysaccharides et les hétéropolysaccharides (De Vuyst & Degeest, 1999).

Les homopolysaccharides sont constitués de la répétition d'un seul type de monosaccharide. Par exemple, les molécules de fructane sont constituées de résidus fructose tel que le levane produit par *Streptococcus salivarius* et *S. mutans*, où ces résidus sont associés par des liaisons β -2,6 (Figure 5a). Deux autres grandes catégories d'homopolysaccharides peuvent être observées : les glucanes (polymère de résidus glucose, figure 5b) et les polygalactanes (polymère de résidus galactose). La masse moléculaire des homopolysaccharides est de l'ordre de $7 \cdot 10^6$ Da (Ruas-Madiedo *et al.*, 2002).

Les hétéropolysaccharides sont composés d'unités répétées qui contiennent généralement du D-glucose, D-galactose, L-rhamnose et dans quelques cas du N-acétylglucosamine, N-acétylgalactosamine ou de l'acide glucuronique (Figure 6). Par exemple, l'unité répétée des EPS de *S. thermophilus* Sfi39 est tétrasaccharidique et est composée de D-glucose et D-galactose, en ratio équimolaire (Lemoine *et al.*, 1997). Parfois, les hétéropolysaccharides contiennent également des molécules non-glucidiques de phosphate, acétyl ou glycérol (Ruas-Madiedo *et al.*, 2002). La masse moléculaire des hétéropolysaccharides est comprise entre $4 \cdot 10^4$ et $6 \cdot 10^6$ Da (De Vuyst & Degeest, 1999).

Les conditions de culture peuvent influencer sur la composition des EPS. Ainsi, le pH ou la source de carbone peuvent modifier les ratios relatifs de monosaccharides au sein des molécules d'EPS (Petry *et al.*, 2000). Par exemple, les EPS produits par *Lb. delbrueckii* subsp. *bulgaricus* NCFB2772 en présence de glucose sont constitués de galactose, glucose et rhamnose avec un ratio de 6,8:1:0,7 alors qu'en présence de fructose, ils sont composés uniquement de galactose et glucose avec un ratio 2,5:1 (Grobbe *et al.*, 1995) ; (Grobbe *et al.*, 1996).

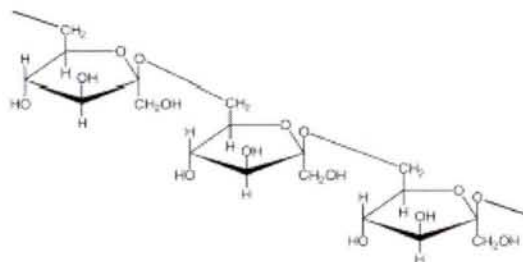
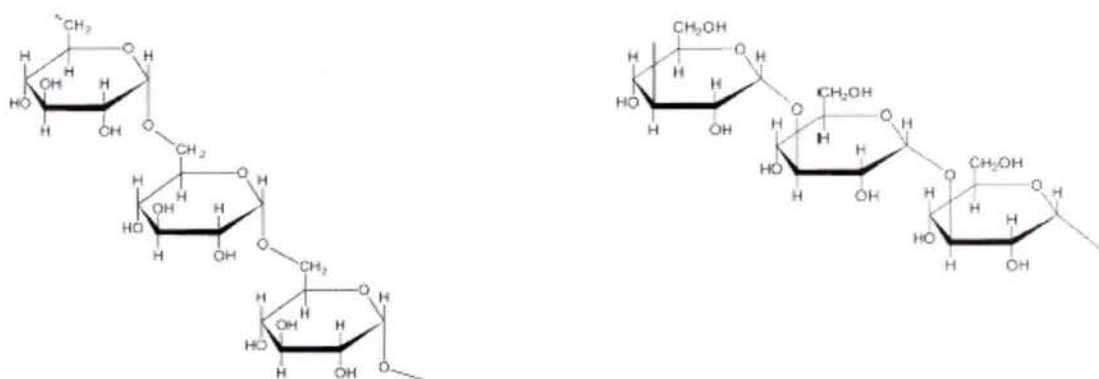
a**b**

Figure 5 : Structure d'homopolysaccharides (d'après Monsan *et al.*, 2001)

a: Fructane : levane

b: Glucane : dextrane à gauche et mutane à droite

Souches de bactéries lactiques

Unités répétées

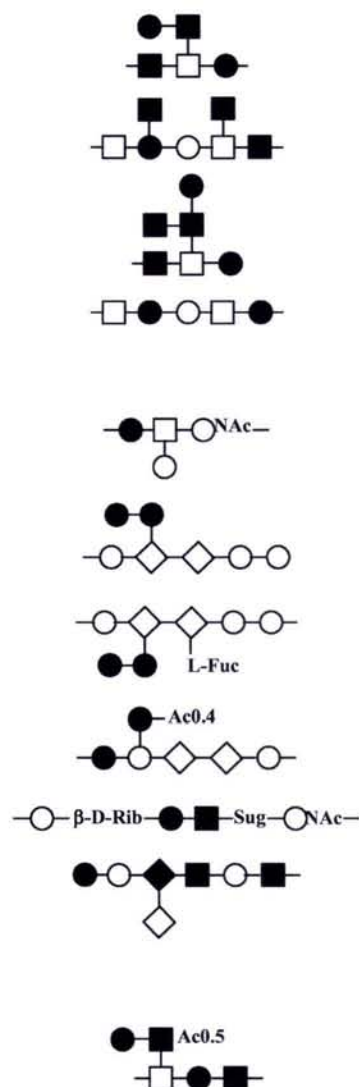
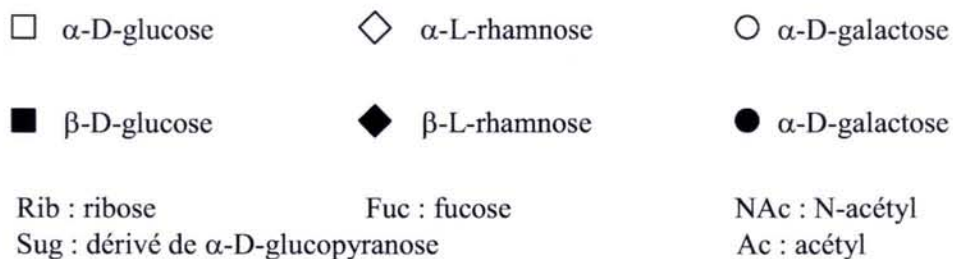
*Lactobacillus**Lb. delbueckii*
subsp. *bulgaricus* 291*Lb. helveticus* Lb161*Lb. helveticus* K16*Lb. rhamnosus* C83*Streptococcus**S. thermophilus* Sfi20*S. thermophilus* Rs*S. thermophilus* Sts*S. thermophilus* MR-1C*S. thermophilus* S3*S. thermophilus* 8Sb*S. thermophilus* EU20*Lactococcus**L. lactis* subsp. *cremoris* NIZO B891

Figure 6 : Structure des hétéropolysaccharides produits par différentes souches de bactéries lactiques (compilation effectuée par Ruas-Madiedo *et al.*, 2002).



1.3.1 Avantage sélectif conféré aux bactéries pathogènes par la variabilité de leurs EPS

La très grande variabilité structurale des EPS observée chez les bactéries est liée au polymorphisme génétique des clusters impliqués. Les phénotypes alternatifs en découlant permettent l'adaptation des bactéries à des conditions environnementales jamais rencontrées préalablement. Cette situation s'applique aux bactéries pathogènes, qui sont confrontées au système immunitaire de leurs hôtes et sont donc soumises à des conditions environnementales variables.

Ainsi, la virulence de *Staphylococcus aureus* est corrélée à la présence d'une capsule polysaccharidique, qui protège de l'opsonisation. La variabilité de la capsule chez cette bactérie est illustrée par l'isolement de onze sérotypes différents (Lin *et al.*, 1994).

La pathogénicité de *Streptococcus pneumoniae*, bactérie impliquée dans des otites, sinusites, pneumonies, méningites, est attribuée à différents facteurs dont la capsule. Celle-ci confère une protection contre la phagocytose et présente une importante variabilité puisque plus de quatre vingt dix types différents de capsule ont été décrits (Garcia *et al.*, 2000). Un phénomène de variation phénotypique entre un type opaque et un type transparent a été observé. Le type opaque présente une virulence beaucoup plus importante que le type transparent (Figure 7). Cette propriété est corrélée à une quantité plus importante de polysaccharide de capsule chez les variants opaques (Kim & Weiser, 1998).

Du fait de leur implication dans le développement de maladies, les EPS des bactéries pathogènes sont particulièrement étudiés. Ainsi, pour limiter l'impact des maladies dues à *S. pneumoniae*, une approche vaccinale a été développée. La génération actuelle de vaccins conjugués protéine-polysaccharide ne permet pas de prendre en compte les différents sérotypes de cette espèce. La stratégie thérapeutique actuelle s'oriente vers le blocage de la

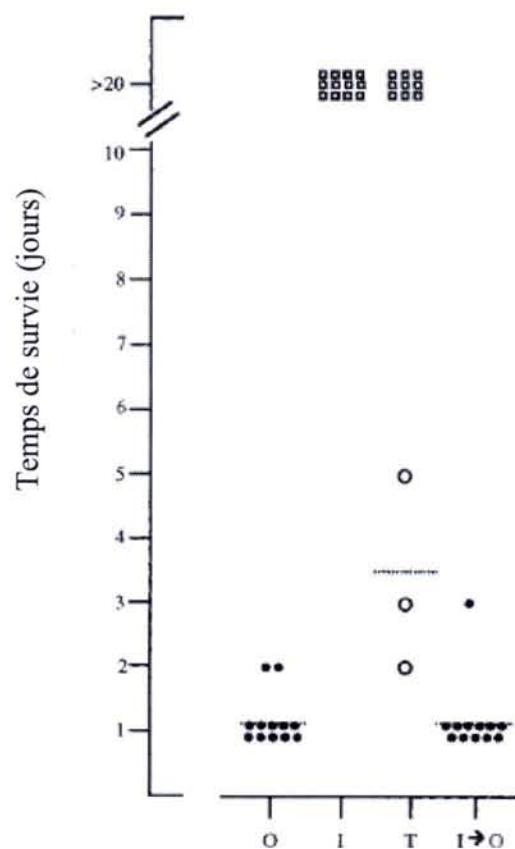


Figure 7 : Relation entre la virulence et la morphologie des colonies chez *S. pneumoniae* de type 6A (d'après Kim et Weiser, 1998).

Des souris adultes sont inoculées intrapéritonéalement avec 10^7 CFU de types opaque (O), transparent (T), intermédiaire (I) et révertant (I→O).

Lorsque les cercles sont noirs, la morphologie des colonies observées *post-mortem* est similaire à celle des colonies de départ. Lorsqu'il sont blancs, la morphologie diffère. Les animaux ayant survécu plus de vingt jours sont représentés par des carrés. Le temps moyen de survie est figuré par une trait pointillé.

production de la capsule (Garcia *et al.*, 2000) et s'accompagne de nombreuses études portant sur la fonctionnalité des clusters *cps* (CF paragraphes 1.4 à 1.7).

L'avantage sélectif apporté aux bactéries lactiques non pathogènes par la variabilité de leurs EPS fait l'objet de différentes hypothèses. Elle pourrait par exemple leur conférer un avantage sélectif face aux attaques de bactériophages. Néanmoins, la variabilité des EPS de ces bactéries présente un intérêt industriel important.

1.3.2 Utilisation industrielle de la variabilité des EPS produits par les bactéries lactiques non pathogènes

La structure de nombreux EPS produits par les bactéries lactiques utilisées dans l'industrie agro-alimentaire, notamment pour la production de yaourt, fromage et autres produits lactés, a été déterminée. Ces EPS présentent des propriétés rhéologiques leur permettant de contribuer à la texture, aux propriétés gustatives et à la stabilité du produit final. C'est particulièrement important lors de la fabrication de yaourts brassés, où la présence d'EPS permet de réduire le phénomène de synérèse (Duboc & Mollet, 2001). D'autres approches technologiques peuvent être utilisées pour contribuer à la texture et à la stabilité du produit, par exemple en ajoutant des molécules stabilisantes telles que la gélatine, la pectine, l'alginate et les carraghénanes (De Vuyst *et al.*, 2001). Mais l'usage de ces molécules n'est pas toujours autorisé par la législation (De Vuyst & Degeest, 1999). L'utilisation de souches produisant *in-situ* des EPS permet de s'affranchir de ces restrictions industrielles.

De nombreuses études ont permis de constater que plusieurs facteurs sont impliqués dans les propriétés rhéologiques des EPS : quantité produite, conditions de fermentation et nature de l'EPS (Laws & Marshall, 2001). Cette constatation pose la question de la relation existant entre la structure des EPS et les propriétés rhéologiques que ces molécules confèrent (De Vuyst & Degeest, 1999). Etant donnée l'importance industrielle des EPS des bactéries

lactiques, de nombreux travaux concernant la fonctionnalité des clusters *eps* de ces bactéries ont été réalisés (CF paragraphes 1.4 à 1.6).

1.4 Biosynthèse des EPS

La biosynthèse des EPS est un mécanisme coûteux en énergie qui implique successivement des enzymes catalysant la production des précurseurs (molécules d'oses nucléotidylées, c'est-à-dire associées à un nucléotide), une glycosyl-1-phosphate transférase (catalysant la liaison de la première molécule osidique sur l'accepteur lipidique), une ou plusieurs glycosyltransférases (catalysant l'ajout séquentiel de molécules osidiques pour former l'unité répétée), ainsi que des enzymes impliquées dans la régulation des gènes de biosynthèse des EPS, la translocation membranaire, la détermination de la longueur de chaîne et de la polymérisation des unités répétées synthétisées (Broadbent *et al.*, 2003).

L'étude de la biosynthèse des EPS chez *E. coli* a permis d'élaborer un modèle de biosynthèse qui est transposable dans la plupart des bactéries lactiques (Whitfield & Paiment, 2003).

1.4.1 Mécanisme ubiquitaire de biosynthèse des EPS

Chez *E. coli*, les unités répétées sont synthétisées par adjonction successive de molécules glucidiques sur un accepteur lipidique, une molécule d'undecaprényl diphosphate qui se trouve au niveau du feuillet interne de la membrane cytoplasmique. Cet assemblage est catalysé par des glycosyltransférases, à partir d'oses nucléotidylés (Whitfield & Paiment, 2003). La classification proposée en 1999 par le même auteur (Whitfield & Roberts, 1999) permet de distinguer quatre groupes de capsules qui se différencient notamment par le mode de polymérisation des unités répétées. Ainsi, la polymérisation des unités répétées des polysaccharides de capsule des groupes 1 et 4 implique les protéines Wzx et Wzy, alors que

les unités répétées des polysaccharides de capsule des groupes 2 et 3 sont assemblées par une voie métabolique nécessitant l'intervention de transporteurs ABC (Whitfield & Paiment, 2003).

Les études de génétique fonctionnelle et la comparaison des produits des gènes des clusters *eps* suggèrent que les mécanismes de biosynthèse des EPS chez *L. lactis* (Figure 8) (van Kranenburg *et al.*, 1997) (van Kranenburg *et al.*, 1999a) (van Kranenburg *et al.*, 1999b) et *S. thermophilus* (Stingele *et al.*, 1996a) (Stingele *et al.*, 1999a) sont similaires aux mécanismes mis en évidence chez *E. coli*. Par exemple, la biosynthèse des EPS chez *S. thermophilus* Sfi6 nécessite une molécule acceptrice lipidique sur laquelle les unités répétées sont formées par action séquentielle de glycosyltransférases (Stingele *et al.*, 1999a). La première étape est catalysée par EpsE, une galactosyl-1-phosphate transférase, qui catalyse la liaison d'une molécule de galactosyl phosphate activée par un nucléotide à la molécule porteuse lipidique. Puis trois autres glycosyltransférases, EpsG, EpsI et EpsF, catalysent la liaison séquentielle d'une molécule nucléotidylée de N-acétylgalactosamine, de glucose et de galactose pour former l'unité répétée de l'EPS produit.

La biosynthèse des capsules de structure complexe chez les Streptocoques pathogènes nécessite également un accepteur lipidique et suit un mécanisme comparable (Broadbent *et al.*, 2003).

Peu de connaissances sont disponibles sur la polymérisation des EPS et leur translocation chez *S. thermophilus* et les autres bactéries à coloration de Gram positive. Cependant, certaines protéines potentielles codées par les clusters *eps* des bactéries lactiques présentent une faible homologie avec les protéines Wzy et Wzx d'*E. coli*, suggérant que ces étapes suivent un mécanisme similaire. Les protéines EpsJ de *S. thermophilus* Sfi6, EpsK de *Lb. delbrueckii* subsp. *bulgaricus* Lfi5 et EpsI de *L. lactis* NIZOB40 présentent respectivement 25, 19 et 22 % d'identité avec la protéine Wzy. Des pourcentages d'identité

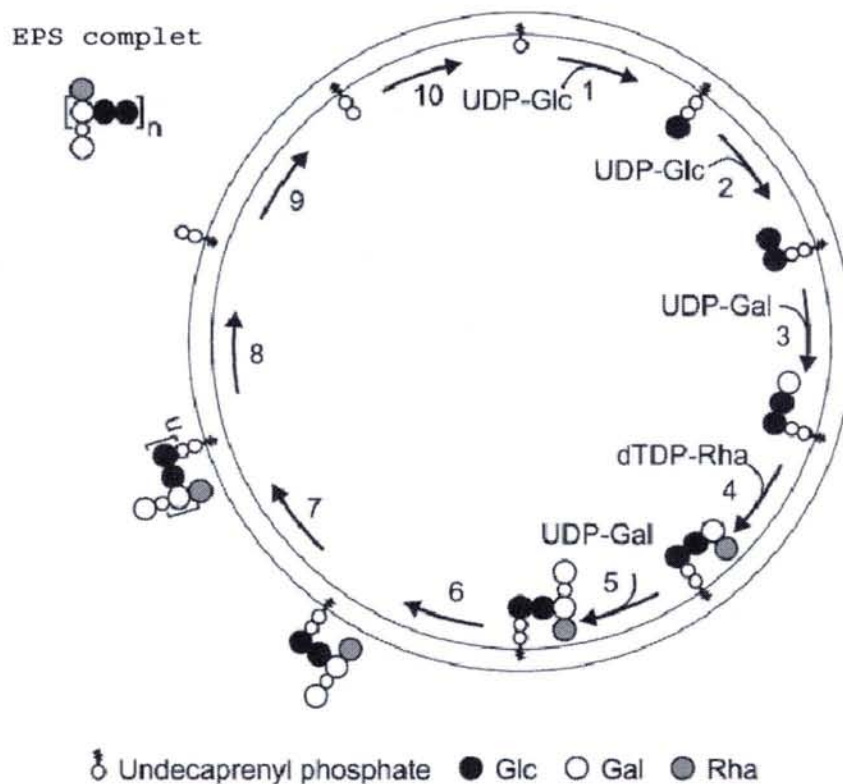


Figure 8: Modèle de la synthèse d'EPS chez *L. lactis* NIZOB40 (Boels *et al.*, 2001).

La protéine EpsD (1) assure la fixation d'un glucose phosphate issu d'une molécule d'UDP-glucose sur le porteur lipidique. Après l'intervention d'EpsE et d'EpsF (2) qui catalysent la fixation d'une nouvelle molécule de glucose, EpsG (3) catalyse l'ajout d'une molécule de galactose. L'unité répétée est achevée par l'ajout de rhamnose (4) et de galactose phosphate (5) grâce à EpsH et EpsJ. Puis sa translocation est catalysée par EpsK (6). EpsI catalyse la polymérisation des sous-unités répétées alors qu'EpsA (7) et EpsB (8) déterminent la longueur de la chaîne. La molécule porteuse subit une nouvelle translocation (9) et est déphosphorylée (10).

dTDP : déoxyThymidine DiPhosphate
 UDP : Uridine DiPhosphate

Glc : Glucose
 Gal : Galactose
 Rha : Rhamnose

comparables sont relevés entre la protéine Wzx et les protéines EpsM de *S. thermophilus* Sf16, EpsN de *Lb. delbrueckii* subsp. *bulgaricus* Lfi5 et EpsK de *L. lactis* NIZOB40 (Jolly & Stingeles, 2001).

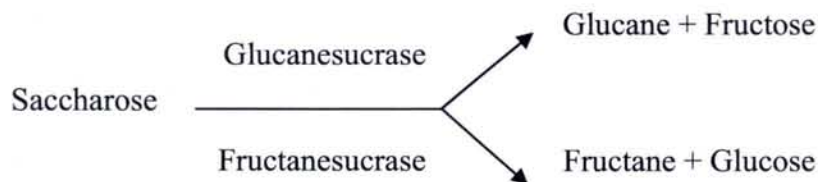
Les taux d'EPS produits varient considérablement selon la souche, les conditions de culture et la composition du milieu. Ils varient entre 50 et 550 mg/l pour *S. thermophilus*, entre 60 et 150 mg/l pour *Lb. delbrueckii* subsp. *bulgaricus*, entre 26 et 600 mg/l pour *L. lactis* subsp. *cremoris* et entre 50 et 60 mg/l pour *Lactobacillus casei* (Cerning, 1995) #123}. La souche LY03 de *S. thermophilus* produit jusqu'à 1,5 g/l lorsqu'elle bénéficie d'un ratio carbone/azote optimal (Degeest *et al.*, 2001). Ces quantités sont faibles par rapport à celles produites par d'autres bactéries, comme *Xanthomonas campestris* qui produit entre 10 et 25 g/l de xanthane (Becker *et al.*, 1998). Cette constatation est à l'origine de travaux visant à augmenter le taux de production d'EPS chez les bactéries lactiques, ce qui nécessite une meilleure compréhension des voies métaboliques impliquées.

1.4.2 Le cas particulier de la biosynthèse des EPS simples

La synthèse d'EPS simples tels que la capsule de type 3 de *S. pneumoniae* et les molécules d'acide hyaluronique de *S. pyogenes* ne nécessite aucune liaison à un accepteur lipidique. La capsule de type 3 est ainsi polymérisée par un mécanisme d'addition de molécules de glucose et d'acide glucuronique grâce à une synthase associée à la membrane (Cartee *et al.*, 2000). De même, la synthèse des molécules d'acide hyaluronique est catalysée par une hyaluronate synthase associée à la membrane, à partir d'acide glucuronique et de N-acétylglucosamine (Crater & van de Rijn, 1995).

De même, la biosynthèse de la plupart des homopolysaccharides produits par les bactéries lactiques se différencie du modèle décrit précédemment. Ils sont ainsi synthétisés au niveau

extracellulaire par des glycanesucrases (glucanesucrase et fructanesucrase) qui utilisent les molécules de saccharose comme donneuses de molécules de fructose ou glucose :



Ces enzymes sont des transglycosylases qui utilisent l'énergie de la liaison osidique pour catalyser le transfert d'un résidu glycosyl (Monsan *et al.*, 2001).

1.4.3 Interconnexion de la biosynthèse des EPS avec d'autres voies métaboliques

La biosynthèse des EPS peut faire intervenir des protéines, dont les fonctions ne sont pas spécifiques à cette biosynthèse. C'est le cas des enzymes impliquées dans la biosynthèse des molécules porteuses lipidiques et des oses nucléotidylés, qui sont nécessaires à l'assemblage d'autres polymères tels que le peptidoglycane ou les acides téichoïques (Broadbent *et al.*, 2003).

Un des facteurs limitant la biosynthèse des EPS pourrait être la compétition s'exerçant au niveau des molécules d'undécaprényl-phosphate puisqu'elles interviennent également dans la production du peptidoglycane et des acides téichoïques. Ainsi, chez *L. lactis*, la production d'EPS est réduite lorsque les conditions de culture favorisent la croissance et les divisions cellulaires, alors qu'elle est augmentée lorsque les conditions de culture réduisent la synthèse de nouvelle paroi cellulaire (Looijesteijn & Hugenholtz, 1999). Cette interconnexion de la voie de biosynthèse des EPS avec celle du peptidoglycane est illustrée par les travaux de Stingle et Mollet (Stingle & Mollet, 1996b) chez *S. thermophilus* Sfi6. Quatre des mutants obtenus par mutagenèse insertionnelle avec le transposon Tn916 présentent une morphologie altérée et un caractère non producteur d'EPS. L'analyse de ces mutants a révélé que Tn916 s'est intégré dans le gène *pbp2b*, qui code une protéine appartenant à la famille des protéines

de liaison à la pénicilline (PBP). Elles présentent une activité transglycosylase nécessaire à la polymérisation des brins de glycane et/ou une activité transpeptidase impliquée dans la formation des chaînes peptidiques du peptidoglycane. Chez ces mutants, l'absence de la protéine PBP 2b se traduirait par une accumulation de précurseurs du peptidoglycane liés à des molécules d'undécaprényl-phosphate, ce qui réduirait le nombre de molécules disponibles pour la synthèse des EPS.

Cependant, des résultats contradictoires ont été obtenus chez *S. thermophilus*. Ainsi, il a été montré que la production d'EPS est maximale quand le taux de croissance est lui aussi maximal (De Vuyst *et al.*, 1998). Il a également été montré que la bacitracine, un composé qui bloque le recyclage des molécules d'undécaprényl-phosphate, n'inhibe pas la production d'EPS chez *S. thermophilus* (Stingele *et al.*, 1999a). Ces résultats suggèrent que cette bactérie utilise une molécule acceptrice alternative ou que la disponibilité en undécaprényl-phosphate n'est pas limitante.

1.5 Variabilité des clusters *eps/cps* chez les bactéries lactiques et les Protéobactéries

Le cluster *eps* de *S. thermophilus* et des autres bactéries lactiques thermophiles présente une localisation chromosomique (De Vuyst *et al.*, 2001) (Broadbent *et al.*, 2003). Il en est de même pour le cluster *cps* de *S. pneumoniae* (Garcia *et al.*, 2000). Par contre, le cluster *eps* de *L. lactis* présente une localisation plasmidique. Ainsi, la souche NIZOB40 possède un plasmide conjugatif de 43 kb (pNZ4000) qui contient le cluster *eps* (van Kranenburg *et al.*, 1997). Cette localisation plasmidique pourrait expliquer la forte fréquence de perte du caractère producteur d'EPS observé chez cette espèce (De Vuyst *et al.*, 2001). Le même phénomène est observé chez *S. thermophilus* et serait dû à l'activité d'éléments génétiques mobiles ou à une instabilité génomique (Bourgoin *et al.*, 1999).

1.5.1 Complexité du cluster *eps* de *S. thermophilus*

Vingt clusters *eps* ont été séquencés chez *S. thermophilus* : ceux des souches Sfi6 (Stingle *et al.*, numéro d'accèsion U40830), Sfi39 (Germond *et al.*, numéro d'accèsion AF373595), NCFB2393 (Almiron-Roig *et al.*, numéro d'accèsion Y17900), MR-1C (Welker *et al.*, numéro d'accèsion AF448249), MR-2C (Welker *et al.*, numéro d'accèsion AY061649), MTC310 (Broadbent *et al.*, numéro d'accèsion AF448502), MTC330 (Welker *et al.*, numéro d'accèsion AF430847), MTC360 (Welker *et al.*, numéro d'accèsion AF434993), CNRZ368 (Bourgoin *et al.*, numéro d'accèsion Z98171), IP6757 (Pluvinet *et al.*, numéro d'accèsion AJ289861), et NST2280 (Pluvinet *et al.*, numéro d'accèsion AF517575), mais aussi les clusters de types I (Rallu *et al.*, numéro d'accèsion AF410175), III (Rallu *et al.*, numéro d'accèsion AY057915), IV (Rallu *et al.*, numéro d'accèsion AF454495), V (Rallu *et al.*, numéro d'accèsion AF454496), VI (Rallu *et al.*, numéro d'accèsion AF454497), VII (Rallu *et al.*, numéro d'accèsion AF454498), IX (Rallu *et al.*, numéro d'accèsion AF454499), X (Rallu *et al.*, numéro d'accèsion AF454500) et XI (Rallu *et al.*, AF454501). La figure 9 présente cinq clusters *eps* différents de *S. thermophilus*.

Chacune des séquences de ces vingt clusters comprend des ORF codant des produits potentiels qui présentent des homologies avec des protéines impliquées dans la biosynthèse des EPS chez d'autres bactéries. Une ORF distale de ces clusters, *orf14.9*, code un produit potentiel ne présentant aucune homologie avec des protéines connues. Cette ORF est orientée en sens inverse par rapport aux autres gènes du cluster, et est présente dans les différents clusters.

1.5.1.1 Organisation opéronique et modulaire de la première région du cluster *eps* (3' d'*orf14.9*)

L'expression du cluster *eps* de *S. thermophilus* Sfi6 est dirigée par un transcrit unique d'environ 15 kb qui correspond au cluster séquencé. Le site d'initiation de la transcription a

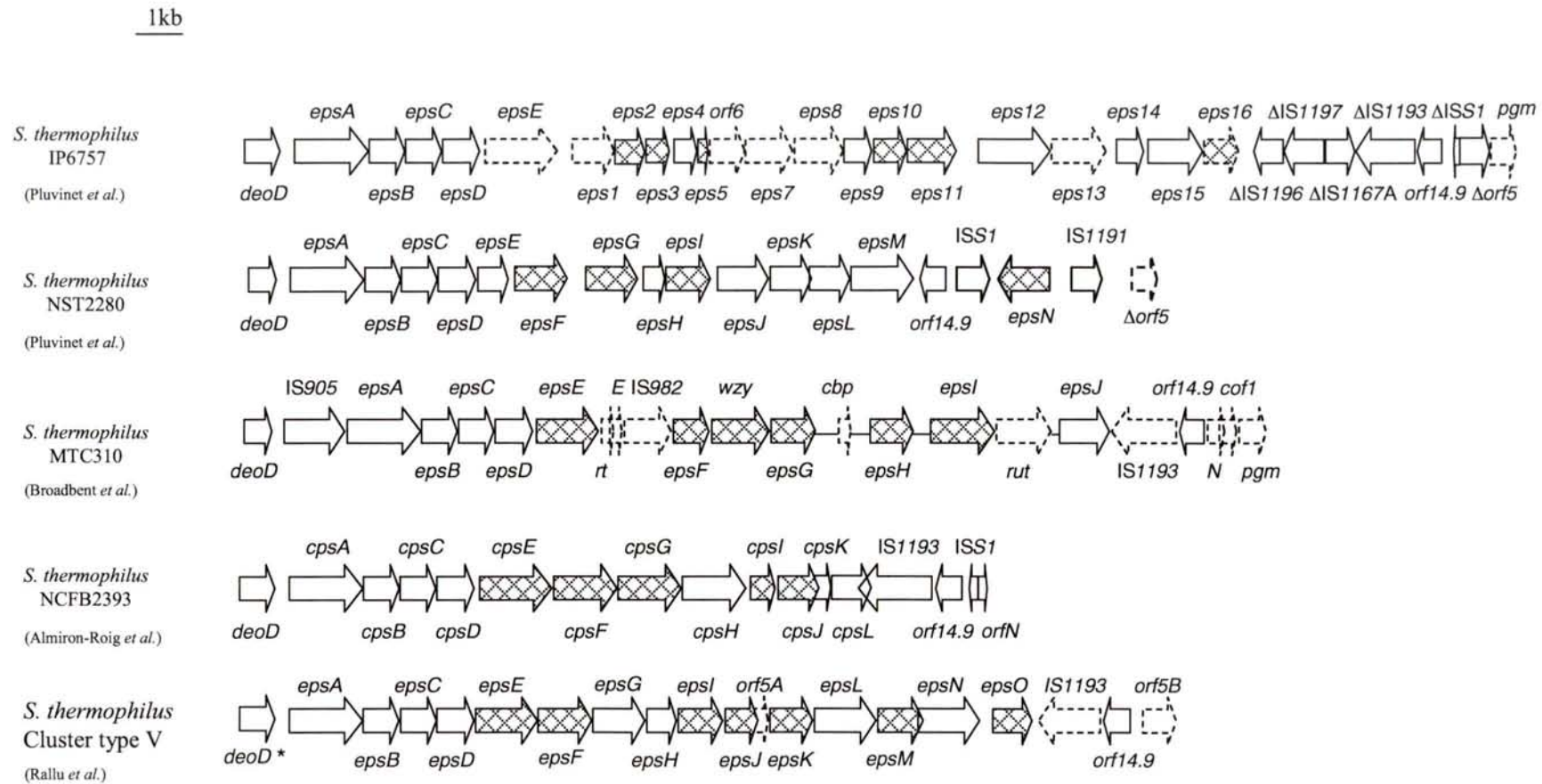


Figure 9 : Organisation du cluster *eps* de *S. thermophilus* IP6757, NST2280, MTC310, NCFB2393 et du cluster *eps* de type V. Les ORF annotées comme codant une glycosyltransférase potentielle sont hachurées. Les ORF dont les flèches sont en pointillés sont des pseudo-ORF.
 * : dans le cluster *eps* de type V, la séquence de *deoD* disponible n'est que partielle.

été localisé en amont d'*epsA* par extension d'amorces (Stingele *et al.*, 1999b). Cette organisation opéronique est illustrée par l'effet polaire observé sur l'expression des deux dernières ORF du cluster (*epsL* et *epsM*) lors de l'intégration du plasmide pSA3 à la fin d'*epsK* (Stingele *et al.*, 1996a). Une organisation opéronique du cluster *cps* de *S. thermophilus* NCFB2393 est également proposée par Almiron-Roig *et al.* (Almiron-Roig *et al.*, 2000).

Les comparaisons de séquence des différents opérons des clusters *eps* de *S. thermophilus* suggèrent qu'ils présentent une organisation modulaire. Ainsi, le premier module, très fortement conservé dans tous les clusters étudiés, contient les quatre premiers gènes de l'opéron qui seraient impliqués dans la régulation (*epsA* et *epsB*), la polymérisation et son contrôle (*epsC* et *epsD*) lors de la production d'EPS (Jolly & Stingele, 2001). Le second module, situé en aval du premier, est très variable, aussi bien par la taille que par le nombre et la séquence des gènes impliqués dans la biosynthèse et l'export des EPS. C'est dans ce module que sont localisés les gènes codant les glycosyltransférases potentielles dont le nombre et la séquence sont très variables selon les souches, à l'exception de la première glycosyltransférase (Jolly & Stingele, 2001). Celle-ci est codée par le premier gène de ce module et présente des homologies avec une undecaprényl-glycosyl-1-phosphate transférase, enzyme qui catalyse l'accrochage de la première molécule osidique sur la molécule porteuse d'undecaprényl-phosphate (Jolly & Stingele, 2001). De plus, cette première glycosyltransférase présente un degré d'homologie important tant chez différentes protéobactéries (*E. coli* et *Klebsiella pneumoniae* notamment) que chez différentes bactéries présentant une coloration de Gram positive appartenant à la division des Firmicutes dont *S. pneumoniae* ou présentant un haut contenu en bases G et C (Wang *et al.*, 1996). Cette conservation de la première glycosyltransférase est par exemple observée au niveau du cluster *eps* de *Streptomyces sp. 139*, bactérie à coloration de Gram positive et présentant un génome à

haut pourcentage en G+C (Wang *et al.*, 2003). Cependant chez *S. thermophilus*, deux gènes différents ont été identifiés. Le premier code une protéine de type CpsE de la souche NCFB2393 qui présenterait une activité bi-fonctionnelle, correspondant à une activité glucosyl-1-phosphate transférase et à une activité de libération de l'unité répétée du transporteur lipidique (Almiron-Roig *et al.*, 2000). Le deuxième code une protéine de type EpsE de la souche Sfi6, qui posséderait seulement une activité galactosyl-1-phosphate transférase (Jolly & Stingle, 2001).

1.5.1.2 Présence d'autres gènes dans le cluster : bordures du cluster

Le premier module des clusters *eps* est situé immédiatement en aval de l'ORF *deoD* qui code une protéine homologue à une phosphorylase de nucléosides puriques et n'est donc pas impliquée dans la biosynthèse des EPS (Broadbent *et al.*, 2003). Cette ORF constitue ainsi une des frontières du cluster *eps* chez *S. thermophilus* (Figure 10). Dans la suite de ce propos, *deoD* sera désignée comme la frontière gauche du cluster *eps* de *S. thermophilus*. Les souches Sfi39 et MTC310 sont les seules souches qui diffèrent de ce schéma puisque la région intergénique séparant *deoD* et le premier gène du cluster *eps* est interrompue par une copie de l'élément IS905.

Pour certains auteurs, l'autre frontière du cluster *eps* chez *S. thermophilus* est localisée au niveau d'*orf14.9* (Jolly & Stingle, 2001). Cependant, des gènes codant des glycosyltransférases potentielles ou d'autres enzymes potentielles impliquées dans la synthèse d'EPS ont été identifiés en aval d'*orf14.9* dans différents clusters *eps* (Figure 10). Ainsi, *eps/cpsV* codant une glycosyltransférase potentielle ainsi que *epsU/cpsU* et *epsX/cpsX* codant des protéines potentiellement impliquées dans la translocation membranaire des EPS ont été identifiés en aval d'*orf14.9* dans le cluster *eps* des souches MR-1C et CNRZ368. De même, les clusters *eps* des souches MR-2C et MTC360 présentent en aval d'*orf14.9*, les ORF

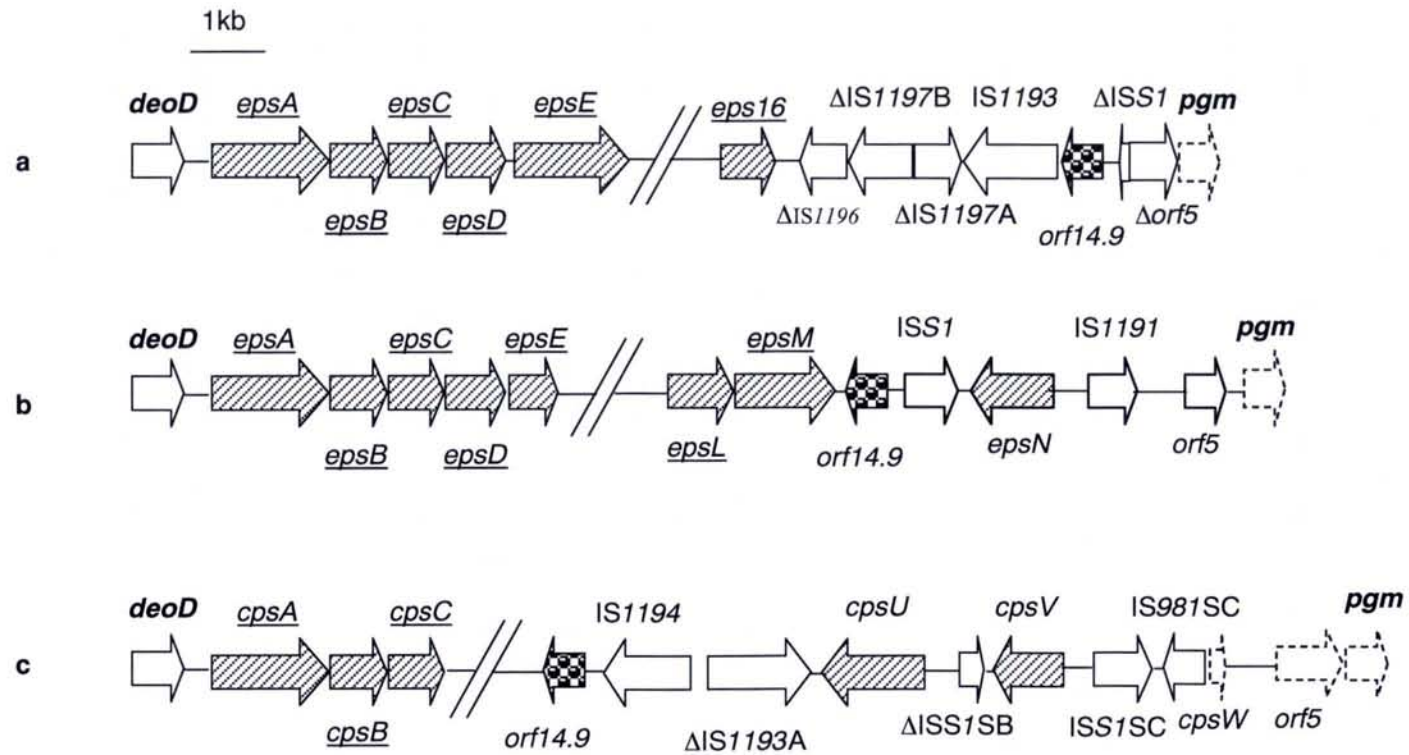


Figure 10 : Localisation des frontières du cluster *eps* de *S. thermophilus* dans les souches IP6757 (Pluvinet *et al.*) et LMG18311 (Bolotin *et al.*) (a), CNRZ1066 (Bolotin *et al.*) (b) et MR-1C (Welker *et al.*) (c).

Le cluster *eps* des souches IP6757 et LMG18311 présentent plus de 99% d'identité. Les clusters *eps* représentés sont les seuls parmi l'ensemble des clusters *eps* séquencés à présenter une séquence complète pour *deoD* et une séquence allant jusqu'à *pgm*. Le cluster *eps* des souche NST2280, MR-2C et MTC360 présentent 99% d'identité avec la séquence du cluster *eps* de la souche CNRZ1066. Le cluster *eps* de la souche CNRZ368 présente 99% d'identité avec le cluster *eps* de la souche MR-1C.

Les ORF hachurées sont impliquées dans la biosynthèse des EPS. Le nom des ORF opéroniques appartenant à la première région des clusters *eps* est souligné. Δ correspond à des ORF délétées ou à des éléments IS dont l'ORF codant la transposase est délétée. La double barre indique que l'échelle n'est pas respectée dans cette région. Les ORF figurées en pointillés sont des pseudo-ORF. L'ORF *orf14.9* longtemps considérée comme la frontière droite du cluster *eps* de *S. thermophilus* est indiquée avec un symbolisme particulier.

epsO/cpsO et *epsX/cpsX* qui codent respectivement une glycosyltransférase potentielle et une protéine potentiellement impliquée dans la translocation membranaire. Ces résultats suggèrent que le cluster *eps* de *S. thermophilus* s'étend au-delà d'*orf14.9*, mais l'implication de ces ORF dans la biosynthèse des EPS repose sur des homologues de séquence. Cependant, il a été montré chez la souche NST2280 qu'*epsN*, un gène situé immédiatement en aval d'*orf14.9* et codant une glycosyltransférase potentielle est impliqué dans la production d'EPS (Pluvinet *et al.*, 2004). Ce résultat amène les auteurs à proposer les ORF *orf5* et *pgm* comme frontière du cluster *eps*. Dans la suite de ce propos, *pgm* sera désignée comme la frontière droite du cluster *eps* chez *S. thermophilus*.

Ainsi, le cluster *eps* de *S. thermophilus* ne se limite pas à l'opéron identifié en amont d'*orf14.9*, mais il comprend également d'autres ORF, dont *orf14.9*, qui constituent la seconde région du cluster *eps*.

1.5.2 Organisation du cluster *eps/cps* des autres bactéries lactiques

1.5.2.1 *Organisation opéronique et modulaire*

De nombreux clusters *eps* ont été caractérisés chez d'autres bactéries lactiques. Par exemple, chez *L. lactis* NIZOB40 (van Kranenburg *et al.*, 1997) et *L. lactis* subsp *cremoris* HO2 (Forde & Fitzgerald, 2003), le cluster *eps* est localisé en aval d'une copie de l'élément IS982 et en amont d'*orfY*, une ORF de fonction inconnue (Figure 11). Le séquençage des clusters *eps* de *L. lactis* NIZOB891 et NIZOB35 n'est que partiel et ne permet pas de localiser avec précision les frontières du cluster *eps* (van Kranenburg *et al.*, 1999a). Chez *S. pneumoniae*, les clusters *cps* correspondant à seize sérotypes différents ont été caractérisés. Le séquençage des clusters correspondants aux autres sérotypes est réalisé à l'Institut Sanger (Lopez & Garcia, 2004). Les séquences correspondantes sont à ce jour considérées comme préliminaires (http://www.sanger.ac.uk/Projects/S_pneumoniae/CPS/). Quatorze des seize

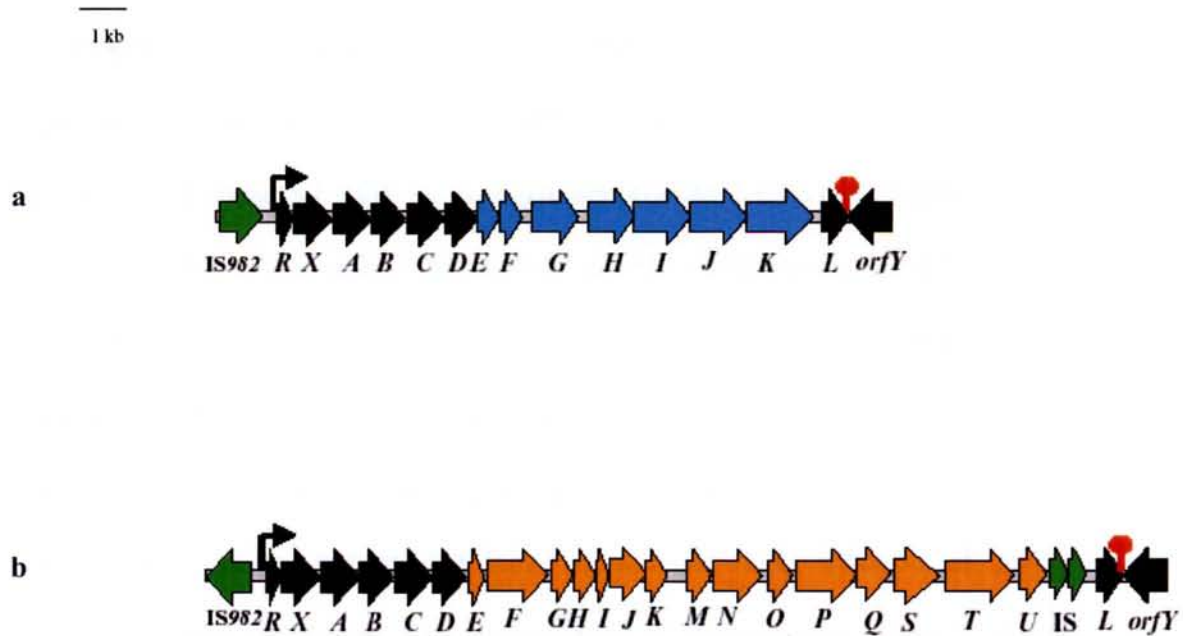


Figure 11 : Représentation schématique du cluster *eps* de *L. lactis* NIZOB40 (a) et de *L. lactis* subsp *cremoris* HO2 (b), d'après van Kranenburg *et al.* (1997) et Forde *et al.* (2003).

La flèche symbolise le promoteur de transcription de ces deux clusters et le terminateur de transcription est figuré entre *epsL* et *orfY* dans les deux clusters. Les frontières de ces deux clusters *eps* sont indiquées.

Les ORF figurées par des flèches noires sont quasiment identiques entre les clusters.

clusters connus présentent comme frontières *dexB* et *aliA*, deux gènes qui ne sont pas impliqués dans la biosynthèse de la capsule (Lopez & Garcia, 2004) (Figure 12).

Comme pour *S. thermophilus*, ces différents clusters présentent une organisation opéronique. Ainsi, le cluster *eps* de *L. lactis* NIZOB40 est transcrit selon un ARN messager unique et polycistronique de 12 kb (van Kranenburg *et al.*, 1997). Le promoteur utilisé est localisé en amont d'*epsR*, la première ORF du cluster. Sa fonctionnalité est montrée par insertion d'un fragment contenant ce promoteur en amont du gène rapporteur *gusA* dépourvu de promoteur. De plus, le point d'initiation de la transcription a été déterminé par extension d'amorce en 5' d'*epsR*. Un terminateur de transcription potentiel est détecté immédiatement en aval d'*epsL*, la dernière ORF du cluster (van Kranenburg *et al.*, 1997). Chez *S. pneumoniae*, les clusters *cps* présentent la même organisation opéronique mais l'activité du promoteur en amont de la première ORF a été montrée expérimentalement seulement pour les cluster conférant les sérotypes 1 (Munoz *et al.*, 1997), 14 et 33 (Llull *et al.*, 1998). Cette organisation opéronique peut être retrouvée chez d'autres bactéries lactiques, par exemple pour le cluster *eps* de la souche Lfi5 de *Lb. delbrueckii subsp bulgaricus* (Lamothe *et al.*, 2002).

L'organisation modulaire des gènes de l'opéron du cluster *eps* décrite chez *S. thermophilus* est également retrouvée chez les autres bactéries lactiques. Ainsi, dans le cluster *eps* de *L. lactis* NIZOB40 (van Kranenburg *et al.*, 1997), NIZOB891 (van Kranenburg *et al.*, 1999a), NIZOB35 (van Kranenburg *et al.*, 1999a) et dans le cluster *eps* du plasmide pCI658 (Forde & Fitzgerald, 2003), le premier module contient les cinq premières ORF identifiées. Quatre de ces ORF sont impliquées dans la régulation (*epsR*), et dans la polymérisation et l'export (*epsA*, *B* et *C*), alors qu'*epsX* qui est intercalée entre *epsR* et *epsA* ne présente pas de fonction connue. Le second module, qui est localisé en aval du premier, est constitué des gènes impliqués dans la biosynthèse et l'export des EPS. Ce module est

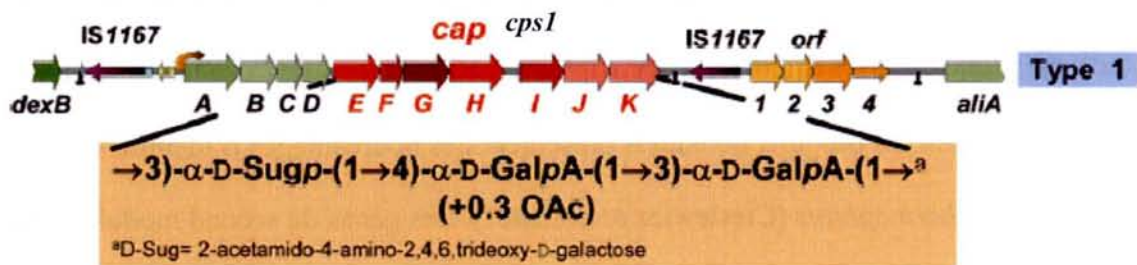


Figure 12 : Représentation schématique du cluster *cps* de *S. pneumoniae* de type 1 d'après Lopez *et al.* (2004). Les frontières *dexB* et *aliA* sont indiquées.

Les clusters *cps* de *S. pneumoniae* de type 2, 4, 6B, 8, 9V, 14, 18C, 19F, 19A, 19B, 19C, 23F et 33F présentent une organisation comparable ainsi que des frontières identiques.

La composition des unités répétées de la capsule de la souche de type 1 est également indiquée.

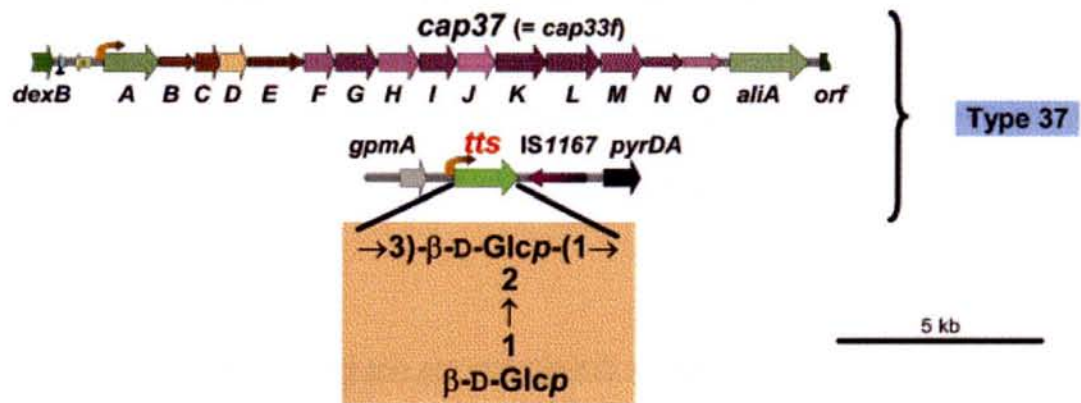
spécifique à chaque cluster, mais sa première ORF qui code la première glycosyltransférase (EpsD) est très conservée (van Kranenburg *et al.*, 1997) (van Kranenburg *et al.*, 1999a). Les clusters *cps* de *S. pneumoniae* présentent la même organisation modulaire. Ainsi, le premier module est très conservé et contient les quatre premiers gènes des clusters *cps* (Lopez & Garcia, 2004), dont les produits potentiels présentent des pourcentages d'identité significatifs avec ceux de *S. thermophilus* (Cieslewicz *et al.*, 2001). Les gènes du second module, situés en aval, sont sérotype spécifiques et comprennent ceux codant les glycosyltransférases (Lopez & Garcia, 2004). Comme chez *S. thermophilus*, le gène codant la première glycosyltransférase (CpsE) est hautement conservé dans les clusters conférant les sérotypes 2, 4, 14, 19F/A/B/C, 23F et 33F (Lopez & Garcia, 2004). Lorsque ce gène code une glucosyltransférase, deux classes de gènes *cpsE* peuvent être distinguées selon leurs séquences (Morona *et al.*, 1999).

1.5.2.2 Organisation différente des clusters des souches de *S. pneumoniae* type 3 et type 37

Les souches de *S. pneumoniae* produisant une capsule de sérotype 37 présentent un cluster *cap* localisé entre *dexB* et *aliA*, et présentant la même organisation modulaire et opéronique que les autres souches de *S. pneumoniae*, mais il n'est pas impliqué dans la production de la capsule. Un gène situé hors de ce cluster *cap*, le gène *tts*, code la synthase nécessaire à la biosynthèse de la capsule de cette souche (Figure 13a) (Llull *et al.*, 1999) (CF paragraphe 1.7.2).

Le cluster *cap* des souches de *S. pneumoniae* type 3 diffère du schéma général et ne présente pas l'organisation modulaire décrite plus haut. Sa localisation immédiatement en aval de *dexB* n'est pas respectée puisque d'autres ORF sont intercalées (Figure 13b) et *aliA* n'a pas été identifié (Arrecubieta *et al.*, 1995). Le cluster *cap* de ces souches ne contient pas les quatre premières ORF habituellement présentes au sein des clusters *cap* de *S. pneumoniae*, même si des séquences homologues contenant de nombreuses mutations et non transcrites

a



b

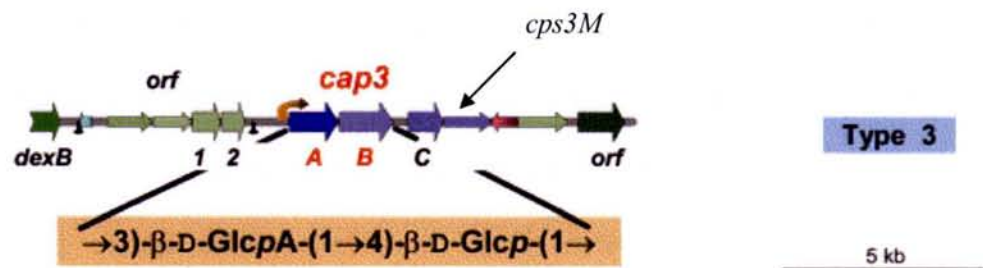


Figure 13 : Organisation des clusters *cap* de *S. pneumoniae* de type 37 (a) et de type 3 (b), d'après Lopez et Garcia (2004).

La composition des unités répétées de la capsule de ces deux souches est également indiquée.

(Arrecubieta *et al.*, 1995) sont présentes immédiatement en amont du cluster *cap* des souches de type 3. Leur cluster est constitué seulement de quatre ORF opéroniques: *cap3A*, *cap3B*, *cap3C* (ou *cap3U*) (Figure 13b). Le promoteur impliqué est localisé immédiatement en amont de *cap3A* (Arrecubieta *et al.*, 1995) (Caimano *et al.*, 1998). Seules les deux premières ORF sont impliquées dans la biosynthèse de la capsule de type 3, *cap3C* étant dispensable et *cap3M* codant un produit présentant une activité potentielle de phosphoglucomutase (Arrecubieta *et al.*, 1995) (Dillard & Yother, 1994).

1.5.3 Organisation modulaire des clusters *cps* des Protéobactéries

Un cluster *cps* permettant de synthétiser une capsule de groupe 1 a été isolé chez *Escherichia coli* E69 (O9a:K30). Il peut être divisée en deux modules : le premier comporte les gènes impliqués dans la translocation des unités répétées et le second les gènes impliqués dans la synthèse et la polymérisation des unités répétées. Le premier module est très conservé parmi les souches d'*E. coli* synthétisant une capsule appartenant au groupe 1, à l'inverse du second qui varie selon les souches. Dans ce second module, quatre gènes codant des glycosyltransférases potentielles ont été identifiés (Drummelsmith & Whitfield, 1999). Les capsules produites par les souches du genre *Klebsiella* ont des structures similaires à celles du groupe 1 d'*E. coli*. La comparaison des clusters *cps* de six souches de *Klebsiella pneumoniae* et de douze souches d'*E. coli* appartenant au groupe 1 révèle qu'ils présentent la même localisation chromosomique à proximité des gènes *his* et *rfb* (Drummelsmith & Whitfield, 1999) et que le module impliqué dans la translocation est très conservé. Ces résultats suggèrent l'intervention de transferts horizontaux entre ces organismes (Rahn *et al.*, 1999).

Les clusters *cps* codant les capsules des groupes 2 et 3 sont localisés à proximité du gène *serA* (Clarke *et al.*, 1999). L'analyse des clusters impliqués dans la biosynthèse des capsules du groupe 2 révèle une organisation en trois modules distincts (Figure 14). Les modules 1 et 3

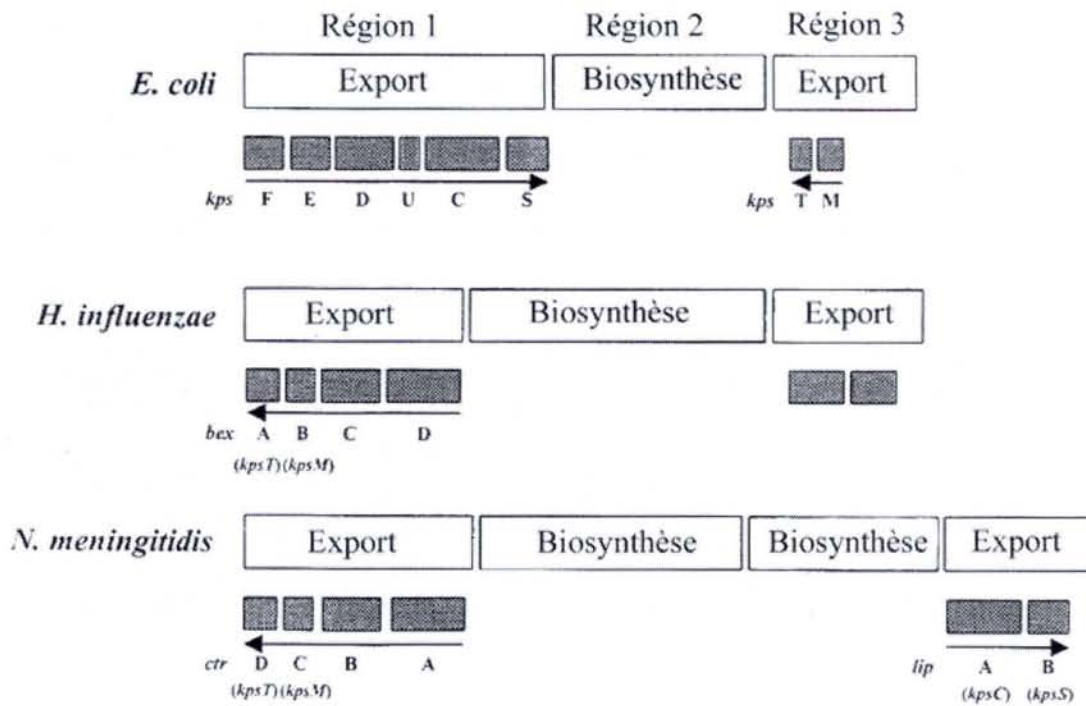


Figure 14 : Organisation du cluster *cps* codant la capsule de type 2 chez *Escherichia coli* et des clusters *cps* d'*Haemophilus influenzae* et de *Neisseria meningitidis*, d'après Clarke *et al.* (1999).

La région 2 figurée est sérotype spécifique.

Le nom des gènes est indiqué et les flèches indiquent le sens de la transcription. Les gènes de biosynthèse de la capsule n'ont pas été représentés. Les gènes d'*H. influenzae* et *N. meningitidis*, homologues aux gènes *kps* d'*E. coli*, sont indiqués entre parenthèse.

sont conservés et contiennent les gènes impliqués dans l'export des unités répétées. Le module 2 est sérotype-spécifique et contient les gènes impliqués dans la synthèse du polysaccharide (Clarke *et al.*, 1999). L'étude des clusters impliqués dans la biosynthèse de la capsule du groupe 3 montre une organisation modulaire comparable à celle décrite pour les clusters des souches du groupe 2. Le module central est sérotype-spécifique et est bordé par deux modules conservés dont les gènes présentent des homologies avec ceux des modules 1 et 3 des clusters du groupe 2 mais sont organisés différemment (Clarke *et al.*, 1999).

Les clusters impliqués dans la biosynthèse de la capsule chez *Neisseria meningitidis* et *Haemophilus influenzae* présentent une organisation comparable à celle décrite pour les clusters des groupes 2 et 3 chez *E. coli*. Ainsi, deux modules impliqués dans l'export de la capsule bordent un module central, sérotype-spécifique, impliqué dans sa biosynthèse (Figure 14) (Clarke *et al.*, 1999).

1.6 Deux niveaux de régulation de la production d'EPS chez les bactéries lactiques

1.6.1 Régulation codée par le cluster *eps/cps*

La première ORF des clusters *eps/cps* des bactéries lactiques (*epsA* chez *S. thermophilus*, *cpsA* chez *S. pneumoniae*, *Streptococcus agalactiae* et *S. thermophilus* NCFB2393 et *epsR* chez *L. lactis*) code une protéine présentant 20 à 35% d'identité avec la protéine LytR de *Bacillus subtilis* (Jolly & Stingle, 2001) (Broadbent *et al.*, 2003). Cette petite protéine basique de 35 kDa régule négativement la transcription de l'opéron *lytABC* qui code les molécules d'autolysine (Lazarevic *et al.*, 1992). Cette protéine agit également comme atténuateur de l'expression de l'opéron *lytR* (Lazarevic *et al.*, 1992). L'implication de la protéine CpsA dans la régulation de l'expression du cluster *cps* a été montrée expérimentalement chez *S. agalactiae*. L'inactivation du gène *cpsIaA* montre que la protéine

CpsA exercerait un contrôle positif sur l'expression du cluster puisqu'elle s'accompagne de la présence de deux à trois fois moins de transcrits du cluster *cps* que dans la souche sauvage et d'une diminution de 24% de la production de capsule (Cieslewicz *et al.*, 2001). Cette fonction d'activateur transcriptionnel de CpsA est également retrouvée chez *S. pneumoniae* (Morona *et al.*, 2000). Chez *S. thermophilus* et *L. lactis*, de telles études n'ont pas été menées, mais la fonction de régulation de EpsA et EpsR respectivement, est probablement conservée du fait des pourcentages d'identité élevés de ces protéines avec LytR de *B. subtilis* et CpsA de *S. pneumoniae* et *S. agalactiae* (Broadbent *et al.*, 2003). De plus, la protéine EpsR de *L. lactis* NIZOB40 est homologue aux protéines régulatrices Xre, un répresseur identifié chez *B. subtilis*, et PtrR, un régulateur négatif identifié chez *Pseudomonas aeruginosa* (van Kranenburg *et al.*, 1997).

Une autre régulation codée par les gènes *cpsB*, *cpsC* et *cpsD* du cluster *cps* de *S. pneumoniae* s'exerce au niveau post-transcriptionnel. La protéine CpsD est une tyrosine kinase autophosphorylante qui interagit avec la protéine CpsC. Le complexe formé par ces deux protéines permet la fixation d'ATP, puis un taux maximal de polymérisation des EPS. Quand la protéine CpsD s'autophosphoryle en utilisant l'ATP lié, elle se dissocie de CpsC, et provoque une réduction à un taux basal de la production d'EPS. La protéine CpsB, qui possède une activité de phosphatase, réactive la protéine CpsD et lui permet d'interagir à nouveau avec la protéine CpsC, pour augmenter le taux de production/ polymérisation des EPS (Morona *et al.*, 2000) (Morona *et al.*, 2002). La protéine CpsB inhibe également l'autophosphorylation de CpsD (Bender & Yother, 2001) et joue donc un rôle important dans la régulation de la biosynthèse des CPS chez *S. pneumoniae*. Les interactions proposées pour ces trois protéines constituent un modèle pour le fonctionnement des ORF homologues chez les autres bactéries lactiques (Broadbent *et al.*, 2003).

1.6.2 Régulation codée à l'extérieur du cluster : la protéine RegM

L'étude de la régulation catabolique de l'expression de gènes de virulence chez *S. pneumoniae* a permis de mettre en évidence l'implication de la protéine RegM, codée par un gène situé hors du cluster *cps*, dans la régulation de la production de capsule (Giammarinaro & Paton, 2002). La protéine RegM présente 46% d'identité avec le répresseur catabolique CcpA de *Staphylococcus aureus*. Les études montrent un lien entre la protéine RegM et le métabolisme glucidique mais aussi la virulence de *S. pneumoniae*. En effet, la virulence des mutants présentant une délétion de *regM* est drastiquement réduite par rapport à la souche sauvage. Cette diminution est due à une diminution de la transcription du cluster *cps* de ces mutants, qui est observée en présence de glucose ou de saccharose comme source glucidique (Figure 15). Elle se traduit par une production insuffisante de capsule qui empêche les mutants de résister au système immunitaire de l'hôte. Par ailleurs, des éléments CRE (*Catabolite Response Element*) permettant la fixation de la protéine RegM ont été identifiés en 5' du cluster *cps*.

Des homologues du gène *regM* de *S. pneumoniae* sont présents dans le génome de *S. thermophilus* (*ccpA*) et *S. mutans* (*regM*), dont les produits présentent respectivement 75 et 76% d'identité avec la protéine RegM (Giammarinaro & Paton, 2002). A l'heure actuelle, aucune étude n'a porté sur l'implication du gène *regM* dans la production d'EPS chez *S. thermophilus*. Par contre, chez *S. mutans*, le rôle d'interface de la protéine RegM entre le métabolisme glucidique et la production d'EPS est également observé. Lorsque *regM* est inactivé, l'expression d'un gène rapporteur contrôlé par le promoteur du cluster *gtfBC* (opéron codant deux glucosyltransférases) et celui du gène *ftf* (codant une fructosyltransférase) est drastiquement réduite. Un élément CRE est identifié 238 pb en amont du codon start du gène *ftf* et à 51 pb du codon start de *gtfB* de l'opéron *gtfBC* (Brownngardt *et al.*, 2004).

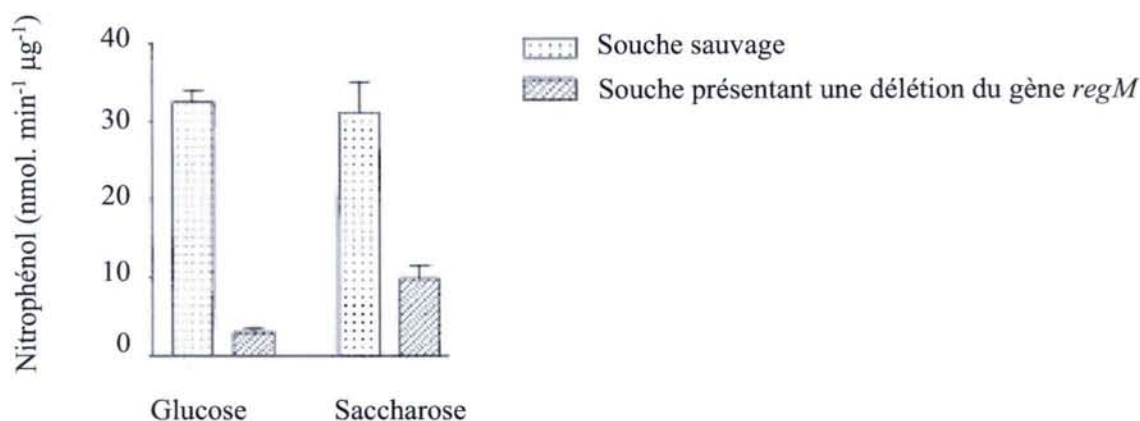


Figure 15 : Expression de du cluster *cps* dans une souche sauvage et dans un mutant délété dans le gène *regM* de *S. pneumoniae*, en présence de glucose et de saccharose, d'après Giammarinaro et Paton (2002).

Cette expression est déterminée par la mesure de l'activité du gène rapporteur *lacZ* figurée en ordonnée. Elle correspond à la quantité de nitrophénol libéré dans le milieu à partir du substrat *p*-nitrophényl α -D-glucopyranoside.

Les écart-types obtenus lors de trois mesures indépendantes sont indiqués.

1.7 Intervention de gènes situés à l'extérieur du cluster dans la production des EPS/CPS

D'autres études effectuées chez *S. pneumoniae* ont permis de mettre en évidence l'implication de gènes situés à l'extérieur du cluster *cps* dans la biosynthèse de la capsule : les gènes *pgm*, *galU* et *tts*.

1.7.1 Intervention de gènes du métabolisme cellulaire dans la biosynthèse de la capsule chez *S. pneumoniae* de type 3

1.7.1.1 Importance du gène *pgm*

Les études menées sur la fonctionnalité du cluster *cap* chez *S. pneumoniae* de type 3 ont montré que les deux premières ORF de ce cluster (*capA* et *capB*) sont indispensables à la production de capsule chez cette souche (Dillard *et al.*, 1995). La quatrième ORF du cluster, *cps3M*, code une phosphoglucomutase potentielle qui convertit le glucose-6-P en glucose-1-P, un intermédiaire métabolique nécessaire à la formation de la capsule de type 3. Un mutant, obtenu par mutagenèse insertionnelle dans *cps3M*, ne présente aucune modification de la production de capsule par rapport à la souche sauvage (Dillard *et al.*, 1995), suggérant que d'autres gènes présents dans le chromosome de *S. pneumoniae* peuvent compléter cette mutation.

Hardy *et al.* (2000) ont étudié un mutant spontané produisant quatre fois moins de capsule que la souche sauvage. La mutation à l'origine de ce phénotype n'est pas localisée dans le cluster *cap* de type 3 mais dans le gène *pgm*, qui code un homologue de Cap3M. Ce gène est localisé à au moins 450 kb du cluster *cap* et sa fonction de phosphoglucomutase a été montrée par des études enzymatiques et des expériences de complémentation (Hardy *et al.*, 2000). L'inactivation insertionnelle de *pgm* s'accompagne d'une importante réduction de la production de capsule par rapport à la souche sauvage. Cette étude démontre que l'essentiel

de l'activité phosphoglucomutase requise pour la production de la capsule de type 3 est assurée par la protéine Pgm, qui est codée par un gène situé hors du cluster. L'activité du gène *cps3M* n'est donc pas nécessaire à la production de capsule (Hardy *et al.*, 2000) et à la virulence de *S. pneumoniae* de type 3 (Hardy *et al.*, 2001).

1.7.1.2 Importance du gène *galU*

L'ORF *cap3C* du cluster *cap* de *S. pneumoniae* de type 3 code une UDP-Glucose pyrophosphorylase (UDPG:PP), qui est nécessaire à la formation d'UDP-Glucose, un intermédiaire clef de la biosynthèse de la capsule chez toutes les souches de *S. pneumoniae*. Or, il a été montré que le gène *cap3C* est dispensable à la production de la capsule chez *S. pneumoniae* de type 3 (Arrecubieta *et al.*, 1995) (Dillard *et al.*, 1995). Un autre gène coderait donc une UDPG:PP, qui serait de plus commun à toutes les autres souches de *S. pneumoniae*. En effet, aucun gène identifié au sein des clusters *cps* de *S. pneumoniae* ne coderait une UDPG:PP (Mollerach *et al.*, 1998).

La recherche d'homologue à *cps3C* dans une séquence partielle du génome de *S. pneumoniae* de type 4 a permis d'identifier le gène *galU*. Ce gène, conservé dans le génome des différents sérotypes analysés, code une protéine dont la fonction d'UDPG:PP est démontrée expérimentalement (Mollerach *et al.*, 1998). Des mutants insertionnels de *galU* obtenus chez *S. pneumoniae* de type 3 mais aussi de type 1 sont incapables de produire une capsule (Mollerach *et al.*, 1998). Le gène *galU* est le gène extérieur au cluster *cps* qui est impliqué dans la synthèse d'UDP-Glucose et qui joue donc un rôle essentiel dans la biosynthèse de la capsule chez *S. pneumoniae* de type 3 mais aussi des autres types (Mollerach *et al.*, 1998).

1.7.2 Intervention du gène *tts* dans la biosynthèse des CPS chez *S. pneumoniae* de type 37

Les souches de type 37 contiennent un cluster *cps* identique à celui des souches de sérotype 33f. Cependant, des mutations interrompant les phases de lecture sont identifiées

dans les ORF *cps37B*, *cps37E*, *cps37N* et *cps37O*. Lors des expériences de transformation d'une souche de *S. pneumoniae* ne produisant aucun CPS avec l'ADN du cluster *cps* de type 37, aucune capsule de type 37 n'est détectée chez les transformants. Ces résultats suggèrent que des gènes localisés hors du cluster *cps* sont impliqués dans la biosynthèse de la capsule de type 37, dont la structure est homopolysaccharidique (Llull *et al.*, 1999).

L'analyse d'un fragment d'ADN génomique de 7 kb, qui confère le caractère producteur d'une capsule de type 37 a permis d'identifier une copie de l'élément IS1167 et 3 ORF (Figure 16) : *gpmA* qui code une phosphoglycéromutase potentielle, *tts* qui ne présente pas d'homologies avec les séquences des banques de données, et *pyrDA* qui code une dihydroorotate dehydrogénase potentielle. Des répétitions inversées de 105 pb sont localisées de part et d'autre de *tts* et de la copie de l'élément IS1167 (Llull *et al.*, 1999).

L'analyse de la séquence partielle du chromosome de *S. pneumoniae* de type 4 montre que les ORF *gpmA* et *pyrDA* sont séparées par plus de 380 kb, suggérant un remaniement génomique majeur dans le clone de type 37 analysé (Llull *et al.*, 1999).

Lors de la transformation d'une souche non productrice de CPS avec le gène *tts*, les transformants isolés présentent une capsule de type 37. Le gène *tts* serait donc le seul gène impliqué dans la biosynthèse de la capsule de type 37, une situation observée seulement chez cette souche. Cependant, les auteurs soulignent que la non implication du cluster *cps* identifié n'a pas été montrée (Llull *et al.*, 1999).

L'ORF *tts* code une protéine potentielle de 509 acides aminés présentant 6 hélices membranaires. Elle présente également des motifs de séquence caractéristiques de β -glycosyltransférases et de cellulose synthases (Llull *et al.*, 1999). Des études expérimentales ont montré que la protéine Tts présentait ces deux activités et qu'elle permettait la synthèse d'une capsule de type 37. Le mécanisme de biosynthèse assuré par la synthase Tts pourrait être identique à celui permettant la biosynthèse de la capsule de type 3 chez *S. pneumoniae* et

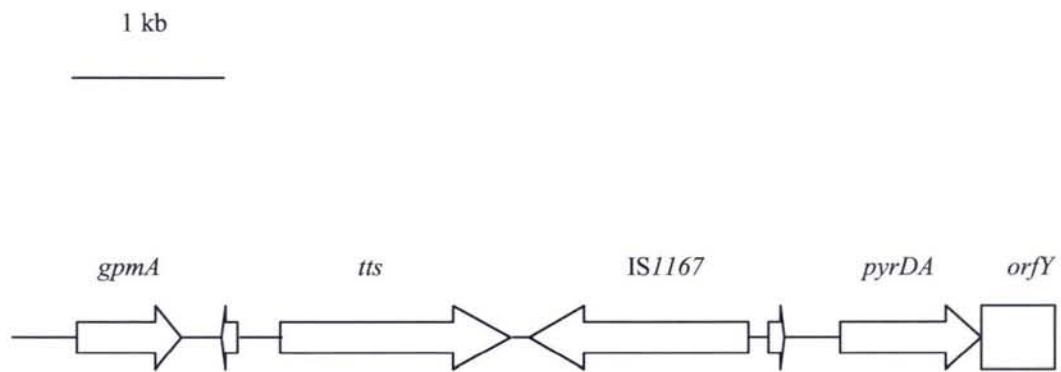


Figure 16 : Organisation génétique du fragment d'ADN génomique de *S. pneumoniae* de type 37 qui confère le caractère producteur d'une capsule de type 37, d'après Llull *et al.* (1999).

 : élément répété de 105 pb

des molécules d'acide hyaluronique chez *S. pyogenes* (CF paragraphe 1.4.2) (Llull *et al.*, 2001).

2 POLYMORPHISME INTRASPECIFIQUE DE LA BACTERIE LACTIQUE *STREPTOCOCCUS THERMOPHILUS*

2.1 Le groupe polyphylétique des bactéries lactiques

Des bactéries de genres différents sont regroupées sous le terme de « bactéries lactiques » et appartiennent à la division des Firmicutes. Elles présentent une coloration de Gram positive, et sont asporulées et anaérobies aérotoles. Leur chromosome se caractérise par un faible pourcentage en G+C. Les bactéries lactiques sont également dépourvues de catalase, de nitrate réductase et de cytochrome oxydase. Elles produisent de l'acide lactique selon un métabolisme homofermentaire ou hétérofermentaire. Dans le premier cas, le catabolisme du glucose aboutit à la formation de deux molécules d'acide lactique alors que dans le second, il s'accompagne de la production d'une molécule d'acide lactique, de dioxyde de carbone et d'éthanol ou d'acide acétique.

Les bactéries lactiques regroupent des bactéries pathogènes et des bactéries utilisées dans l'industrie agro-alimentaire, qualifiées en général de ferments lactiques. Les premières se retrouvent essentiellement au sein du genre *Streptococcus* comme *S. pneumoniae* et *S. pyogenes*. Certains hôtes normaux de la cavité buccale (bactéries du genre *Lactobacillus* et *Streptococcus salivarius*) et de l'intestin (bactéries des genres *Enterococcus* et *Lactobacillus*) peuvent avoir exceptionnellement un caractère pathogène opportuniste quand elles pénètrent dans le système circulatoire. Les bactéries non pathogènes pour l'homme sont qualifiées de « GRAS » : (*Generally Recognised As Safe species*) et correspondent à de nombreuses espèces de bactéries lactiques comme *S. thermophilus*, *L. lactis*, *Lb. delbrueckii*, *Oenococcus oeni*, *Streptococcus macedonicus*. Elles sont utilisées pour leurs propriétés fermentaires dans l'industrie agro-alimentaire et se retrouvent dans des aliments variés comme la charcuterie, le vin, ou la choucroute par exemple. Ces bactéries lactiques sont également présentes dans les produits laitiers fermentés.

2.2 *S. thermophilus*, espèce non pathogène du genre *Streptococcus*

Les cellules de *S. thermophilus* sont de forme ovoïde et sont groupées en chaînettes (Figure 17). Des études menées au laboratoire ont permis de montrer que le nombre de cellules par chaînette est spécifique de la souche et est directement lié à l'allèle du gène *cse* (Borges *et al.*, 2005). Ces cellules présentent une température optimale de croissance de 42°C avec un métabolisme homofermentaire.

S. thermophilus est une des deux espèces appartenant au genre *Streptococcus* à être utilisée dans l'industrie agro-alimentaire, l'autre étant *S. macedonicus* (Tsakalidou *et al.*, 1998). Dans l'industrie, *S. thermophilus* est utilisé en co-culture avec *Lb. delbrueckii* subsp *bulgaricus* pour la fabrication des yaourts et est associé à d'autres espèces de bactéries lactiques dont *L. lactis* pour la fabrication de certains fromages.

La présence de *S. thermophilus* n'a été observée que dans le lait ou les produits laitiers. Les études phylogénétiques effectuées au sein des Streptocoques révèlent une très grande proximité de cette espèce avec deux Streptocoques oraux, commensaux de l'homme, *S. salivarius* et *Streptococcus vestibularis*. *S. thermophilus* serait en fait issu d'un clone de *S. salivarius* qui se serait adapté au milieu lait après l'apparition de la fabrication des produits laitiers (Hols *et al.*, 2005).

Le chromosome de *S. thermophilus* est circulaire et présente une taille comprise entre 1,8 et 1,9 Mb (Roussel *et al.*, 1994) (Roussel *et al.*, 1997). Son contenu en G+C est compris entre 37,2 et 39,8% (Farrow & Collins, 1984). Les souches de *S. thermophilus* analysées possèdent rarement des plasmides et quand ils sont présents, ils sont petits et cryptiques (Mercenier, 1990). Les génomes des souches CNRZ1066 (numéro d'accès : NC_006449) et LMG18311 (numéro d'accès : NC_006448) ont été séquencés (Bolotin *et al.*, 2004). L'analyse de ces séquences révèle l'absence de gènes de virulence. Par ailleurs, 10% des gènes des deux souches sont des pseudogènes, dont les gènes impliqués dans le métabolisme du carbone. Par ailleurs, 185 sur 1915 et 166 sur 1890 ORF codant une protéine potentielle de

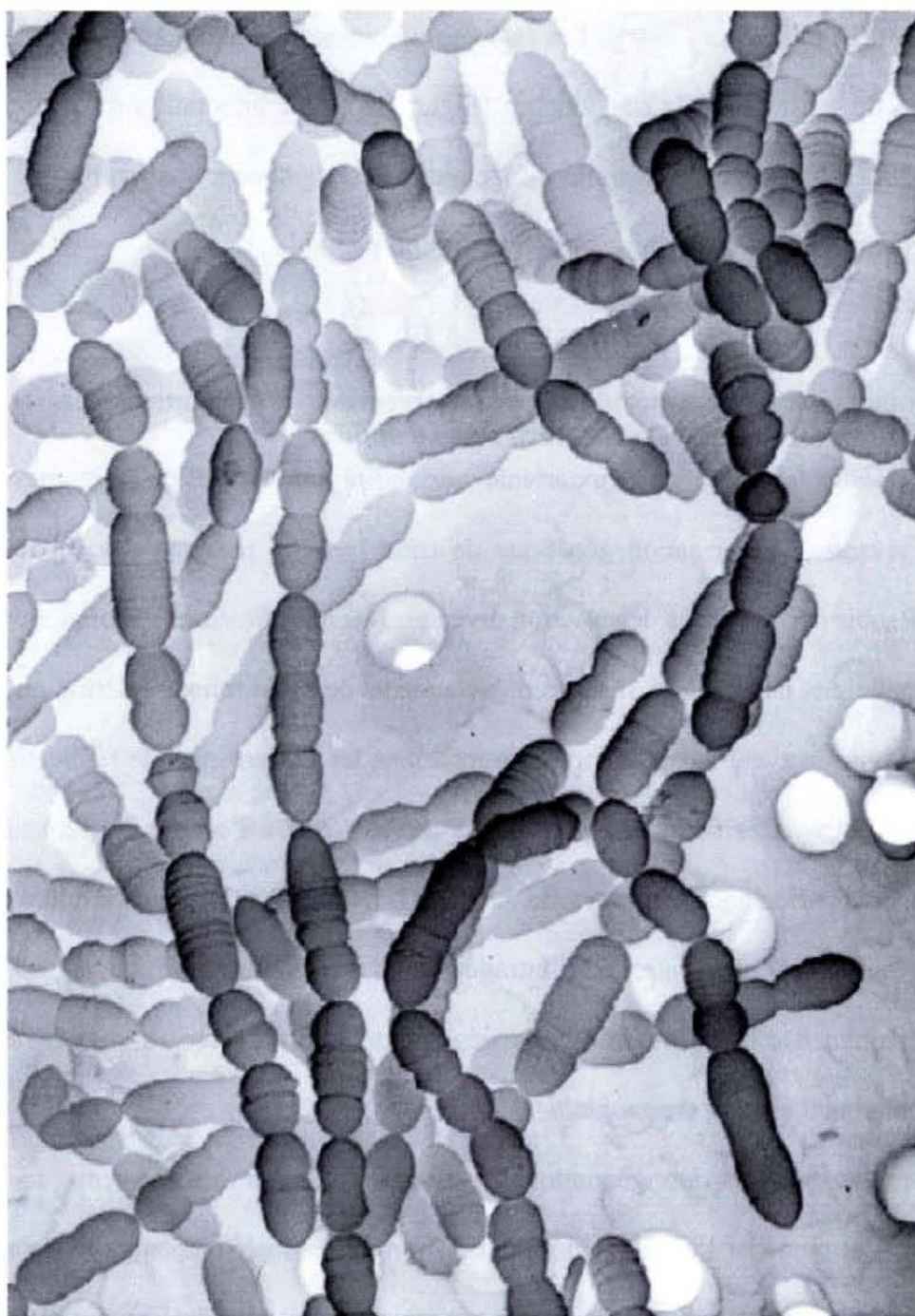


Figure 17 : Photographie de cellules de *S. thermophilus* CNRZ368 obtenue par microscopie électronique à balayage (A. Fernandez, com. pers.).
Les cellules sont cultivées en milieu TPPY jusqu'en phase stationnaire de croissance, puis traitées selon Luft (1961).

fonction inconnue sont identifiés dans le génome des souches CNRZ1066 et LMG18311 respectivement (Bolotin *et al.*, 2004). Les données obtenues permettent également de décrire l'organisation du cluster *eps* de ces souches (Figure 18). Une annotation du génome de la souche LMD-9, effectuée à l'université du Nebraska, est également disponible sur Internet (http://genome.jgi-psf.org/draft_microbes/strth/strth.home.html).

2.3 Polymorphisme intraspécifique et intervention d'événements de transferts horizontaux dans l'évolution de *S. thermophilus*

S. thermophilus présente une importante variabilité intra-spécifique comme tous les organismes vivants. L'information génétique de cette bactérie présente des capacités à se modifier et à subir des pressions de sélection diverses. Les modifications génomiques peuvent résulter de mutations ponctuelles. Il s'agit d'événements de substitution, délétion ou addition de nucléotides qui peuvent survenir lors d'erreurs dans les mécanismes de réplication ou de réparation de l'ADN. Ces mutations ponctuelles peuvent être à l'origine de l'apparition de codons stop et de décalages du cadre de lecture. Les modifications génomiques peuvent également résulter de réarrangements intragénomiques et intergénomiques, associés au transfert horizontal. De telles modifications génomiques ont été décrites chez les bactéries lactiques, notamment chez *S. thermophilus*.

Les réarrangements intragénomiques peuvent impliquer les éléments IS dont la transposition peut modifier l'expression de gènes, soit en les interrompant, soit en apportant ou en permettant la création d'un promoteur. Leur transposition peut également provoquer des délétions ou des fusions de réplicons. En particulier, la transposition répllicative d'*ISSI* permet l'intégration aléatoire du plasmide pGhost9::*ISSI* dans le génome pour générer des mutants insertionnels chez certaines bactéries lactiques (Maguin *et al.*, 1996). Des événements de recombinaison homologue entre différentes copies d'un élément IS peuvent également provoquer la formation de cointégrats, des délétions ou des inversions

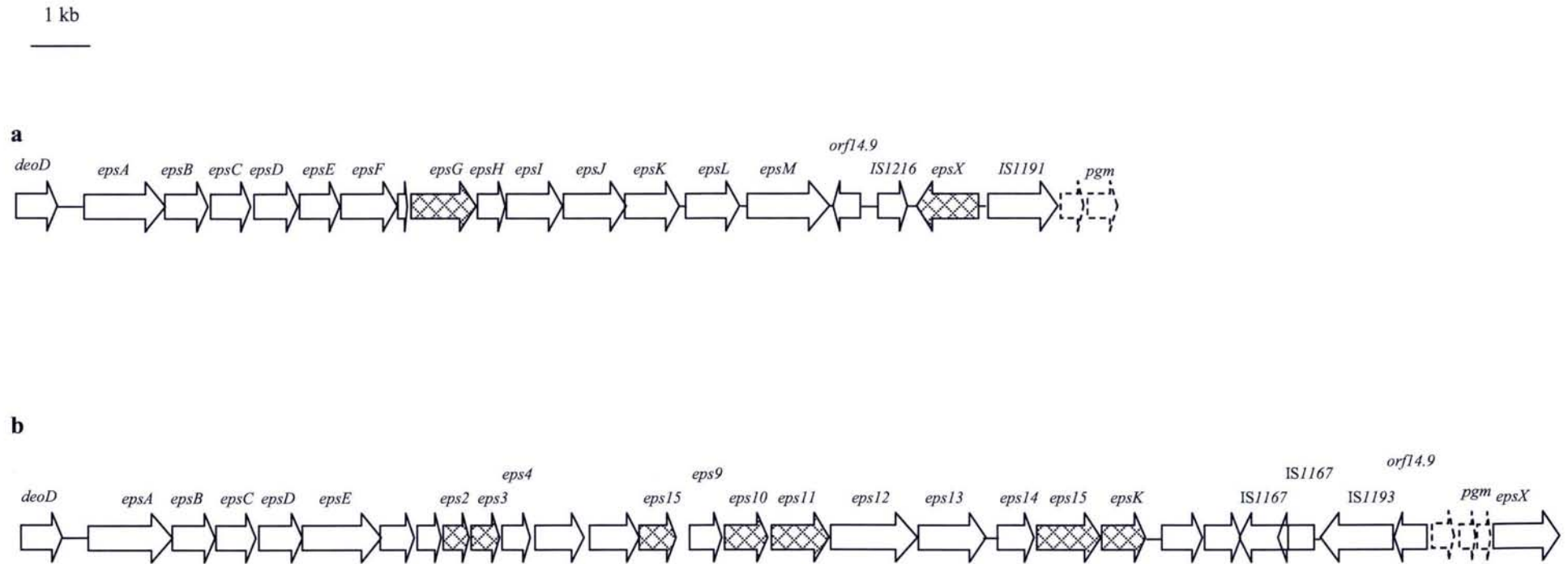


Figure 18 : Organisation du cluster *eps* de *S. thermophilus* CNRZ1066 (a) et LMG18311 (b) d'après Bolotin *et al.* (2004). Les ORF annotées comme codant une glycosyltransférase sont hachurées.

(Galas & Chandler, 1989). Ainsi, un événement de recombinaison homologue entre deux copies d'IS905 a provoqué l'inversion d'un fragment correspondant à la moitié du chromosome chez *L. lactis* subsp. *cremoris* MG1363 (Daveran-Mingot *et al.*, 1998).

Des remaniements intergénomiques associés au transfert horizontal peuvent également être observés. Au sein du pool génétique de leur écosystème, les bactéries peuvent établir des échanges génétiques par transferts horizontaux, ce qui leur permet de s'adapter rapidement aux variations du milieu dans lequel elles évoluent par sélection. L'échange d'information génétique par transfert horizontal peut intervenir entre deux souches d'une même espèce (transfert intraspécifique) ou entre deux souches d'espèces différentes (transfert interspécifique).

La mise en contact de *S. thermophilus* avec d'autres bactéries lactiques, dont *L. lactis*, lors de la fabrication de produits laitiers s'est probablement accompagnée de tels transferts horizontaux. Ainsi, l'étude d'éléments IS chez *S. thermophilus* et *L. lactis* révèle la présence d'éléments ISS1 de séquences presque identiques chez toutes les souches de *L. lactis* étudiées et seulement chez une partie des souches de *S. thermophilus* étudiées. Un transfert d'élément ISS1 se serait donc produit de *L. lactis* vers *S. thermophilus* (Bourgoin *et al.*, 1996). Des éléments ISS1 de séquences presque identiques ont également été identifiés chez *Leuconostoc mesenteroides*, *Lactobacillus sake*, *Enterococcus faecalis*, *Enterococcus faecium* et *Enterococcus hirae*, suggérant que les transferts de tels éléments ne se limitent pas à ces deux espèces (Guedon *et al.*, 2000). De même, la distribution de l'élément IS981 chez *S. thermophilus* et *L. lactis* en fonction de l'origine des souches (isolées du fromage ou du yaourt) suggère que cet élément IS aurait été transféré de *L. lactis* vers *S. thermophilus* lors de la fabrication de fromages. A l'inverse, des éléments IS1191/IS905 sont retrouvés dans toutes les souches de *S. thermophilus* testées et seulement dans quelques souches de *L. lactis*, ce qui suggère un transfert de *S. thermophilus* vers *L. lactis* (Guedon *et al.*, 1995).

2.4 Mécanismes impliqués dans le transfert horizontal chez *S. thermophilus*

Trois mécanismes permettant l'échange de matériel génétique entre les bactéries sont classiquement distingués : la conjugaison, la transformation et la transduction. Quel que soit le mécanisme, l'acquisition effective de nouvelles séquences nécessite leur maintien dans le génome des bactéries receveuses.

La conjugaison implique l'établissement d'un contact physique entre la cellule donneuse et la cellule receveuse grâce à un pilus. Puis, un élément conjugatif de type plasmide, transposon ou élément intégratif conjugatif (ICE) est transféré vers la cellule receveuse. Lors de la mobilisation de l'élément conjugatif, d'autres séquences peuvent être transférées. Une des équipes du laboratoire s'intéresse particulièrement à la conjugaison chez *S. thermophilus*. Chez *S. thermophilus* CNRZ368, un élément intégratif potentiellement conjugatif, *ICES_{StI}*, a été identifié (Burrus *et al.*, 2000). L'organisation de cet élément est modulaire. Ainsi, la partie gauche de 16,4 kb contient des ORF dont les fonctions potentielles ne semblent pas associées à la conjugaison alors que la partie droite, de 18,3 kb, contient des ORF codant des protéines homologues à celles de transposons conjugatifs tel que *Tn916* (Burrus *et al.*, 2002a). Il a été montré que l'élément *ICES_{StI}* pouvait s'exciser du chromosome en libérant une molécule circulaire (Burrus *et al.*, 2002b). Les études actuelles tendent à mettre en évidence un transfert par conjugaison de cet élément.

La transformation pourrait aussi être un mécanisme permettant le transfert horizontal chez *S. thermophilus*. Elle suppose la libération d'ADN donneur ainsi que sa dispersion et son maintien dans l'environnement. La fabrication de produits laitiers fermentés peut s'accompagner de lyses cellulaires importantes (de Ruyter *et al.*, 1997), permettant la libération de matériel génétique dans le milieu. Cependant, son maintien dans le milieu nécessite que les nucléases soient présentes en faible concentration. De plus, l'entrée d'ADN dans une cellule réceptrice nécessite que celle-ci présente un état de compétence. La

transformation est particulièrement efficace chez *S. pneumoniae*, résultat établi depuis les travaux historiques de Griffith en 1928.

Le troisième mécanisme pouvant intervenir dans le transfert horizontal pourrait être la transduction, qui implique la présence de bactériophages. La transduction généralisée s'accompagne du transfert de n'importe quelle région du génome d'une cellule lysée par un phage. Les particules phagiques générées portent alors soit l'ADN du phage soit un fragment de l'ADN bactérien. La transduction spécialisée s'accompagne seulement du transfert des séquences localisées à proximité du site d'intégration du prophage. L'ADN contenu par les particules phagiques générées provient du phage et de la cellule lysée. Les phages peuvent être à l'origine de la destruction de cultures bactériennes lors de la fabrication des produits laitiers ou être intégrés à l'état de prophage dans le génome des souches utilisées. L'analyse des séquences de *S. thermophilus* CNRZ1066 et LMG18311 a permis de mettre en évidence la présence d'un prophage chez CNRZ1066 et l'absence de tels éléments chez LMG18311 (Bolotin *et al.*, 2004). Le séquençage du génome du phage ϕ Sfi21 isolé de *S. thermophilus* révèle que plusieurs protéines potentielles codées par ce phage présentent un degré d'homologie significatif avec des protéines codées par des phages de *L. lactis* (Brussow *et al.*, 1998). Ces données suggèrent que les phages de *S. thermophilus* pourraient évoluer par transfert horizontal de gènes avec des phages de bactéries présentes dans la même niche écologique (Brussow *et al.*, 1998).

Pour que le transfert de séquence puisse être observé, la séquence acquise devra ensuite être transmise aux cellules filles de génération en génération. Le maintien des séquences acquises constitue donc une étape importante du transfert horizontal. Ces séquences se maintiendront soit par répllication autonome (dans le cas des plasmides) soit par intégration dans le génome de la cellule réceptrice par recombinaison homologue ou illégitime. A plus long terme, la sélection naturelle joue un rôle clef dans le maintien des séquences transférées.

La probabilité de maintien, au sein de la population, de séquences pouvant conférer un avantage sélectif dans certaines conditions environnementales sera plus grande que des séquences ne conférant pas d'avantage sélectif.

2.5 Implication du transfert horizontal dans la constitution des clusters *eps* chez *S. thermophilus*

Les clusters *eps* de *S. thermophilus* étant très variables, ils constituent de bons candidats pour l'étude de l'intervention du transfert horizontal dans leur constitution. Des traces potentielles de tels événements peuvent être retrouvées au sein des clusters *eps*. Par exemple, les clusters *eps* des souches IP6757 et NST2280 de *S. thermophilus* sont très différents ; alors que l'arbre phylogénétique de différentes souches de *S. thermophilus* montre que ces deux souches sont très proches phylogénétiquement (Figure 1 du préambule). Ces résultats indiquent une évolution rapide du cluster *eps* de la souche IP6757 qui pourrait résulter d'un remplacement de séquence (Pluvinet *et al.*, 2004). Un tel mécanisme s'accompagne d'une évolution rapide du cluster *eps*. L'intervention d'un phénomène comparable a été proposé pour expliquer l'observation de variants de la souche *S. pneumoniae* type 23F dont la seule différence avec la souche parentale réside dans leurs clusters *cps* qui confère le sérotype 19F (Coffey *et al.*, 1998).

D'autres traces potentielles d'événements de transferts horizontaux peuvent également être observées au sein des clusters *eps* de *S. thermophilus*. Ainsi, le pourcentage en G+C du cluster *eps* de *S. thermophilus* CNRZ368 varie de 26,2 à 41,9% selon les ORF (Bourgoin *et al.*, 1999), alors que celui du génome total est compris entre 37,2 et 39,8% (Farrow & Collins, 1984). Cette variation du contenu en G+C se retrouve au sein de clusters d'autres souches de *S. thermophilus*. Ceci suggère que plusieurs gènes de ce cluster auraient été acquis par transfert horizontal à partir d'autres espèces bactériennes. En particulier, une région de 13,6 kb contenant plusieurs copies d'*IS981* et d'*ISS1* et deux ORF quasiment identiques à *epsL* et

orfY du cluster *eps* de *Lactococcus lactis* NIZOB40 aurait été transférée entre *S. thermophilus* et *L. lactis* (Bourgoin *et al.*, 1999).

L'étude des clusters *eps/cps* suggère l'intervention d'événements de transferts horizontaux, mais le ou les mécanisme(s) ayant permis le transfert des séquences concernées n'est (ne sont) pas connu(s). Le maintien des séquences ainsi transférées suggère qu'elles confèrent un avantage sélectif à la bactérie receveuse, ce qui amène à s'interroger sur le rôle des EPS chez *S. thermophilus*.

RESULTATS

Notre étude s'est focalisée sur le cluster *eps* de la souche IP6756 de *S. thermophilus*. Les gènes *epsC* et *epsD* du cluster *eps* constamment présents dans les autres souches étudiées (Broadbent *et al.*, 2003) sont absents du cluster *eps* de cette souche. Rappelons que *epsC* et *epsD* appartiennent au premier module, le module de régulation, des clusters *eps* de *S. thermophilus* qui est très conservé (Jolly & Stinglele, 2001). Ainsi, le premier module du cluster *eps* de la souche IP6756 ne contiendrait que les deux premières ORF *epsA* et *epsB*.

L'étude de la structure du cluster *eps* de cette souche a été réalisée et le caractère producteur d'EPS de la souche IP6756 a été recherché. Cette étude permet d'aborder de nombreuses questions sur la fonctionnalité et plus généralement sur l'organisation des clusters *eps* chez *S. thermophilus*. Les résultats présentés ont permis la rédaction d'un article sous presse dans *International Dairy Journal* (Annexe 1). Cette étude permet notamment de s'interroger sur la fonctionnalité d'*orf14.9*, une ORF présente dans tous les clusters *eps* séquencés de *S. thermophilus*. Elle est localisée en 3' et orientée en sens inverse des gènes organisés en opéron et impliqués dans la biosynthèse des EPS. Les résultats présentés ont permis la rédaction d'un article soumis à *Archives of Microbiology* (Annexe 2).

1 SPECIFICITES STRUCTURALES DU CLUSTER *EPS* DE LA SOUCHE IP6756

1.1 Frontières du cluster *eps* de la souche IP6756

Le cluster *eps* de la souche IP6756 a été amplifié par LR-PCR en utilisant des amorces spécifiques des régions conservées qui jouxtent les clusters *eps* chez *S. thermophilus* (Stingle et al., 1996a) (Pluvinet et al., 2004). La première amorce est spécifique de la région intergénique localisée entre les gènes *deoD* et *epsA* et la seconde de l'ORF *pgm* (Figure 19). Le séquençage du produit obtenu a permis d'obtenir une séquence de 8119 pb. Elle comprend les séquences localisées entre la région intergénique située entre les gènes *deoD* et *epsA* et *pgm* (Figure 19). L'ORF *deoD* constituerait ainsi la frontière gauche du cluster *eps* de la souche, comme c'est le cas dans les autres clusters *eps* de *S. thermophilus* (CF paragraphe 1.5.1.2 de l'introduction). Par ailleurs, l'hybridation de sondes spécifiques de *deoD* et d'*epsA* de *S. thermophilus* IP6757 révèle un fragment *EcoRI* de 6,4 kb de l'ADN génomique de la souche IP6756, mais également de 25 autres souches de *S. thermophilus* (Pluvinet, 2003). Ces résultats indiquent que *deoD* et *epsA* sont localisées dans la même région du chromosome chez l'ensemble des souches testées et en particulier chez IP6756. De même, la frontière droite du cluster *eps* de la souche IP6756 n'est pas différente de celles des autres clusters et correspondrait à *pgm*.

L'analyse de la séquence obtenue révèle la présence d'un site potentiel de fixation du ribosome (RBS) localisé en amont des quatre ORF identifiées : *epsA*, *epsB*, *epsZ* et *epsJ*, ainsi qu'en amont d'*orf14.9*, de *pgm* et de l'ORF codant la transposase de la séquence d'insertion $\Delta IS1193$. Elle révèle également l'organisation originale du cluster *eps* de la souche IP6756. De plus, le gène *orf14.9* est présent au sein du cluster comme c'est le cas pour tous les autres clusters *eps* analysés chez *S. thermophilus*.

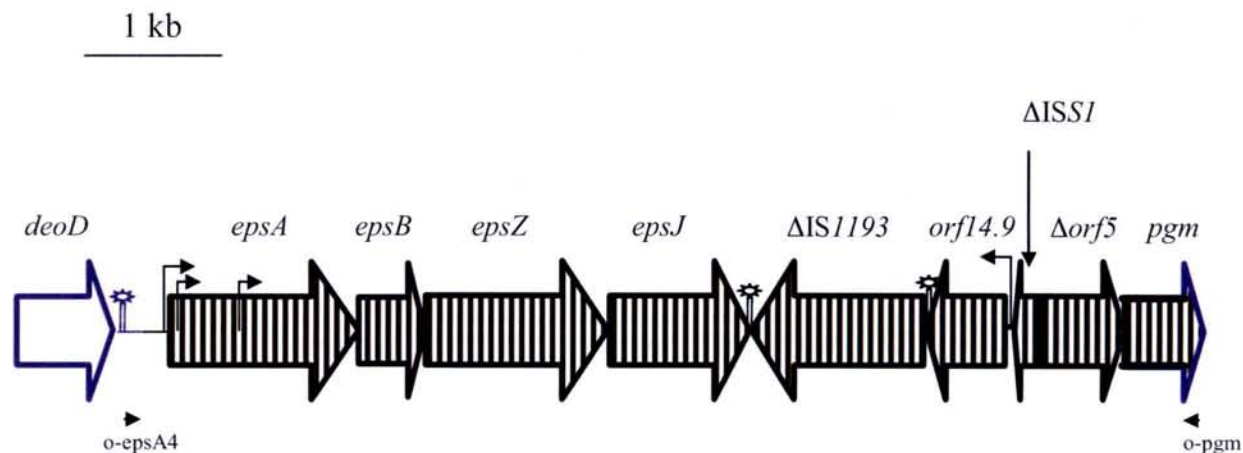


Figure 19 : Organisation du cluster *eps* de *S. thermophilus* IP6756.

La séquence de ce cluster est référencée sous le numéro d'accèsion AY916787 dans la banque de données GenBank/EMBL.

Les promoteurs potentiels de transcription sont indiqués par une flèche.

Aucun promoteur potentiel de transcription n'a été détecté en amont de $\Delta orf5$ ni de *pgm*.

Les terminateurs potentiels de transcription sont symbolisés par $\star$.

Les séquences figurées en bleu sont hypothétiques.

La localisation des amorces utilisées lors de l'amplification du cluster *eps* par LR-PCR est indiquée par : 5' 3'

o-epsA4 : GGAAGAAGAAGACTTATAATA

o-pgm : CCTGAATTCGACTGAAAGGCAC

1.2 Organisation modulaire particulière de la première région du cluster *eps* de la souche IP6756

La première région du cluster *eps* de *S. thermophilus* englobe les ORF situées entre *deoD* et le terminateur d'*orf14.9*. Cette première région présente une organisation modulaire : un premier module impliqué dans la régulation de la production d'EPS et un second module qui contient les ORF codant les enzymes nécessaires à la biosynthèse et à l'export des EPS (CF paragraphes 1.5.1.1 et 1.5.1.2 de l'introduction).

La séquence de 6051 pb située en 3' d'*orf14.9* contient les quatre ORF *epsA*, *epsB*, *epsZ* et *epsJ* ainsi qu'une séquence d'insertion $\Delta IS1193$ (Figure 19). La séquence de $\Delta IS1193$ présente 94% d'identité avec la séquence d'*IS1193* du cluster *eps* des souches Sfi39 et MTC310. Elle comporte des répétitions inversées terminales (IRR et IRL) de 24 pb, mais elle présente une délétion de 64 pb dans la partie 5' de l'ORF codant la transposase de cet élément (Figure 20). Cette délétion génère un décalage du cadre de lecture de cette ORF entraînant l'apparition de codons stop. La copie d'*IS1193* de la souche IP6756 serait donc défective pour la transposition.

Cette région présente une organisation modulaire mais elle se révèle très différente de celles décrites dans les autres clusters *eps* de *S. thermophilus*. Ainsi, le premier module du cluster *eps* de IP6756 qui correspond au module de régulation, n'est constitué que des ORF *epsA* et *epsB*. La séquence de l'allèle *epsA* de la souche IP6756 est très conservée et présente 95 à 98% d'identité avec celle des allèles des vingt deux autres clusters *eps* séquencés de *S. thermophilus* (Tableau1). La fonction de régulateur transcriptionnel de son produit potentiel serait ainsi conservée. La seconde de ces ORF, *epsB*, présente une taille réduite (489 pb) par rapport à celle des allèles des autres clusters *eps* (732 pb). Seuls les 465 premiers nucléotides d'*epsB* présentent au moins 96% d'identité avec les séquences correspondantes des vingt deux autres clusters *eps* de *S. thermophilus*, les 24 nucléotides restant ne présentant

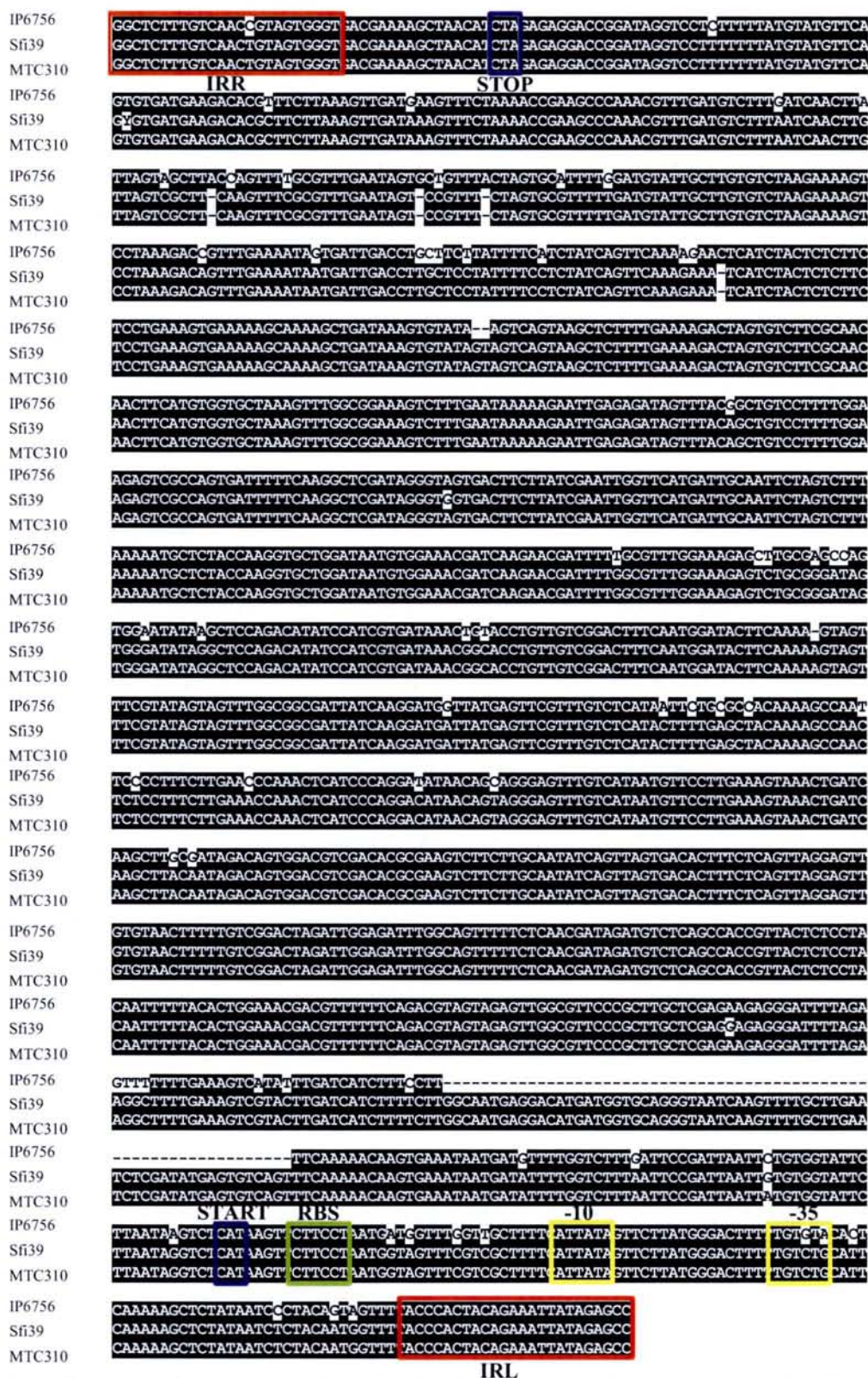


Figure 20 : Alignement des séquences d'IS1193 des clusters *eps* de *S. thermophilus* IP6756, MTC310 et Sfi39.

Les codons d'initiation (START) et d'arrêt (STOP) de traduction de la transposase sont encadrés en bleu. Les boîtes promotrices -10 et -35 sont encadrées en jaune et le RBS de la transposase est encadré en vert. IRR et IRL correspondent respectivement aux répétitions terminales inversées droite et gauche d'IS1193. Les séquences noircies correspondent à des séquences identiques. Les tirets correspondent à des décalages introduits lors de l'alignement.

Tableau 1 : clusters *eps* séquencés chez *S. thermophilus*

Souche ou type de cluster <i>eps</i>	Numéro d'accession	Référence
NST2280	AF517575	Pluvinet <i>et al.</i> 2004
CNRZ1066	NC_006449	Bolotin <i>et al.</i> 2004
CNRZ368	Z98171	Bourgoin <i>et al.</i> 1999
IP6756	AY916787	ce travail
IP6757	AJ289861	Pluvinet <i>et al.</i> 2004
LMG18311	NC_006448	Bolotin <i>et al.</i> 2004
MR-1C	AF448249	Broadbent <i>et al.</i> 2003
MR-2C	AY061649	Broadbent <i>et al.</i> 2003
MTC310	AF448502	Broadbent <i>et al.</i> 2003
MTC330	AF430847	Welker <i>et al.</i> ^b
MTC360	AF434993	Broadbent <i>et al.</i> 2003
NCFB2393	Y17900	Almiron-Roig <i>et al.</i> 2000
Sfi39	AF373595	Germond <i>et al.</i> 2001
Sfi6	U40830	Stingele <i>et al.</i> 1996
type I ^a	AF410175	Rallu <i>et al.</i> ^b
type III ^a	AY057915	Rallu <i>et al.</i> ^b
type IV ^a	AF454495	Rallu <i>et al.</i> ^b
type V ^a	AF454496	Rallu <i>et al.</i> ^b
type VI ^a	AF454497	Rallu <i>et al.</i> ^b
type VII ^a	AF454498	Rallu <i>et al.</i> ^b
type IX ^a	AF454499	Rallu <i>et al.</i> ^b
type X ^a	AF454500	Rallu <i>et al.</i> ^b
type XI ^a	AF454501	Rallu <i>et al.</i> ^b

^a : Les souches dans lesquelles ces clusters ont été caractérisés ne sont pas précisées.

^b : Données non publiées

pas d'homologie avec les séquences des banques de données. Comme l'indiquaient les résultats d'hybridation, *epsC* et *epsD* ne sont pas présents dans le module de régulation du cluster. Cette organisation suggère que ce module de régulation aurait un fonctionnement différent de celui des autres clusters *eps* de *S. thermophilus*, qui serait similaire à celui du cluster *cps* de *S. pneumoniae*. Chez cette bactérie, le taux de polymérisation des CPS serait corrélé à la phosphorylation du complexe formé par les protéines CpsC et CpsD (homologues respectivement des protéines EpsC et EpsD de *S. thermophilus*). La protéine EpsD pourrait s'autophosphoryler à partir des molécules d'ATP fixées ce qui se provoquerait sa dissociation de la protéine EpsC et une réduction à un taux basal de la production de CPS. La protéine CpsB (homologue de la protéine EpsB) qui posséderait une activité phosphatase interviendrait en déphosphorylant la protéine CpsD ce qui se permettrait la reformation du complexe protéique CpsC-CpsD et donc une augmentation du taux de production des CPS. La protéine CpsB pourrait également inhiber l'autophosphorylation de la protéine CpsD. Elle joue donc un rôle important dans la régulation de la biosynthèse des EPS (CF paragraphe 1.6.1 de l'introduction).

La séquence d'*epsZ* présente 98 et 97% d'identité avec respectivement celles des ORF *eps10L* (cluster de type X) et *eps6L* (cluster de type VI) (Tableau 1). Ces deux ORF présentent une taille identique à celle d'*epsZ*. La séquence d'*epsZ*, à l'exception des 86 premiers nucléotides de cette séquence, présente 98 % d'identité avec les 1195 pb de l'extrémité 3' du pseudogène *rut* (*repeating unit transporter*) du cluster *eps* de la souche MTC310. A partir du nucléotide 87, la séquence d'*epsZ* présente également 98% d'identité avec une région du cluster *eps* de la souche Sfi39 englobant une séquence intergénique de 207 pb puis une pseudo-ORF (nommée *epsI**) et une séquence intergénique de 148 pb. De plus, la protéine potentielle EpsZ, de 471 acides aminés, présente 74% d'identité sur toute sa longueur avec les protéines Cps33L et Cps37L de *S. pneumoniae* type 33 et type 37 respectivement

(Llull *et al.*, 1998) (Llull *et al.*, 1999). Ces différentes protéines potentielles ayant une fonction de transporteur d'unités répétées, cette fonction pourrait être conservée par le produit putatif de l'ORF *epsZ*. La séquence d'*epsJ* présente 98% d'identité avec celle d'*epsJ* du cluster *eps* des souches Sfi39 et MTC310. Par ailleurs, la protéine potentielle EpsJ de la souche IP6756 (de 365 acides aminés) présente 94% d'identité avec la protéine potentielle Cps33N de *S. pneumoniae* de type 33. La fonction d'UDP-galactopyranose mutase attribuée aux protéines potentielles EpsJ de Sfi39 et MTC310 et à la protéine potentielle Cps33N de *S. pneumoniae* de type 33 pourrait ainsi être conservée chez la souche IP6756. Les séquences homologues à *epsZ* et *epsJ* appartenant au second module de leurs clusters *eps* respectifs, le second module de la première région du cluster *eps* de la souche IP6756 n'est donc constitué que par les ORF *epsZ* et *epsJ* (Figure 19). Ce module correspond au module de biosynthèse et d'export des EPS.

Outre sa petite taille comparée à celle des vingt deux autres clusters *eps* de *S. thermophilus*, ce second module présente une autre particularité : il ne contient pas d'ORF codant une glycosyltransférase potentielle.

1.3 La seconde région du cluster *eps* de la souche IP6756 contient *orf14.9* ainsi que d'autres ORF situées en 5' d'*orf14.9*

La séquence de 2068 pb restant à analyser contient *orf14.9*, deux ORF partiellement délétées $\Delta orf5$ et $\Delta ISS1$ ainsi que l'extrémité 5' de l'ORF *pgm*, qui constitue la bordure « droite » de la séquence obtenue et du cluster *eps* (Figure 19). Cette séquence de 2068 pb présente 98% et 97% d'identité avec la région similaire du cluster *eps* des souches IP6757 et MTC310. La juxtaposition de $\Delta ISS1$ et $\Delta orf5$ résulterait d'un événement de recombinaison site-spécifique entre les séquences initiales de ces deux éléments, comme dans le cluster *eps* de la souche IP6757 (Pluvinet, 2003). *pgm* peut être considérée comme une pseudo-ORF, puisque la présence de plusieurs mutations frameshift ne permettra pas la synthèse d'une

protéine fonctionnelle. Néanmoins, la séquence protéique obtenue avec le programme BlastX présente des identités significatives avec des phosphoglycérate mutases, notamment 53% d'identité avec une protéine à activité phosphoglycérate mutase de *S. mutans* UA159. L'interrogation des séquences nucléotidiques des banques de données avec la séquence nucléotidique de *pgm* de la souche IP6756 révèle la présence d'une séquence présentant un pourcentage d'identité compris entre 97 et 100% à l'extrémité droite des clusters *eps* des souches MTC310, MTC 330, MTC360, LMG18311, MR-1C, MR-2C, CNRZ368 et CNRZ1066. Dans ces clusters, *pgm* présente également un caractère de pseudo-ORF.

Dans la seconde région du cluster *eps* de la souche IP6756, *orf14.9* est la seule ORF à présenter un caractère potentiellement fonctionnel. Cette ORF présente un contenu en bases G et C de 34,2% et est orientée en sens inverse par rapport aux gènes impliqués dans la biosynthèse des EPS situés en 3' d'*orf14.9*. Elle présente une séquence très conservée au sein des clusters *eps* de *S. thermophilus*. Ainsi, la séquence de l'allèle *orf14.9* de la souche IP6756 présente 96 à 100% d'identité avec les allèles des autres souches. De plus, un produit de PCR de 451 pb interne à *orf14.9* a été obtenu (à l'aide des amorces o-*orf14.9*-2 et o-*orf14.9*-4) puis séquencé chez treize souches de *S. thermophilus* pour lesquelles aucune séquence n'est disponible (A054, ATCC19258^T, ATCC19987, CNRZ302, CNRZ307, CNRZ308, CNRZ385, CNRZ388, CNRZ391, CNRZ455, CNRZ702, CNRZ1402, NST1). Une séquence interne à *orf14.9* de 260 pb a pu ainsi être comparée entre ces treize souches, ce qui a permis de relever un pourcentage d'identité similaire. La comparaison du degré de conservation de la séquence d'*orf14.9* à celui des gènes de ménage permet de déterminer sa significativité biologique. Les comparaisons des séquences des gènes de ménage *lacS* et *lacZ* de 29 souches de *S. thermophilus* révèlent une variabilité comparable (Ercolini *et al.*, 2005). Ces gènes codent respectivement la perméase permettant l'entrée du lactose dans la cellule et la β -galactosidase permettant l'hydrolyse du lactose. Comme la séquence *orf14.9* présente une conservation

similaire à celle de gènes de ménage du chromosome de *S. thermophilus*, cette ORF conférerait un avantage sélectif à *S. thermophilus*.

1.4 Comparaison de l'ensemble du cluster *eps* de la souche IP6756 avec les autres clusters *eps* de *S. thermophilus*

L'alignement de la séquence du cluster *eps* de la souche IP6756 avec celle des souches MTC310 et Sfi39 qui présentent 99% d'identité mais sont annotés différemment (les clusters *eps* de ces souches diffèrent notamment au niveau de l'extrémité 3' de la première région) montre que le cluster *eps* de la souche IP6756 peut être divisé en trois régions (Figure 21). Les frontières entre ces différentes régions sont précisées dans la figure 22. La première région comporte les premières 2091 pb et inclut les 165 nucléotides correspondant à la région intergénique *deoD-epsA*, *epsA* et les 465 premiers nucléotides d'*epsB* (Figure 21). Elle présente 98% d'identité avec les séquences correspondantes des clusters *eps* de MTC310 et Sfi39. Dans le cluster *eps* des souches Sfi39 et MTC310, la région intergénique *deoD-epsA* est interrompue par une séquence d'insertion IS905 qui ne présente aucune homologie avec la séquence de la région intergénique *deoD-epsA* obtenue dans la souche IP6756. Par ailleurs, la première région présente un contenu en bases G et C de 36,2 %.

La seconde région présente une taille de 125 pb et englobe les 24 derniers nucléotides d'*epsB*, une région intergénique de 15 pb ainsi que les 86 pb de l'extrémité 5' d'*epsZ* (Figure 21). Cette région ne présente pas d'homologie avec le cluster *eps* des souches Sfi39 et MTC310. Seuls les 86 premiers nucléotides d'*epsZ* présentent 98 et 97 % d'identité avec les séquences correspondantes d'*eps10L* et *eps6L*. Le contenu en bases G et C de cette région est de 32%.

La troisième région correspond aux 5903 nucléotides restants. Elle contient une séquence comprenant les 1330 pb de l'extrémité 3' d'*epsZ*, *epsJ*, *orf14.9*, Δ *orf5* et *pgm* ainsi que la

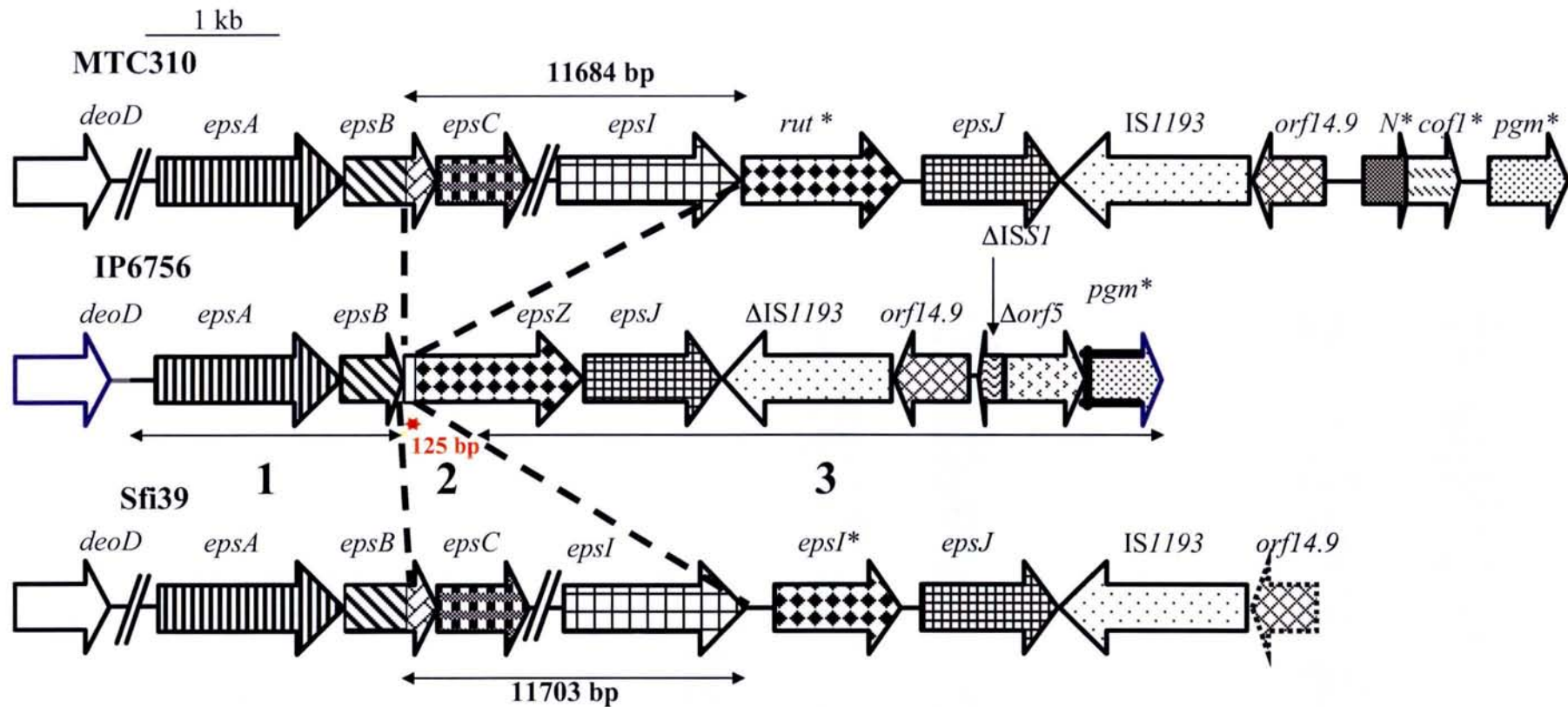


Figure 21 : Comparaison des clusters *eps* de *S. thermophilus* MTC310, IP6756 et Sfi39.

Chaque ORF est représentée avec un symbole différent. Les séquences des régions 1 et 3 de chaque cluster *eps* présentent plus de 98% d'identité. Les 11 premiers nucléotides du pseudogène *rut* (cluster *eps* de la souche MTC310) ne présentent pas d'homologie avec la séquence d'*epsZ* (cluster *eps* de la souche IP6756).

La localisation de la région 2 de chaque cluster est précisée (rectangle blanc pour IP6756 et flèche pour les deux autres souches). Les 3 régions définies dans le cluster *eps* de la souche IP6756 sont indiquées.

L'astérisque signale des pseudo-ORF. Les séquences figurées en bleu sont hypothétiques. La région intergénique entre *deoD* et *epsA* est interrompue dans le cluster *eps* des souches MTC310 et Sfi39 par une séquence d'insertion IS905. La séquence disponible pour le cluster *eps* de la souche Sfi39 ne contient qu'une séquence partielle d'*orf14.9*.

// : indique que la séquence n'est pas représentée.

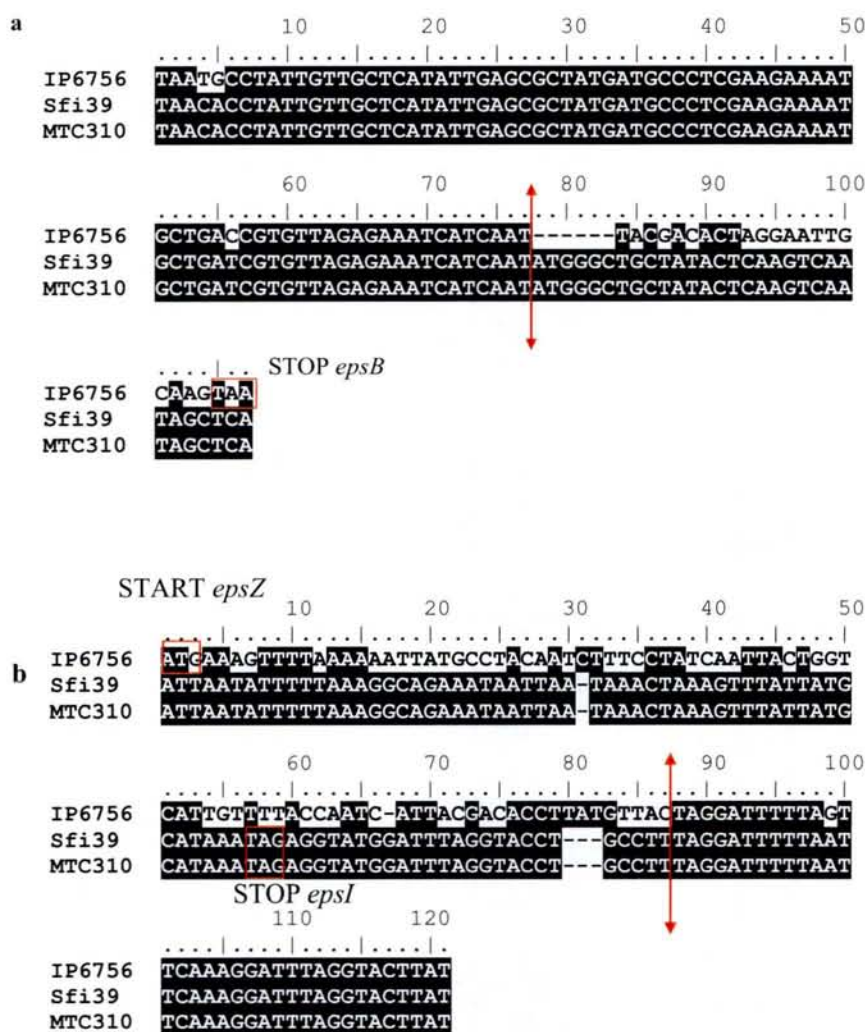


Figure 22 : Frontières des régions d'homologie entre les clusters *eps* de *S. thermophilus* IP6756, Sfi39 et MTC310.

a : Localisation de la frontière (flèche rouge) entre les régions 1 et 2.

Le codon STOP de traduction d'*epsB* d'IP6756 est indiqué, la séquence d'*epsB* étant plus longue chez Sfi39 et MTC310.

b : Localisation de la frontière (flèche rouge) entre les régions 2 et 3.

Le codon START de traduction d'*epsZ* d'IP6756 est indiqué ainsi que le codon STOP de traduction d'*epsI* de Sfi39 et MTC310. La séquence du pseudogène *rut* commence 13 nt en aval du codon STOP de traduction d'*epsI*. Aucun codon START de traduction ne peut être identifié.

Les tirets correspondent à des décalages introduits lors de l'alignement.

copie de $\Delta IS1193$ et $\Delta ISS1$ (Figure 21) et présente 98% d'identité avec les séquences correspondantes des clusters *eps* des souches SFi39 et MTC310. Le contenu en bases G et C de cette troisième région est de 37,5%. Chez Sfi39, la séquence correspondante contient une séquence intergénique de 207 pb, puis une pseudo-ORF (*epsI**), une séquence intergénique de 148 pb, *epsJ*, *IS1193* et *orf14.9* (Figure 21). Chez MTC310, l'extrémité 5' de cette séquence a été annotée différemment puisqu'elle englobe les 1195 derniers nucléotides du pseudogène *rut*, une séquence intergénique de 146 pb puis *epsJ* (Figure 21).

La comparaison de la séquence du cluster *eps* d'IP6756 au-delà d'*orf14.9* n'est possible qu'avec celle du cluster de la souche MTC310, puisque le séquençage du cluster *eps* de Sfi39 est limité à *orf14.9*. Elle présente 97% et 98% d'identité avec la séquence correspondante du cluster *eps* des souches MTC310, qui est annotée différemment, et IP6757.

Les régions 1 et 3 qui bordent la région 2, spécifique à IP6756, présentent donc la même organisation et la même séquence dans le cluster *eps* des souches IP6756, Sfi39 et MTC310. Le cluster *eps* de la souche IP6756 se caractérise par une ORF mosaïque : *epsZ*. En effet, les 86 premiers nucléotides de la séquence de cette ORF appartiennent à la région 2 de 125 pb, alors que les 1330 nucléotides restant appartiennent à la région 3. La présence simultanée des ORF *epsZ* et *epsJ* a été testée par l'hybridation de sondes spécifiques à ces ORF sur l'ADN génomique de 22 souches de *S. thermophilus* (IP6757, CNRZ7, CNRZ308, CNRZ368, CNRZ385, CNRZ388, CNRZ391, CNRZ404, CNRZ445, CNRZ464, CNRZ702, CNRZ1066, CNRZ1068 et CNRZ1042, LMG18311, ATCC19987, NST1, NST3, NST5, NST10, NST11 et NST12). Seule la souche CNRZ388 présente des signaux d'hybridations avec ces deux sondes, alors que la souche CNRZ445 présente un signal d'hybridation seulement avec la sonde spécifique d'*epsZ*.

Le cas des clusters *eps* de type VI et X (Tableau 1) qui contiennent respectivement les ORF *eps6L* et *eps10L*, de même taille qu'*epsZ* et présentant respectivement 97 et 98 %

d'identité avec la séquence d'*epsZ*, est différent. En effet, l'organisation en 3' de ces deux clusters diffère très nettement de celle décrite pour les clusters des souches SFi39, MTC310 et IP6756. Dans le cluster *eps* de type X, une pseudo-ORF (*eps10N*) de 993 nucléotides est localisée en aval d'*eps10M*, ORF codant une UDP-galactopyranose mutase. L'élément IS1191 identifié en 3' d'*eps10N* ne présente aucune homologie avec la séquence des éléments IS1193. Dans le cluster de type VI, une organisation comparable est relevée en 3' d'*eps6L*. Il est également important de rappeler ici que l'origine des séquences de ces clusters *eps* n'est pas indiquée dans les banques de données.

2 ETUDE DU LIEN ENTRE LE CARACTERE PRODUCTEUR D'EPS DE *S. THERMOPHILUS* IP6756 ET LE CLUSTER *EPS* IDENTIFIE

2.1 Détermination du caractère producteur d'EPS de la souche IP6756

Quatre extractions indépendantes d'EPS ont été réalisées à partir de cultures en milieu lait de la souche IP6756. Les dosages effectués selon la méthode de Dubois permettent d'établir que le taux moyen d'EPS produit par la souche IP6756 est de 120 mg/l avec un écart type de 32 mg/l.

Des mesures sur d'autres types d'échantillons sont nécessaires afin de déterminer si une telle production permet de considérer la souche IP6756 comme une souche productrice d'EPS. Ainsi, un témoin négatif est réalisé à partir d'un même volume de milieu lait sans inoculation bactérienne. Les deux dosages effectués présentent une valeur moyenne de 11 mg/l. Au-delà de cette valeur, les souches de *S. thermophilus* peuvent être considérées comme productrices d'EPS.

Un témoin positif est constitué par les dosages effectués sur des cultures en milieu lait de la souche NST2280. Cette souche présente le même cluster *eps* que la souche Sfi6, une souche considérée comme productrice d'EPS (Stingle *et al.*, 1996a). La mesure de la production d'EPS de la souche NST2280 a été effectuée à 9 reprises par Pluvinet *et al.* (2004) et lors de cette thèse et elle présente une valeur moyenne de 170 mg/l avec un écart type de 36 mg/l.

Un autre témoin négatif est constitué par un mutant de NST2280, dans lequel le gène *epsN*, qui est situé à droite d'*orf14.9*, est interrompu. Cette souche présente une production moyenne d'EPS de 25 mg/l (Pluvinet *et al.*, 2004).

L'ensemble de ces résultats permet de considérer que la souche IP6756 est une des souches productrice d'EPS de l'espèce *S. thermophilus*. L'analyse par chromatographie sur couche mince (TLC) des molécules d'EPS produites par *S. thermophilus* IP6756 révèle qu'elles contiennent principalement du glucose et du galactose (dans des proportions de

l'ordre de $\frac{3}{4}$ et $\frac{1}{4}$ respectivement) et des traces d'une molécule glucidique non identifiée (Figure 23). Les molécules d'EPS produites par la souche IP6756 sont donc de nature hétéropolysaccharidique.

2.2 Etude de l'activité transcriptionnelle des gènes de la première région du cluster *eps* de *S. thermophilus* IP6756

Chez *S. thermophilus* Sfi6 et *L. lactis* NIZOB40, productrices d'EPS (Stingele *et al.*, 1996a) et (van Kranenburg *et al.*, 1997), les gènes de la première région du cluster *eps* qui présentent une organisation modulaire classique sont organisés en opéron (CF paragraphe 1.5.1.1 de l'introduction). Les gènes *epsA*, *epsB*, *epsZ* et *epsJ* du cluster *eps* de *S. thermophilus* IP6756 sont-ils également organisés en opéron?

2.2.1 Analyse *in silico* : étude de la région promotrice et identification d'un terminateur de transcription potentiel

L'analyse de la séquence en amont d'*epsA* a permis d'identifier un promoteur potentiel P1 (Figure 24). Il présente une boîte -35 (TTGAAA) de séquence presque identique à celle de la boîte -35 consensus des bactéries appartenant à la division des Firmicutes (TTGACA) (Graves & Rabinowitz, 1986). La boîte -10 est localisée à 17 nucléotides (espacement consensus) et présente une identité parfaite avec la séquence de la boîte -10 consensus des bactéries appartenant à la division des Firmicutes (TATAAT) (Graves & Rabinowitz, 1986). Un RBS (ttAGGAGC) de séquence proche de celle du RBS consensus définie chez *Bacillus subtilis*, une bactérie appartenant à la division des Firmicutes, (aaAGGAGG) (Rocha *et al.*, 1999) est également identifié 24 nucléotides en aval de cette boîte -10. Ce RBS est localisé à 7 nucléotides du codon start de traduction d'*epsA*.

La comparaison de cette séquence promotrice avec celle de neuf autres clusters *eps* de *S. thermophilus* (celui des souches MTC310, NCFB2393, MR-1C, MR-2C, Sfi6 et Sfi39 ainsi

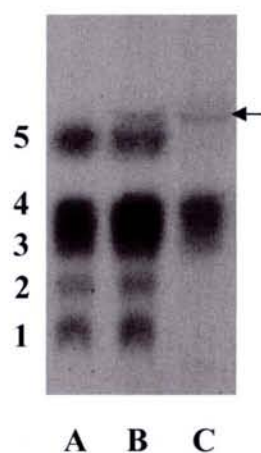


Figure 23 : Analyse par chromatographie sur couche mince de la composition des EPS produits par *S. thermophilus* IP6756.

A: standards : acide galacturonique (1), acide glucuronique (2), galactose (3), glucose (4) et rhamnose (5).

B: mélange d'EPS hydrolysés de IP6756 et des standards

C: EPS hydrolysés de IP6756

← : trace d'une molécule glucidique non identifiée

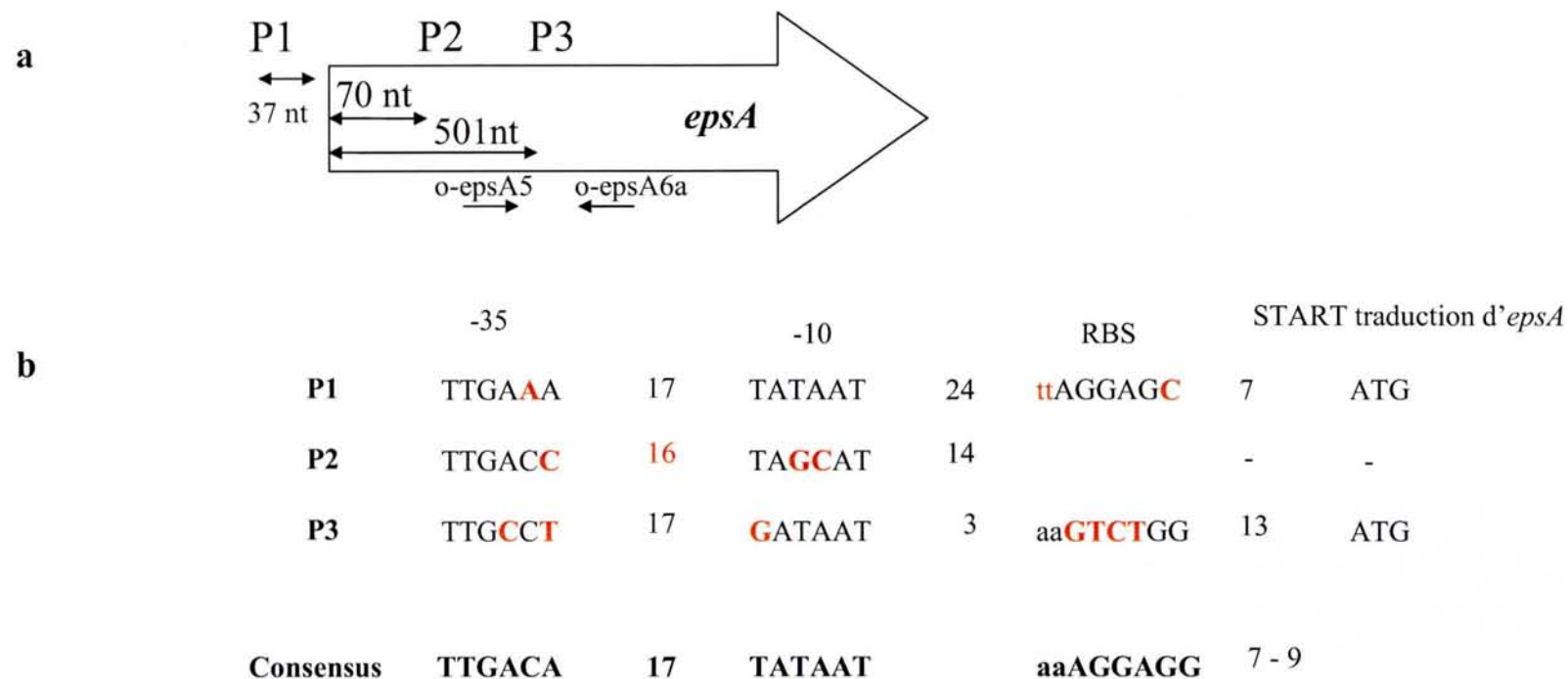


Figure 24 : Caractéristiques du promoteur potentiel de transcription P1 et des deux autres promoteurs potentiels de transcription identifiés dans *epsA* (P2 et P3).

a: Localisation des promoteurs potentiels de transcription P1, P2 et P3

Les distances indiquées s'entendent du codon start de traduction d'*epsA* au premier nucléotide des boîtes -35 (cas P2 et P3) ou au dernier nucléotide de la boîte -10 du promoteur P1.

b: Comparaison des séquences des boîtes -35 et -10 des promoteurs potentiels P1, P2 et P3.

Les séquences consensus représentées pour les boîtes -35 et -10 sont celles décrites chez les bactéries appartenant à la division des Firmicutes (Graves et Rabinowitz, 1986). La séquence consensus du RBS a été établie chez *Bacillus subtilis* par Rocha *et al.* (1999). Les nucléotides différant des nucléotides consensus sont représentés en rouge.

- : aucun RBS et aucun codon TTG, GTG ou ATG n'est identifié en aval de la boîte -10 du promoteur P2.

Les amorces utilisées pour détecter un transcrit spécifique d'*epsA* sont indiquées. Les extrémités 3' des amorces o-epsA5 et o-epsA6a sont situées respectivement 72 nt avant la boîte -35 du promoteur P3 et 205 nt après la boîte -10 du promoteur P3.

que les clusters *eps* de type IV, V et X) révèle qu'un promoteur présentant la même localisation que P1 peut également être identifié. La séquence de sa boîte -35 est identique et la séquence de sa boîte -10 est très conservée puisqu'elle est identique à la séquence consensus pour 6 clusters et ne diffère que d'un nucléotide pour les 4 autres clusters. De plus, le RBS identifié dans ces 10 autres clusters *eps* est identique à celui de la souche IP6756.

Des éléments CRE (*Catabolite Response Element*) potentiels étant présents dans la région promotrice des clusters *cps* de *S. pneumoniae* et en amont de l'opéron *gtfBC* et du gène *fif* de *S. mutans* (CF paragraphe 1.6.2 de l'introduction), de tels éléments ont été recherchés dans la région promotrice du cluster *eps* de la souche IP6756. Ainsi, un élément CRE potentiel présentant 11 nucléotides sur 14 identiques à ceux de la séquence consensus définie chez les bactéries appartenant à la division des Firmicutes (TGWNANCGNTNWCA (Burne *et al.*, 1999)) peut être identifié dans cette région mais également dans la région promotrice de 7 des 9 autres clusters *eps* analysés (Figure 25). Les souches Sfi39 et MTC310 sont les seules à ne pas présenter cet élément. Cette observation est à mettre en relation avec la présence de la séquence d'insertion IS905 qui interrompt la séquence intergénique localisée entre *deoD* et *epsA* dans ces souches. Un autre élément CRE potentiel est également identifié dans les séquences promotrices de toutes les souches analysées. Cet élément dont 9 des 14 nucléotides correspondent à la séquence consensus chevauche la boîte -35 identifiée dans chacune de ces séquences (Figure 25) et est donc conservé dans toutes les souches analysées.

L'analyse de la séquence d'*epsA* permet de détecter deux autres promoteurs potentiels, P2 et P3. Le promoteur P2 (Figure 24) présente une boîte -35 (TTGACC), localisée 70 nucléotides en 3' du codon start de traduction d'*epsA*. Sa séquence est séparée par 16 nucléotides de la boîte -10 (TAGCAT). Le promoteur P3 (Figure 24) présente une boîte -35 localisée 501 nucléotides en 3' du start de traduction d'*epsA*. Sa séquence (TTGCCT) est séparée par 17 nucléotides de la boîte -10 (GATAAT). Les promoteurs potentiels P2 et P3

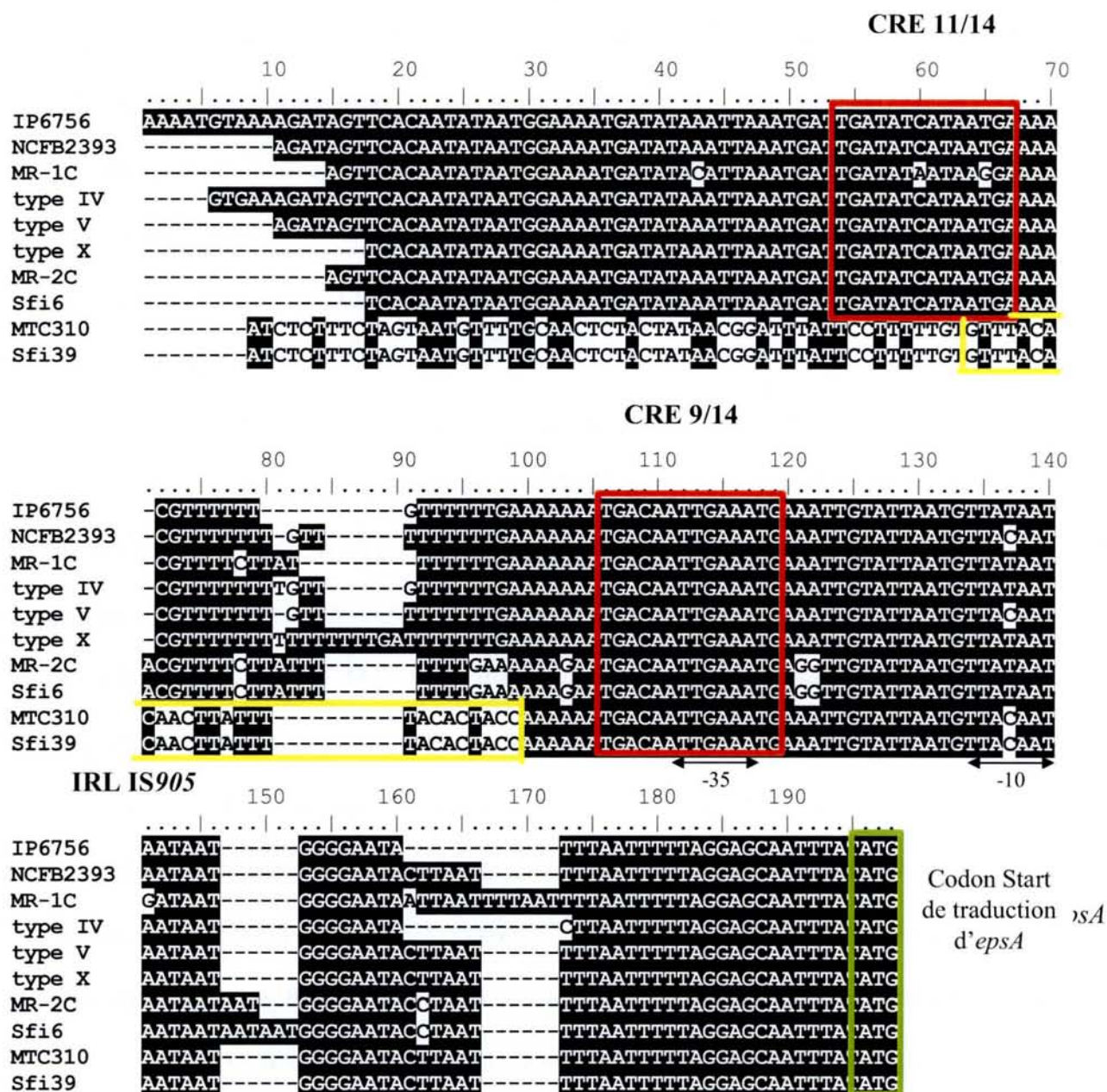


Figure 25 : Localisation des éléments CRE potentiels dans la région promotrice du cluster *eps* de *S. thermophilus* IP6756, NCFB2393, MR-1C, MR-2C, MTC310, Sfi6 et Sfi39 ainsi que dans la région promotrice des clusters *eps* de type IV, V et X.

Comparaison de la séquence intergénique *deoD-epsA* de 165 pb de la souche IP6756 avec les séquences correspondantes de même taille des autres clusters *eps* de *S. thermophilus* analysés.

Les éléments CRE potentiels sont encadrés en rouge. La séquence consensus des éléments CRE chez les bactéries appartenant à la division des Firmicutes est TGWNANCGNTNWCA (Burne *et al.*, 1999). Le codon start de traduction d'*epsA* est encadré en vert dans toutes les séquences. La localisation des boîtes -35 et -10 est également indiquée. La localisation de la répétition terminale inversée gauche (IRL) de la séquence d'insertion IS905 est encadrée en jaune. Les tirets correspondent à des décalages introduits lors de l'alignement.

présentent donc des boîtes -35 et -10 de séquences différant de 1 à 2 nucléotides de leurs séquences consensus (Graves & Rabinowitz, 1986). Par contre, immédiatement en aval de la boîte -10, aucun RBS ne peut être identifié dans le cas du promoteur P2 et la séquence du RBS (aaGTCTGG) identifié dans le cas du promoteur P3 diffère de quatre nucléotides de la séquence consensus. Par ailleurs, un codon start de traduction ATG respectant la phase de lecture d'*epsA* est identifié 13 nucléotides en 3' de ce RBS.

La comparaison de la séquence d'*epsA* de la souche IP6756 avec celle des neuf autres clusters *eps* analysés de *S. thermophilus* (celui des souches MTC310, NCFB2393, MR-1C, Sfi6 et Sfi39 ainsi que les clusters *eps* de type IV, V et X) révèle que deux promoteurs présentant la même localisation que P2 et P3 peuvent également être identifiés. La séquence des boîtes -35 et -10 du promoteur localisé au même endroit que P2 dans les neuf autres clusters est identique à celle des boîtes -35 et -10 du promoteur P2 de la souche IP6756. Les séquences des boîtes -10 et -35 des promoteurs localisés au même endroit que P3 sont identiques à celles du promoteur P3, à l'exception du cinquième nucléotide de la boîte -35 du promoteur identifié dans le cluster de type X.

Un terminateur de transcription potentiel est identifié en 3' d'*epsJ* du cluster *eps* d'IP6756 à l'aide du logiciel mfold3.0. Il comporte deux structures tiges-boucles, la seconde correspondant potentiellement à un terminateur de transcription rho-indépendant se terminant par des bases uraciles (poly U) (Figure 26). Cette structure tige-boucle potentielle présente un ΔG de -33 kcal/mol à 37°C selon le même logiciel. A titre de comparaison, le terminateur de transcription identifié en aval d'*epsM*, la dernière ORF de la première région du cluster *eps* de *S. thermophilus* Sfi6, présente selon le même logiciel un ΔG de -39,2 kcal/mol à 37°C.

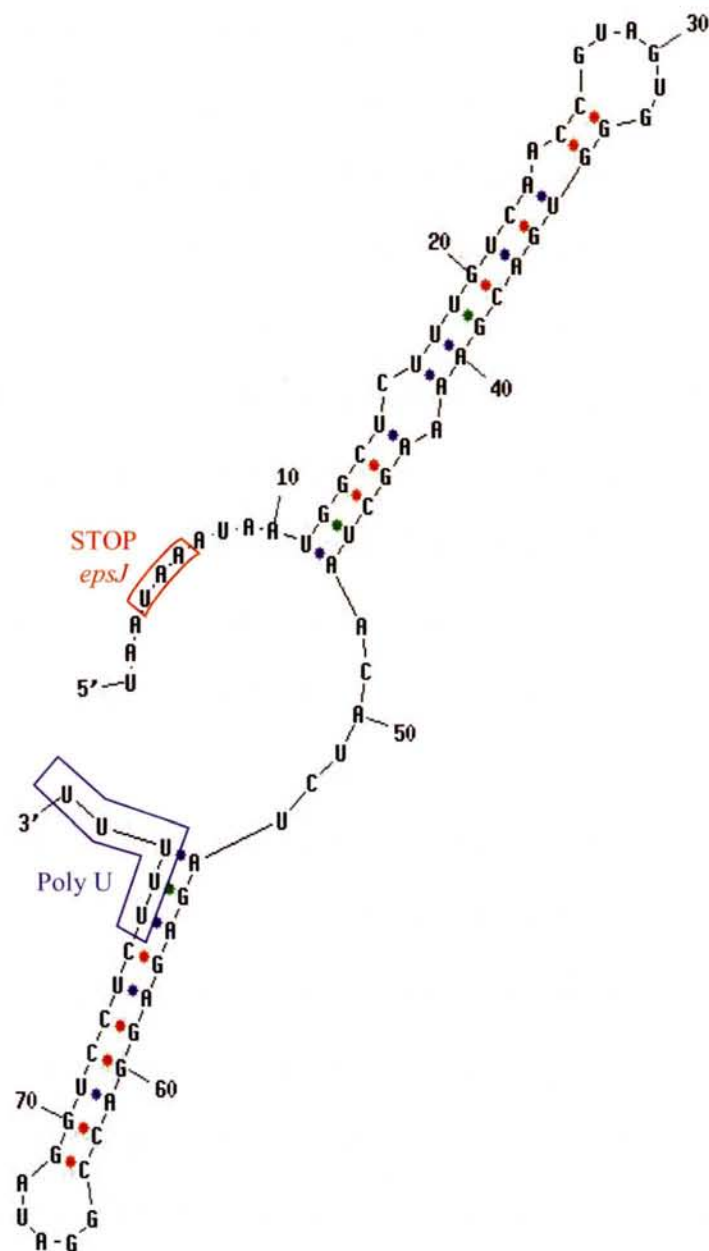


Figure 26 : Structure du terminateur de transcription potentiel identifié en 3' d'*epsJ* à l'aide du logiciel mfold 3.0.

STOP *epsJ* : codon d'arrêt de la traduction d'*epsJ*.

Poly U : bases uraciles.

2.2.2 Les quatre premières ORF du cluster *eps* seraient organisées en opéron

L'analyse transcriptionnelle des ORF *epsA*, *epsB*, *epsZ* et *epsJ* a été entreprise par RT-PCR sur les ARN totaux de la souche IP6756 extraits en début de phase stationnaire de croissance en milieu M17. Lors des premières extractions, les cellules étaient reprises dans le mélange de thiocyanate de guanidine et de phénol puis lysées par quatre sonications de 40W pendant cinq secondes chacune. Les expériences de RT-PCR réalisées sur les ARN ainsi extraits n'ont pas permis d'obtenir des résultats reproductibles. La sonication pouvant nuire à l'intégrité du potentiel transcrit opéronique d'environ 4,5 kb attendu, le protocole d'extraction initial a été adapté. La lyse des cellules est obtenue en les vortexant à vitesse maximale pendant 80 secondes. Les résultats obtenus avec ce nouveau protocole se sont révélés reproductibles. Chaque transcrit identifié par RT-PCR a été obtenu au moins trois fois et a été obtenu à chaque expérience. Les amorces utilisées lors des expériences de RT-PCR sont présentées dans le tableau 5 de la partie matériel et méthodes et sont localisées figure 27.

Les expériences de RT-PCR réalisées sur des ARN totaux extraits selon ce protocole ont permis d'identifier un transcrit spécifique pour chacune des ORF *epsA*, *epsB*, *epsZ* et *epsJ* (Figure 28). La localisation des amorces utilisées pour détecter le transcrit spécifique d'*epsA* est précisée sur la figure 24. Elles permettent donc de détecter un transcrit en amont du promoteur potentiel P3.

Des expériences de RT-PCR ont également été réalisées pour mettre en évidence d'éventuels transcrits spécifiques des régions chevauchant chaque couple d'ORF. Elles ont ainsi permis de mettre en évidence un transcrit correspondant aux jonctions *epsA-epsB*, *epsB-epsZ* et *epsZ-epsJ* (Figure 29).

Les résultats obtenus suggèrent très fortement que cette région du cluster de la souche IP6756 située en 3' d'*orf14.9* présente une organisation opéronique comme les régions

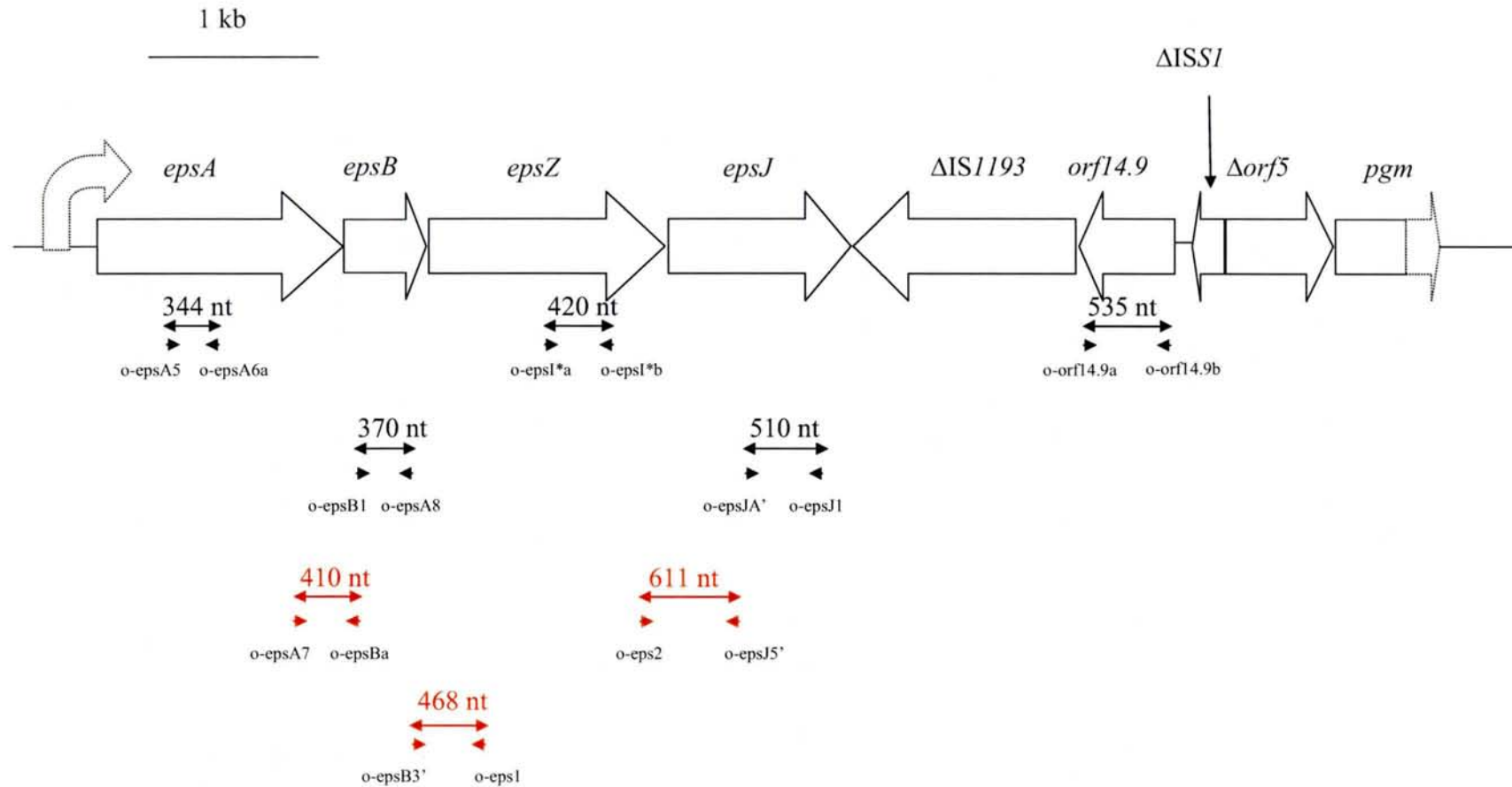


Figure 27 : Localisation des amorces utilisées lors des réactions de RT-PCR pour la détection des transcrits spécifiques à *epsA*, *epsB*, *epsZ*, *epsJ* et *orf14.9* (flèches noires) et des transcrits de jonction chevauchant *epsA* et *epsB*, *epsB* et *epsZ* ainsi qu'*epsZ* et *epsJ* (flèches rouges). La taille des transcrits attendus est indiquée. Le promoteur potentiel de transcription P1 est figuré par la flèche en 5' d'*epsA*. 5'►3': amorces utilisées (CF tableau 2 du matériel et méthodes)

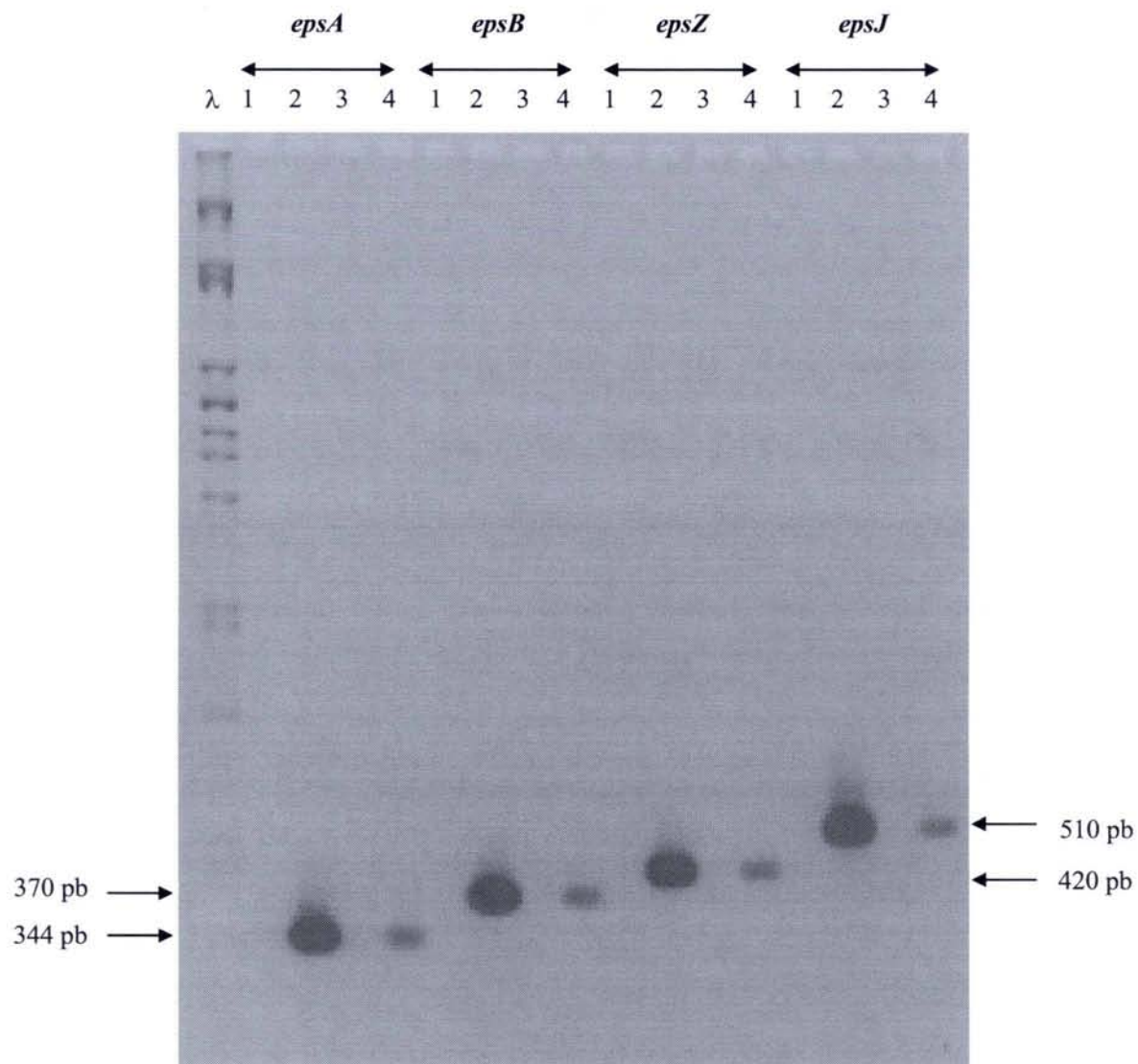


Figure 28 : Détection de transcrits spécifiques à *epsA*, *epsB*, *epsZ* et *epsJ* par PCR chez *S. thermophilus* IP6756

1 : sans matrice d'acides nucléiques

2 : ADN génomique de *S. thermophilus* IP6756

3 : ARN totaux extraits de la souche IP6756 ayant subi le traitement de la DNase

4 : ADNc obtenus par action de la transcriptase inverse

λ : ADN du phage λ digéré par *Pst*I

Electrophorèse réalisée sur un gel TAE 1X contenant 2% d'agarose (migration 12h à 2V/cm)

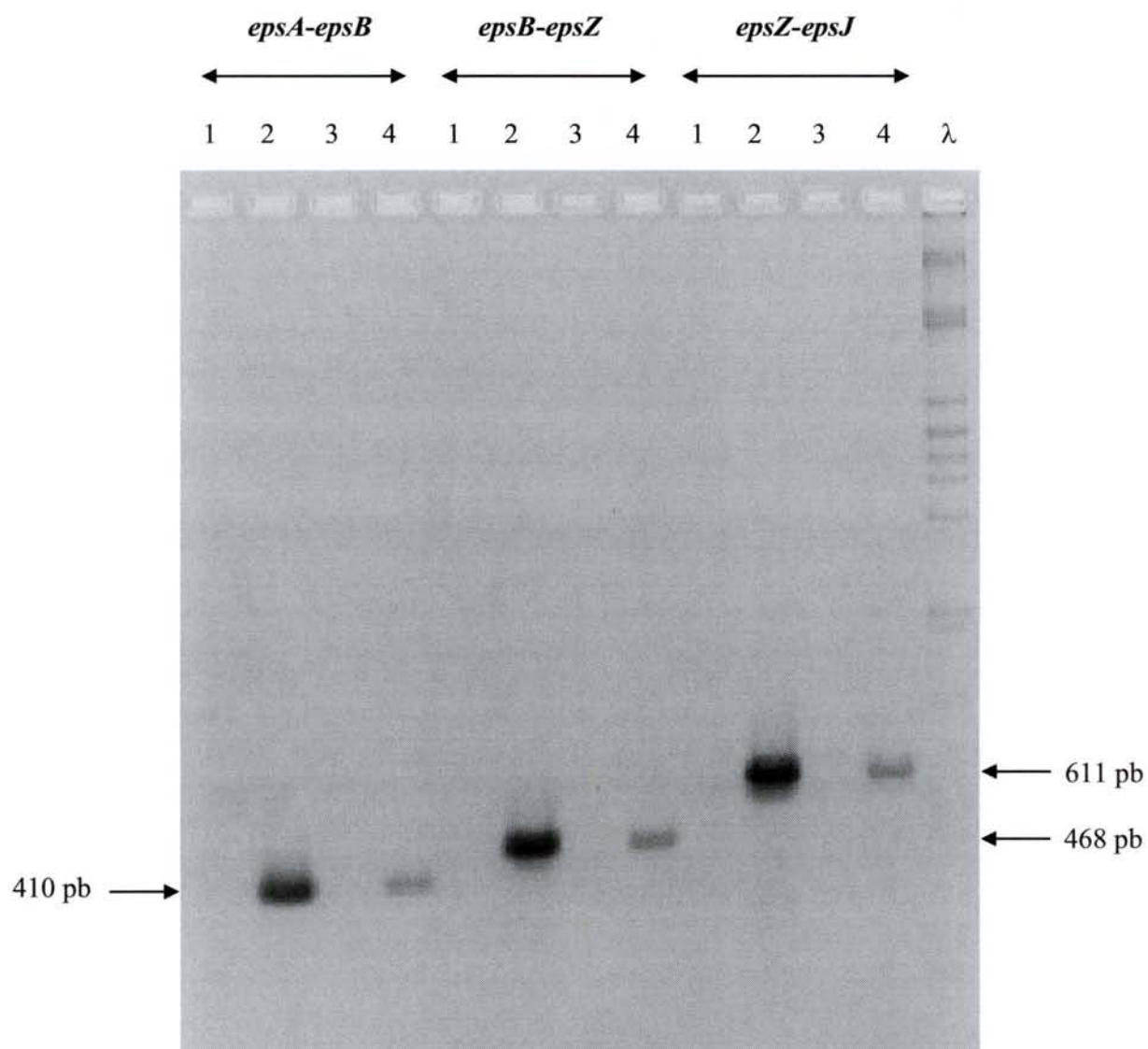


Figure 29 : Détection de transcrits spécifiques des jonctions *epsA-epsB*, *epsB-epsZ* et *epsZ-epsJ* chez *S. thermophilus* IP6756 par PCR

1 : sans matrice d'acides nucléiques

2 : ADN génomique de *S. thermophilus* IP6756

3 : ARN totaux extraits de la souche IP6756 ayant subi le traitement à la DNase

4 : ADNc obtenus par action de la transcriptase inverse

λ : ADN du phage λ digéré par *Pst*I

Electrophorèse réalisée sur un gel TAE 1X contenant 2% d'agarose (migration 12h à 2V/cm).

correspondantes d'autres clusters *eps* de souches productrices. Par ailleurs, les gènes appartenant à la première région du cluster *eps* de *S. thermophilus* CNRZ1066 et LMG18311 sont potentiellement transcrits dans le même sens que la réplication du chromosome. Dans la souche IP6756, il pourrait en être de même.

2.3 Etude de la fonctionnalité du cluster *eps* de la souche IP6756

Les comparaisons de séquences avec les banques de données ont permis d'attribuer une fonction potentielle aux produits des quatre premières ORF du cluster *eps* de la souche IP6756. La protéine EpsA serait impliquée dans la régulation transcriptionnelle du cluster. La protéine EpsB serait également une protéine régulatrice, mais sa fonction pourrait être modifiée puisqu'elle est tronquée et que les protéines associées EpsC et EpsD sont absentes. Les deux ORF suivantes *epsZ* et *epsJ* coderaient respectivement un transporteur d'unité répétée et une UDP-galactopyranose mutase. Ainsi, aucune de ces ORF ne coderait de glycosyltransférase, enzyme qui catalyse l'ajout séquentiel de molécules glucidiques sur les unités répétées en cours d'élongation, alors que la souche IP6756 est productrice d'EPS. Cette constatation amène à s'interroger sur la fonctionnalité de ce cluster et sur la localisation du ou des gènes codant la ou les glycosyltransférases nécessaires à la production d'EPS par cette souche.

2.3.1 Délétion et interruption polaire du cluster *eps* de la souche IP6756

Afin d'étudier l'implication du cluster *eps* de la souche IP6756 dans la production d'EPS, sa délétion a été entreprise. Pour cela, le plasmide à origine de réplication thermosensible pGHost9*epsAorf5* (CF paragraphe 8.1 et figure 45 de la partie matériel et méthodes) qui contient un insert issu de la ligation d'un fragment interne d'*epsA* de 782 pb et d'un fragment interne d'*orf5* de 452 pb ainsi que le terminateur transcriptionnel du gène *ery*

est introduit dans la souche IP6756 par électroporation. Puis ce plasmide est intégré dans son chromosome par recombinaison homologue (figure 30a). La localisation de ces fragments internes est précisée dans la figure 46 du matériel et méthodes. Le mutant IP6756-A17 obtenu qui présente une intégration multiple du plasmide pGHost9*epsAorf5* dans le gène *epsA* (Figure 30b) est utilisé pour entreprendre la délétion du cluster suite à une recombinaison homologue dans *orf5*. Tous les clones isolés à l'issue de cette expérience présentent un cluster *eps* identique au cluster *eps* initial, suite à un événement de recombinaison homologue dans *epsA*. La délétion du cluster *eps* d'IP6756 n'a donc pu être obtenue.

L'intégration du plasmide pGHost9*epsAorf5* dans *epsA* chez le mutant IP6756-A17 s'accompagne de l'intégration du terminateur de transcription porté par le plasmide (Figure 30a). L'intégration du plasmide pGHost9*epsAorf5* entraîne donc potentiellement, par effet polaire, l'inactivation des gènes en aval d'*epsA*. L'étude du mutant IP6756-A17 a été entreprise afin d'étudier l'implication du cluster *eps* dans la biosynthèse des EPS dans la souche IP6756. Dans un premier temps, l'effet de l'intégration du plasmide dans *epsA* sur la transcription des autres ORF du cluster est étudié par RT-PCR. Chez IP6756-A17, aucun transcrit spécifique à *epsB*, *epsZ* et *epsJ* n'a pu être observé, alors qu'un transcrit spécifique à *orf14.9* est observé et constitue un témoin positif (Figure 31). L'intégration du plasmide dans *epsA* présente donc un effet polaire sur la transcription des trois ORF situées en 3' d'*epsA*.

Dans un second temps, l'implication du cluster *eps* localisé entre *deoD* et *pgm* dans la biosynthèse des EPS est étudiée. Les EPS produits par le mutant IP6756-A17 ont été extraits et dosés. La comparaison du taux de production d'EPS de ce mutant par rapport à la souche sauvage ne montre pas de différence statistiquement significative (Figure 32).

De la même façon, la délétion du cluster *eps* de la souche NST2280 a été entreprise avec le plasmide pGHost9*epsAorf5*. Comme précédemment, le mutant NST2280-A8 obtenu par intégration multiple du plasmide dans *epsA* n'a pas permis d'obtenir la délétion du cluster

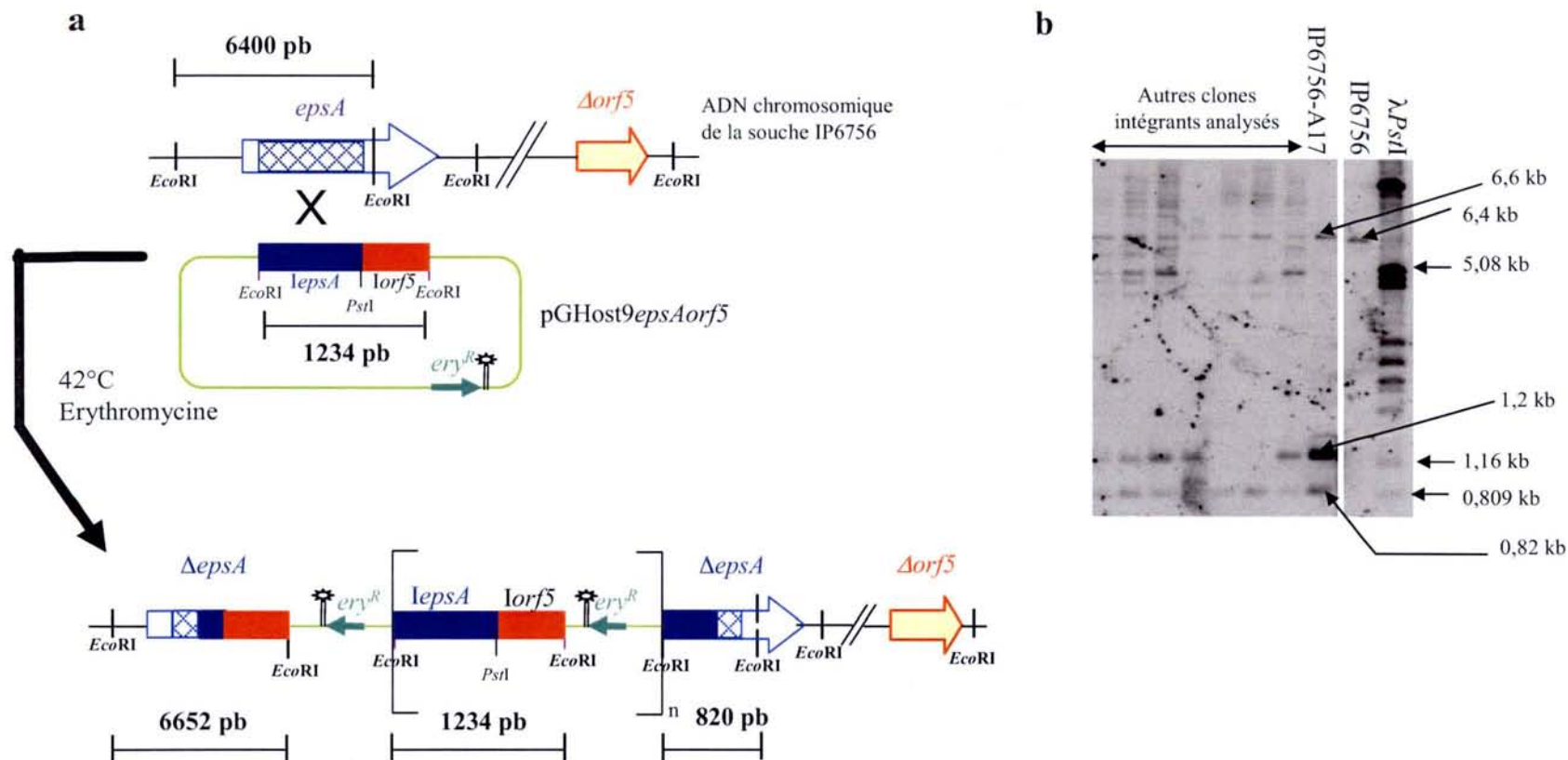


Figure 30 : Caractérisation du mutant IP6756-A17 obtenu par intégration multiple du plasmide pGHost9*epsAorf5* dans *epsA* de *S. thermophilus* IP6756.

a: Schéma de l'intégration multiple du plasmide pGHost9*epsAorf5* par recombinaison homologue dans *epsA*.

EcoRI : site de restriction *EcoRI*. *ery^R* : gène de résistance à l'érythromycine. $\parallel$: terminateur de transcription du gène *ery^R*

IepsA et *Iorf5* : fragment interne à *epsA* et *orf5* respectivement.

[_n] : n copies du plasmide.

b: Caractérisation des clones obtenus.

Hybridation d'une sonde spécifique du fragment interne *epsA* sur des profils de restriction *EcoRI* de l'ADN génomique de *S. thermophilus* IP6756 et des intégrants obtenus dont IP6756-A17.

λ PstI correspond au marqueur de taille.

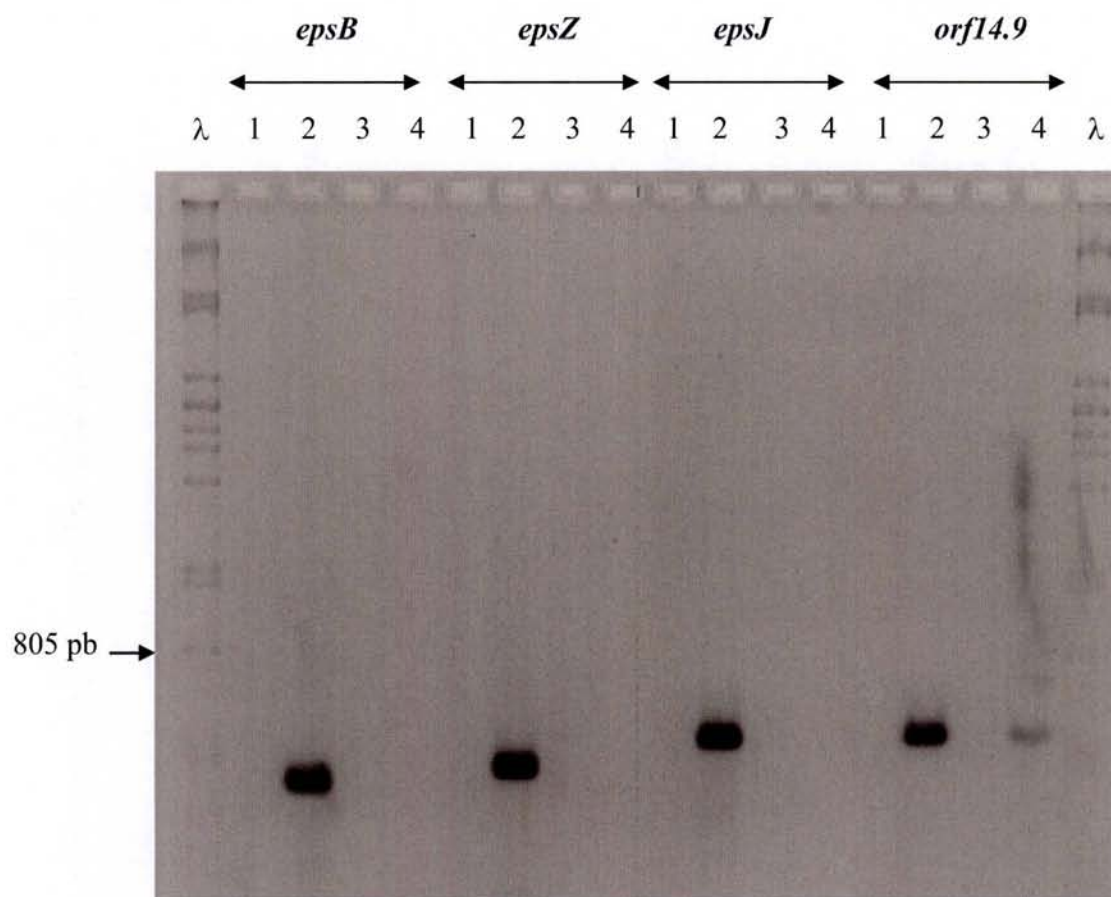


Figure 31 : Détection de transcrits spécifiques d'*epsB*, *epsZ*, *epsJ* et *orf14.9* par PCR chez le mutant IP6756-A17.

1 : sans matrice d'acides nucléiques

2 : ADN génomique de *S. thermophilus* IP6756

3 : ARN totaux extraits de la souche IP6756-A17 ayant subi le traitement de la DNase

4 : ADNc obtenus par action de la transcriptase inverse

λ : ADN du phage λ digéré par *Pst*I

Electrophorèse réalisée sur un gel contenant 2% d'agarose (migration 12h à 2V/cm)

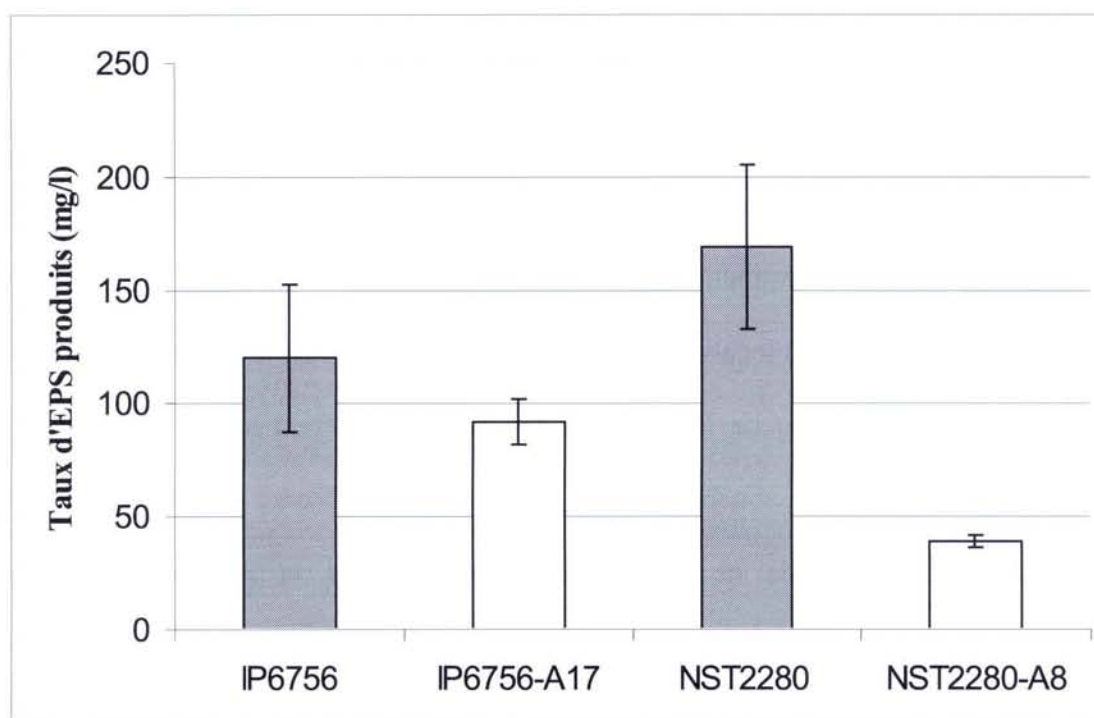


Figure 32 : Taux de production d'EPS des souches IP6756 et NST2280 ainsi que de leurs mutants insertionnels respectifs dans *epsA* IP6756-A17 et NST2280-A8.

Le taux de production d'EPS a été déterminé quatre fois pour les souches IP6756, IP6756-A17 ainsi que NST2280-A8 et neuf fois pour la souche NST2280. Les écarts-types de chaque valeur sont indiqués.

eps (Figure 33). Toutefois, les EPS produits par le mutant NST2280-A8 ont été extraits et dosés. Les résultats obtenus montrent que le mutant NST2280-A8 présente une production d'EPS fortement réduite par rapport au sauvage (Figure 32).

A l'inverse de ce qui est observé dans la souche NST2280, l'inactivation du cluster *eps* de la souche IP6756 s'accompagne du maintien de la production d'EPS. Ce résultat suggère l'intervention de déterminants génétiques autres que le cluster *eps* identifié pour IP6756. Un ou plusieurs gène(s) localisés hors du cluster *eps* serai(en)t donc nécessaire(s) à la biosynthèse des EPS de la souche IP6756. Ce résultat est cohérent avec l'analyse de la séquence du cluster *eps* de la souche IP6756, qui montrait l'absence d'ORF codant une glycosyltransférase au sein de ce cluster et suggérait pour la première fois chez *S. thermophilus* l'intervention d'un ou plusieurs gène(s) extérieur(s) au cluster codant la ou les glycosyltransférases nécessaires à cette biosynthèse.

2.3.2 Recherche du ou des gènes codant la ou les glycosyltransférases nécessaires à la biosynthèse des EPS chez *S. thermophilus* IP6756

2.3.2.1 *Recherche de séquences homologues aux gènes codant les glycosyltransférases de S. thermophilus CNRZ368, IP6757 et NST2280*

Le nombre de glycosyltransférases impliquées dans la biosynthèse des EPS est très variable selon les souches. De plus, les gènes codant ces enzymes présentent une très forte variabilité de séquence, à l'exception de ceux codant la première glycosyltransférase. En effet, cette première glycosyltransférase présente un degré d'homologie important tant chez différentes protéobactéries (*E. coli* et *K. pneumoniae* notamment) que chez différentes bactéries à coloration de Gram positive (bactéries présentant un haut contenu en bases G et C ou appartenant à la division des Firmicutes) dont *S. pneumoniae* (Wang *et al.*, 1996).

De nombreuses sondes spécifiques des ORF codant des glycosyltransférases ont été réalisées à partir des séquences des clusters *eps* des souches CNRZ368

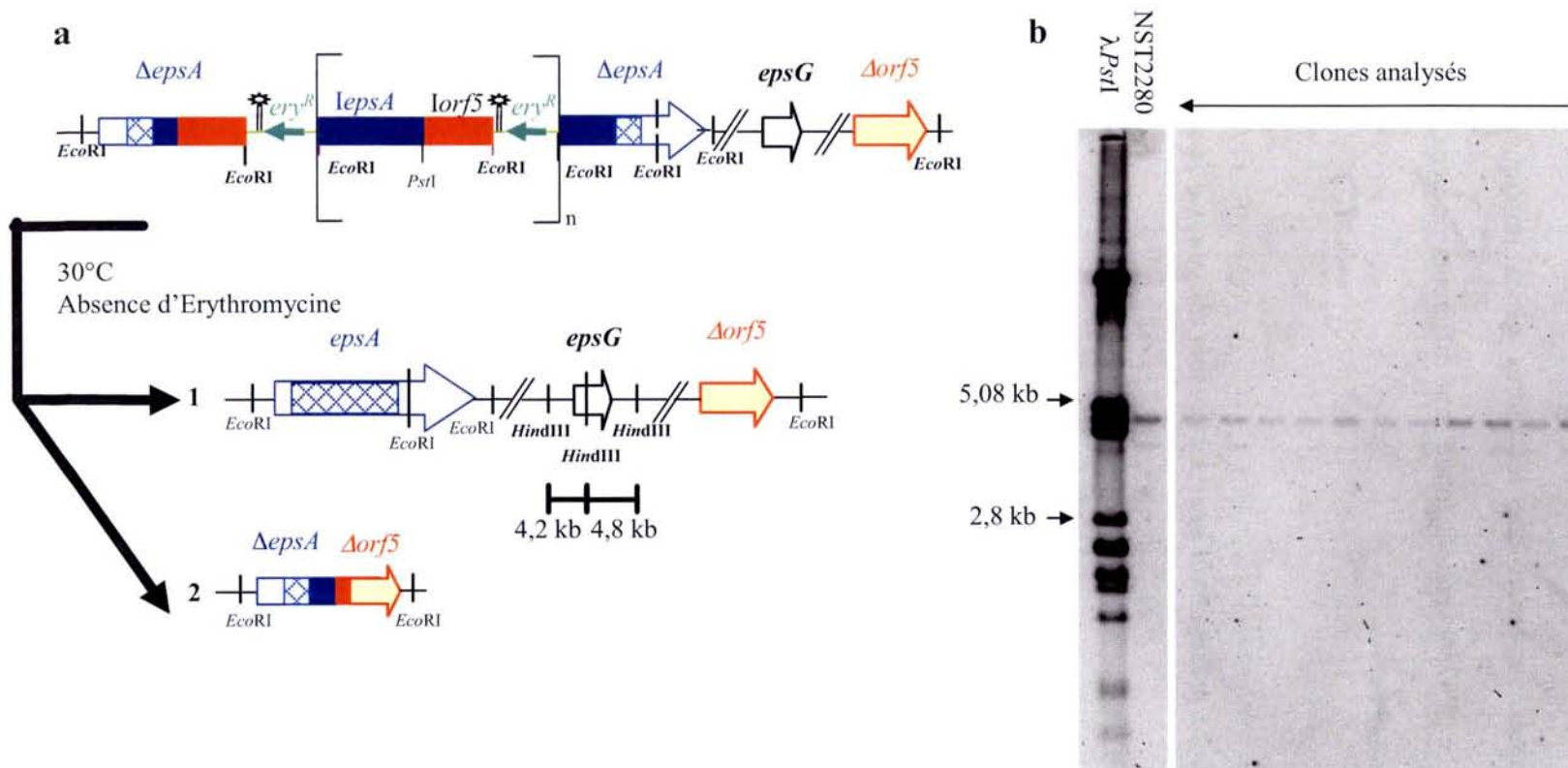


Figure 33 : Caractérisation des clones obtenus lors de l'expérience de délétion du cluster *eps* de *S. thermophilus* NST2280.

a: Excision du plasmide pGHost9*epsAorf5* intégré dans *epsA* de *S. thermophilus* NST2280 par recombinaison homologue au niveau d'*epsA* (**1**) ou au niveau de Δ *orf5* (**2**), à température permissive et sans érythromycine.

Seule la recombinaison au niveau de Δ *orf5* conduit à la délétion des gènes compris entre *epsA* et Δ *orf5*.

ery^R : gène de résistance à l'érythromycine. *lepsA* et *lorf5* : fragment interne à *epsA* et *orf5* respectivement.

EcoRI : site de restriction *EcoRI*. Dans la situation 1 : les sites *HindIII* utilisés par la suite sont représentés.

b: Caractérisation des clones obtenus.

Hybridation d'une sonde spécifique d'*epsG* (utilisée par Pluvinet *et al.*, 2004) sur les profils de restriction *HindIII* de l'ADN génomique de *S. thermophilus* NST2280 et des clones obtenus.

λ*PstI* correspond au marqueur de taille.

(Bourgoin *et al.*, 1999), IP6757 (Pluvinet *et al.*, 2004) et NST2280 (Pluvinet, 2003).

L'hybridation de ces sondes sur l'ADN génomique de la souche IP6756 à forte stringence ne révèle aucun signal. Le génome de la souche IP6756 ne contient donc pas de séquences présentant une homologie significative avec les séquences codant les glycosyltransférases des trois souches testées. Ces résultats suggèrent que les autres gènes impliqués dans la biosynthèse des EPS de la souche IP6756 seraient totalement différents de ceux identifiés dans les clusters *eps* de ces trois souches.

2.3.2.2 Recherche de la présence du gène *tts* au sein du génome de la souche IP6756

Les résultats présentés dans le paragraphe 4.1 sont à rapprocher de la situation décrite chez *S. pneumoniae* de type 37 dont le cluster *cps* n'est pas impliqué dans la biosynthèse de la capsule de type 37 (Llull *et al.*, 1999) (CF paragraphe 1.7.2 de l'introduction). Chez cette souche, c'est le gène *tts*, situé hors du locus *cps* et plus précisément entre les gènes *pyrDA* et *gpmA*, qui est directement impliqué dans la biosynthèse de la capsule de la souche. Ce gène code une protéine potentielle de 509 acides aminés, catalysant la synthèse de la capsule de type 37, et qui contiendrait un peptide signal dont le site de clivage serait localisé entre les acides aminés 32 et 33 ou entre les acides aminés 36 et 37.

Aucune séquence nucléotidique présente dans les banques de données, notamment celle des génomes de *S. thermophilus* LMG18311 et CNRZ1066 entièrement séquencés (numéros d'accension : NC_006448 et NC_006449 respectivement), ne montre d'homologie avec la séquence de *tts* établie par Llull *et al* (numéro d'accension : AJ131985). Ainsi, le gène *tts* n'est pas présent chez ces deux souches, ce que confirme l'analyse de la région contenant les gènes *pyrDA* et *gpmA* chez ces deux souches (Figure 34). Chez ces deux souches, l'extrémité 3' du gène *gpmA* est située à plus de 78 kb de l'extrémité 5' d'*epsA* et donc du cluster *eps*. La distance séparant le gène *gpmA* du cluster *cps* chez *S. pneumoniae* de type 37 n'est pas

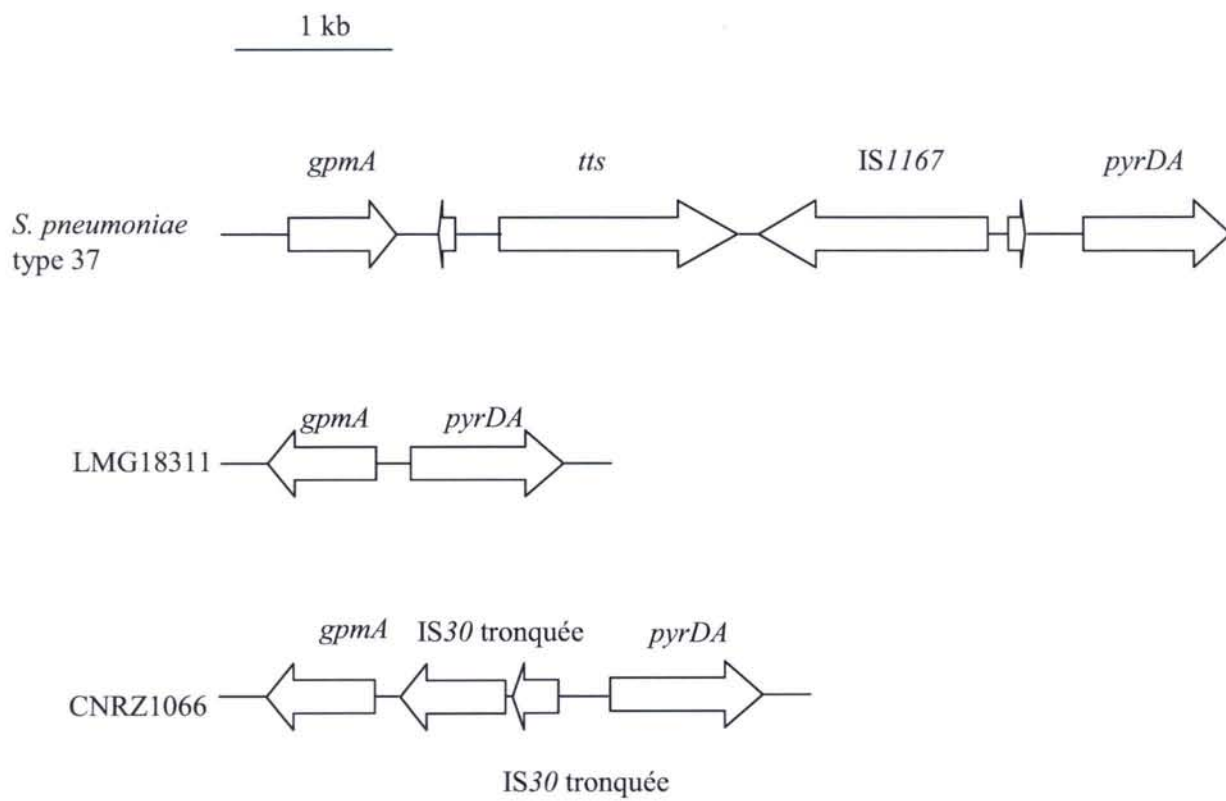


Figure 34 : Organisation de la région comprise entre *gpmA* et *pyrDA* chez *S. pneumoniae* type 37 et *S. thermophilus* LMG18311 et CNRZ1066

□ : élément répété de 105 pb

explicitement mentionnée. A titre de comparaison, la distance séparant *gpmA* de la frontière *aliA* du cluster *cps* de *S. pneumoniae* TIGR4 est de plus de 1,2 Mb (http://www.sanger.ac.uk/Projects/S_pneumoniae/CPS/). Par ailleurs, seule une glycosyltransférase potentielle de *Mycoplasma penetrans* HF-2 (E-value de 10^{-84}) présente 34% d'identité avec la protéine potentielle Tts.

Une sonde de 367 pb spécifique du gène *tts* de *S. pneumoniae* type 37 est obtenue par PCR puis hybridée à forte stringence sur l'ADN génomique digéré par *HindIII* de 4 souches de *S. thermophilus* (IP6756, NST2280, CNRZ1066 et LMG18311). Les résultats obtenus ne révèlent aucun signal, sauf lorsque la sonde est hybridée sur l'ADN génomique de *S. pneumoniae* de type 37 (Figure 35), indiquant qu'aucune séquence présentant une homologie significative avec la séquence de *tts* n'est donc présente dans le génome des quatre souches et en particulier dans celui de *S. thermophilus* IP6756.

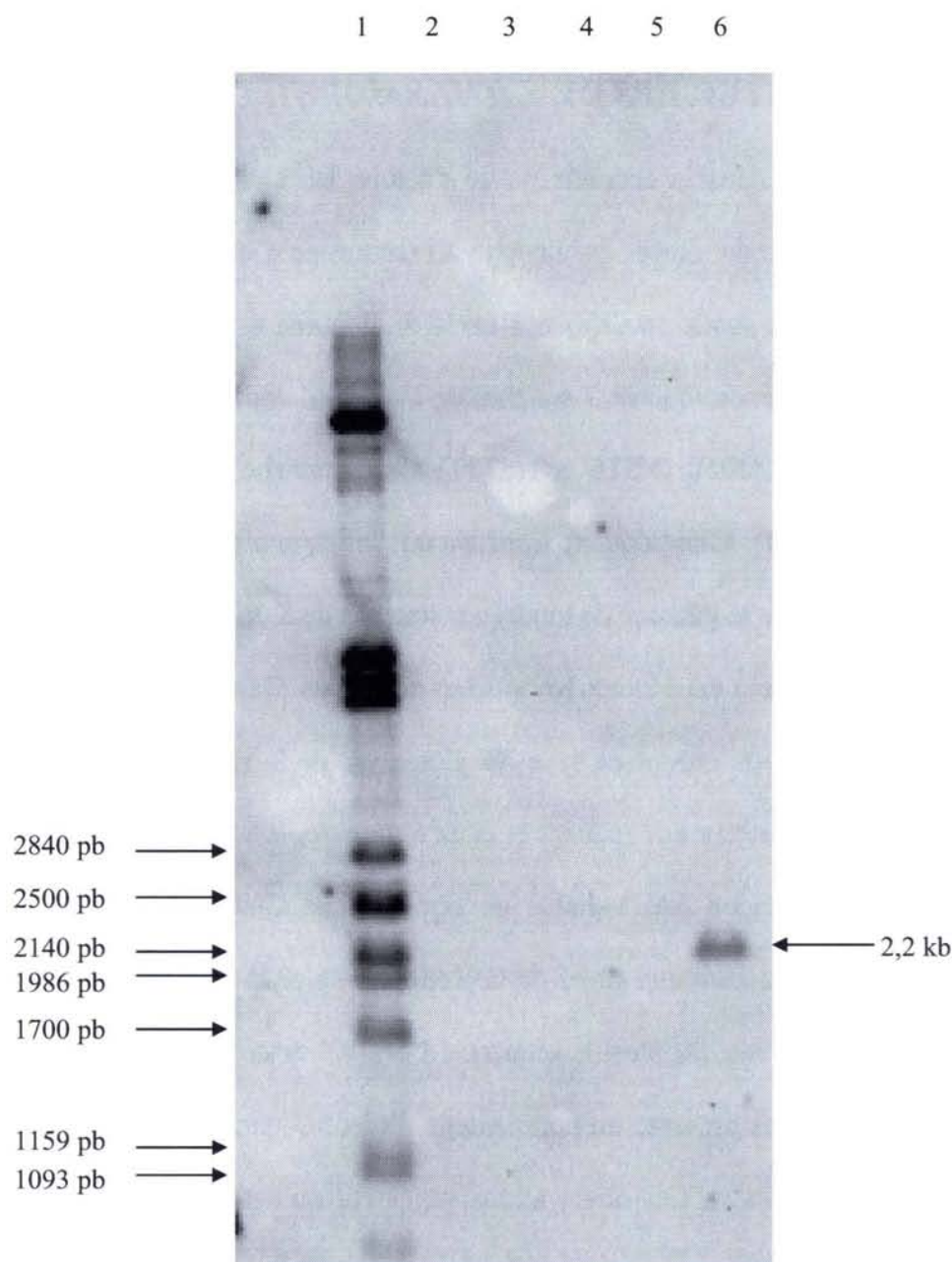


Figure 35 : Hybridation d'une sonde spécifique de *tts* sur un profil de digestion *Hind*III de l'ADN génomique de *S. thermophilus* IP6756 (2), NST2280 (3), CNRZ1066 (4) et LMG18311 (5) et de *S. pneumoniae* type 37 (6).

Une sonde correspondant à l'ADN du phage λ marqué a été utilisée pour révéler le marqueur de taille (ADN du phage λ digéré par *Pst*I (1)).

La sonde utilisée a été obtenue par PCR avec les amorces o-tts1 et o-tts2 sur l'ADN génomique de *S. pneumoniae* de type 37 fourni par l'équipe du Dr Garcia (CF paragraphe 1 et le tableau 2 de la partie matériel et méthodes).

3 FONCTIONNALITE D'ORF14.9, LA SEULE ORF POTENTIELLEMENT FONCTIONNELLE DE LA SECONDE REGION DU CLUSTER EPS DE *S. THERMOPHILUS* IP6756

Cette ORF est présente dans la seconde partie de tous les clusters *eps* séquencés de *S. thermophilus*, à l'exception du cluster *eps* de type XI (séquence d'*orf14.9* non disponible). De plus, l'hybridation d'une sonde spécifique à *orf14.9*, élaborée à partir de l'allèle de la souche IP6757, sur l'ADN génomique de 7 souches de *S. thermophilus* (CNRZ7, CNRZ404, CNRZ445, CNRZ464, CNRZ1068, NST5 et NST11) s'accompagne de l'observation d'un signal (Pluvinet *et al.*, 2004). Une séquence présentant une homologie significative avec *orf14.9* est donc présente dans le génome de toutes ces souches de *S. thermophilus*.

Cette ORF est donc présente dans toutes les souches analysées. Dans la suite de ce propos, nous ne traiterons pas le cas du cluster de type XI. L'analyse de la séquence du cluster *eps* d'IP6756 et des 21 autres clusters *eps* séquencés chez *S. thermophilus* révèle qu'*orf14.9* est localisée dans la seconde région très variable de ces clusters. Cette ORF est également orientée en sens inverse par rapport aux gènes de la première région du cluster *eps* et présente donc une organisation conservée. De plus, la séquence d'*orf14.9* est très conservée puisque la séquence des différents allèles présente un pourcentage d'identité compris entre 96 et 100% (CF paragraphe 2.3 des résultats). Ces observations suggèrent qu'*orf14.9* est soumise à une pression de sélection particulière ce qui amène à s'interroger sur le rôle de cette ORF et en particulier sur l'implication éventuelle d'*orf14.9* dans la biosynthèse des EPS.

3.1 Analyse *in silico* d'*orf14.9*

Chez les 22 souches étudiées par séquençage, *orf14.9* est orientée en sens inverse par rapport aux gènes de la première région du cluster *eps*. Dans les souches CNRZ1066 et LMG18311, elle serait ainsi transcrite en sens inverse de la réplication du chromosome. Cette

propriété s'accompagnerait d'un taux de transcription plus faible, à la différence de celui de la première région du cluster *eps*. La figure 36 présente la localisation d'*orf14.9* au sein du cluster *eps* de *S. thermophilus*. Les trois situations présentées varient selon la présence et le nombre d'ORF impliqués dans la biosynthèse des EPS en 5' d'*orf14.9*.

La comparaison des séquences d'*orf14.9* disponibles révèle que la séquence de 156 pb immédiatement en 5' de la séquence de l'ORF d'*orf14.9* chez la souche IP6756 est très conservée entre les souches. Cette séquence contient un site de fixation potentiel du ribosome (RBS) qui est situé à 9 pb du codon start de traduction d'*orf14.9* et dont la séquence est très conservée entre les souches (généralement aaAGGAGT) et très similaire à la séquence consensus aaAGGAGG, défini par Rocha *et al.* (1999) chez *B. subtilis*. En amont du RBS, une boîte -10 potentielle de séquence TAAAAT ou TAGAAT pour l'allèle de MR-1C est détectée. Elle est située à 14 pb d'une boîte -35 potentielle également de séquence ATGACA chez IP6756 ou ATGAAA pour l'allèle *orf14.9* du cluster *eps* de type VII par exemple. Les séquences des boîtes -10 et -35 de ce promoteur potentiel de transcription sont assez proches des séquences consensus définies chez les bactéries appartenant à la division des Firmicutes (Graves & Rabinowitz, 1986), ce qui suggère que ce promoteur pourrait être fonctionnel.

Un terminateur de transcription potentiel rho-indépendant se terminant par des bases uraciles (polyU) est identifié en 3' d'*orf14.9* chez IP6756 à l'aide du logiciel mfold3.0 (Figure 37). Cette structure tige-boucle potentielle présente un ΔG de -35,6 kcal/mol à 37°C selon le même logiciel. Dans les autres clusters *eps*, l'étude des 150 pb en 3' d'*orf14.9* permet également de détecter un terminateur de transcription potentiel même si cette séquence de 150 pb n'est pas conservée entre les clusters.

La comparaison des 22 séquences disponibles permet également de relever une conservation très importante de la séquence de l'ORF d'*orf14.9* (Tableau 2), suggérant que cette ORF pourrait conférer un avantage sélectif à *S. thermophilus*. Néanmoins, une

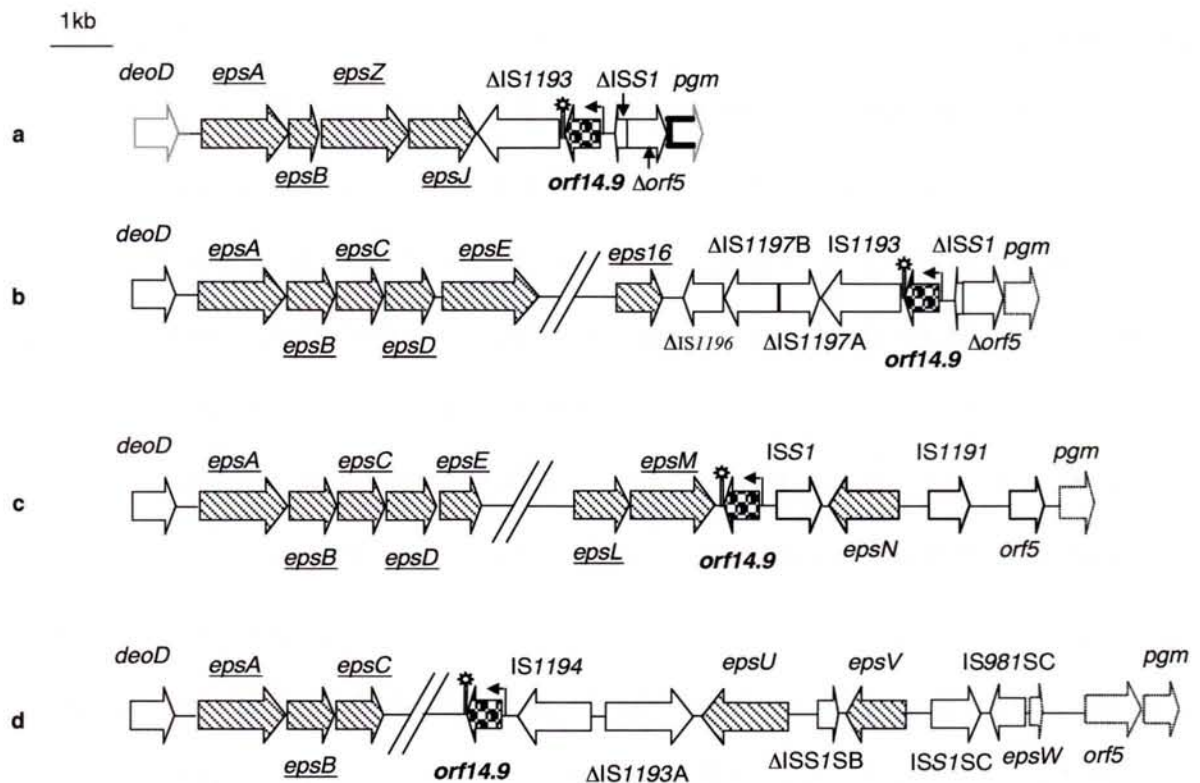


Figure 36 : Localisation d'*orf14.9* dans les clusters *eps* de *S. thermophilus* IP6756 (a), IP6757 et LMG18311 (b), CNRZ1066, MR-2C, MTC330, MTC360 et NST2280 (c), CNRZ368 et MR-1C (d).

Les clusters *eps* des souches regroupées présentent plus de 99% d'identité. Dans les autres clusters *eps*, aucune ORF impliquée dans la biosynthèse des EPS n'est observée en 3'd'*orf14.9* ou n'est séquencée. La séquence de *deoD* des souches MR-2C, MTC330 et MTC360 est partielle. Les promoteurs et terminateurs potentiels de transcription d'*orf14.9* sont symbolisés par une flèche et  respectivement. Les ORF hachurées sont impliquées dans la production d'EPS. Le nom des ORF opéroniques qui appartiennent à la première région des clusters *eps* sont soulignées.

Δ correspond à des ORF partiellement délétées ou à des éléments IS dont la transposase est partiellement délétée. Les deux lignes parallèles ne respectent pas l'échelle. Les ORF représentées en pointillés sont des pseudo-ORF. Les séquences figurées en grisé dans le cluster *eps* de la souche IP6756 sont hypothétiques.

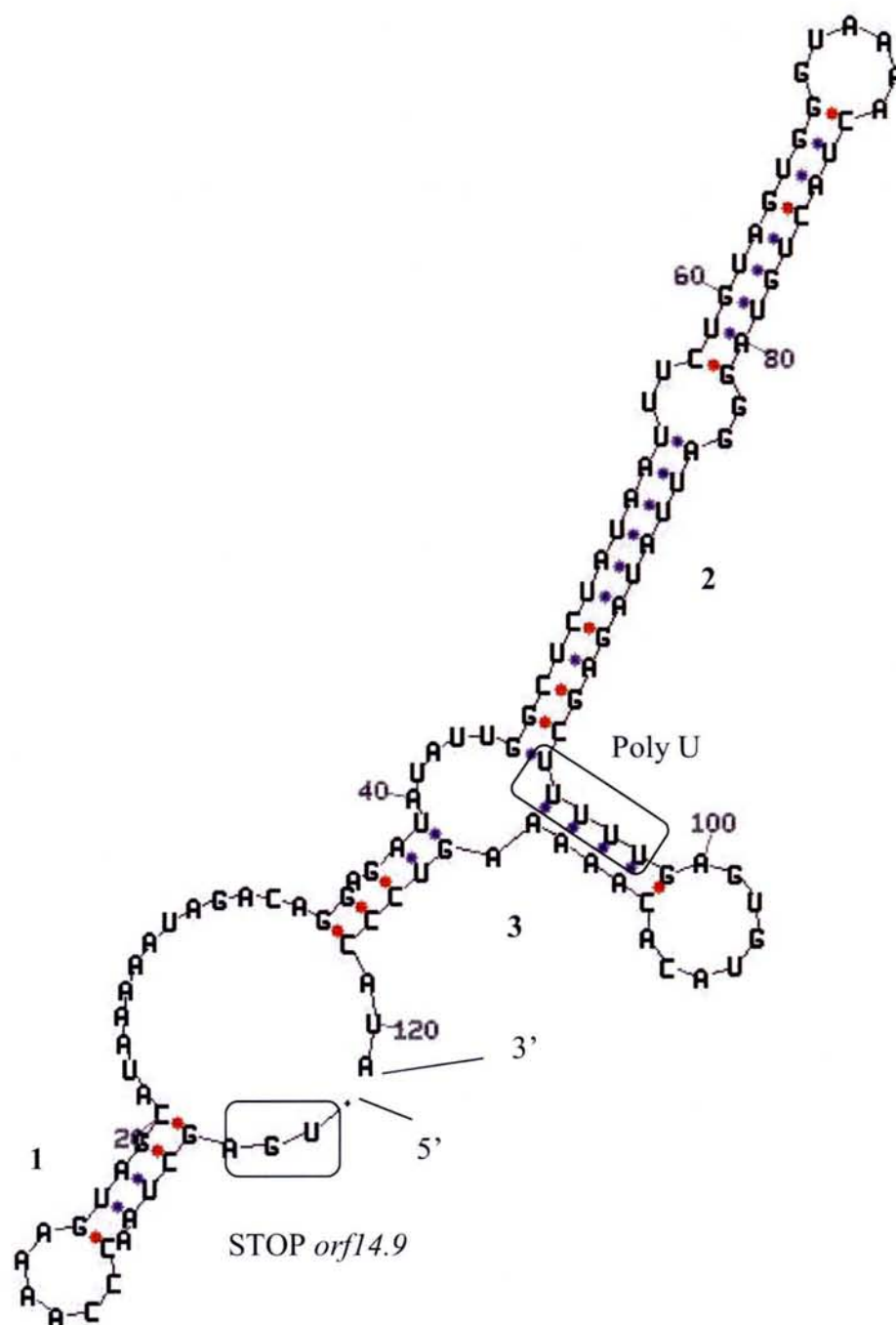


Figure 37 : Structure du terminateur potentiel de transcription identifié en 3' d'*orf14.9* chez IP6756 à l'aide du logiciel mfold3.0.

Trois structures tiges boucles (1, 2 et 3) sont identifiées, la seconde correspondant potentiellement à un terminateur rho-indépendant de transcription se terminant par des bases uraciles (poly U).

STOP *orf14.9* : codon d'arrêt de la traduction d'*orf14.9*.

Tableau 2 : Caractéristiques des allèles d'*orf14.9* des clusters *eps* de *S. thermophilus*.

Souche ou type de cluster <i>eps</i>	Numéro d'accession	Identité ^c (%)	Taille (pb)	Référence
NST2280	AF517575		576	Pluvinet <i>et al.</i> 2004
CNRZ1066	NC_006449	100	576	Bolotin <i>et al.</i> 2004
CNRZ368	Z98171	97	570	Bourgoin <i>et al.</i> 1999
IP6756	AY916787	97	576	ce travail
IP6757	AJ289861	97	576	Pluvinet <i>et al.</i> 2004
LMG18311	NC_006448	97	576	Bolotin <i>et al.</i> 2004
MR-1C	AF448249	97	570	Broadbent <i>et al.</i> 2003
MR-2C	AY061649	100	576	Broadbent <i>et al.</i> 2003
MTC310	AF448502	97	576	Broadbent <i>et al.</i> 2003
MTC330	AF430847	100	576	Welker <i>et al.</i> ^d
MTC360	AF434993	100	576	Broadbent <i>et al.</i> 2003
NCFB2393	Y17900	97	576	Almiron-Roig <i>et al.</i> 2000
Sfi39 ^a	AF373595	97	ND	Germond <i>et al.</i> 2001
Sfi6	U40830	99	576	Stingele <i>et al.</i> 1996
type I ^{a,b}	AF410175	97	ND	Rallu <i>et al.</i> ^d
type III ^b	AY057915	97	576	Rallu <i>et al.</i> ^d
type IV ^b	AF454495	97	570	Rallu <i>et al.</i> ^d
type V ^b	AF454496	97	564	Rallu <i>et al.</i> ^d
type VI ^{a,b}	AF454497	96	ND	Rallu <i>et al.</i> ^d
type VII ^b	AF454498	96	564	Rallu <i>et al.</i> ^d
type IX ^{a,b}	AF454499	100	ND	Rallu <i>et al.</i> ^d
type X ^b	AF454500	97	558	Rallu <i>et al.</i> ^d

^a : La séquence disponible est partielle.

^b : Les souches dans lesquelles ces clusters ont été caractérisés ne sont pas précisées dans les banques de données.

^c : Pourcentage d'identité nucléotidique avec l'allèle de la souche NST2280.

^d : Données non publiées.

ND : non déterminé car la séquence d'*orf14.9* disponible est partielle.

hétérogénéité de longueur de séquence peut être observée. La plupart des allèles *orf14.9* de *S. thermophilus* présentent une taille de 576 nucléotides (Tableau 2). Cependant, certains allèles présentent une longueur de 570 nucléotides (MR-1C, CNRZ368 et cluster *eps* de type IV), 564 nucléotides (clusters *eps* de type V et VII) ou 558 nucléotides (cluster *eps* de type X). La différence de longueur entre les allèles de 576 et 564 nucléotides est due au choix du codon start de traduction effectué par les auteurs : un codon TTG dans les allèles de 576 nucléotides, ou un codon ATG situé 12 nucléotides en aval dans les allèles de 564 nucléotides (Figure 38a). Les allèles de 570 et 576 nucléotides se différencient par la présence d'une délétion d'un nucléotide qui se traduit par un décalage du cadre de lecture au niveau de l'extrémité 3' de la séquence de l'ORF et génère l'apparition d'un codon stop de traduction (TAG) à 6 pb en amont de celui des allèles de 576 nucléotides (Figure 38b). L'allèle de 558 nucléotides du cluster *eps* de type X s'explique par ce type de mutation mais aussi par le choix des auteurs du codon ATG comme start de traduction.

Les allèles de 576 nucléotides coderaient une protéine potentielle de 191 acides aminés, qui ne présente aucune homologie avec des protéines de fonctions connues présentes dans les banques de données. Des prédictions de structure secondaire réalisées à l'aide du logiciel SMART montrent que la protéine potentielle Orf14.9 présenterait 6 hélices transmembranaires entre les acides aminés 7 à 29, 44 à 63, 70 à 92, 112 à 134, 141 à 160 et 164 à 186, suggérant que cette protéine potentielle serait localisée au niveau de la membrane. Cependant, la protéine potentielle Orf14.9 ne semble pas posséder de peptide signal d'après le serveur SignalP3.0. Les recherches de domaines effectuées (programme CDART) permettent de constater que les acides aminés 99 à 191 de la protéine potentielle correspondrait à un domaine VanZ (E-value : $4 \cdot 10^{-12}$). La protéine VanZ confère une faible résistance à l'antibiotique téicoplanine chez *Enterococcus faecium* (Arthur *et al.*, 1995), mais le mécanisme de résistance impliqué

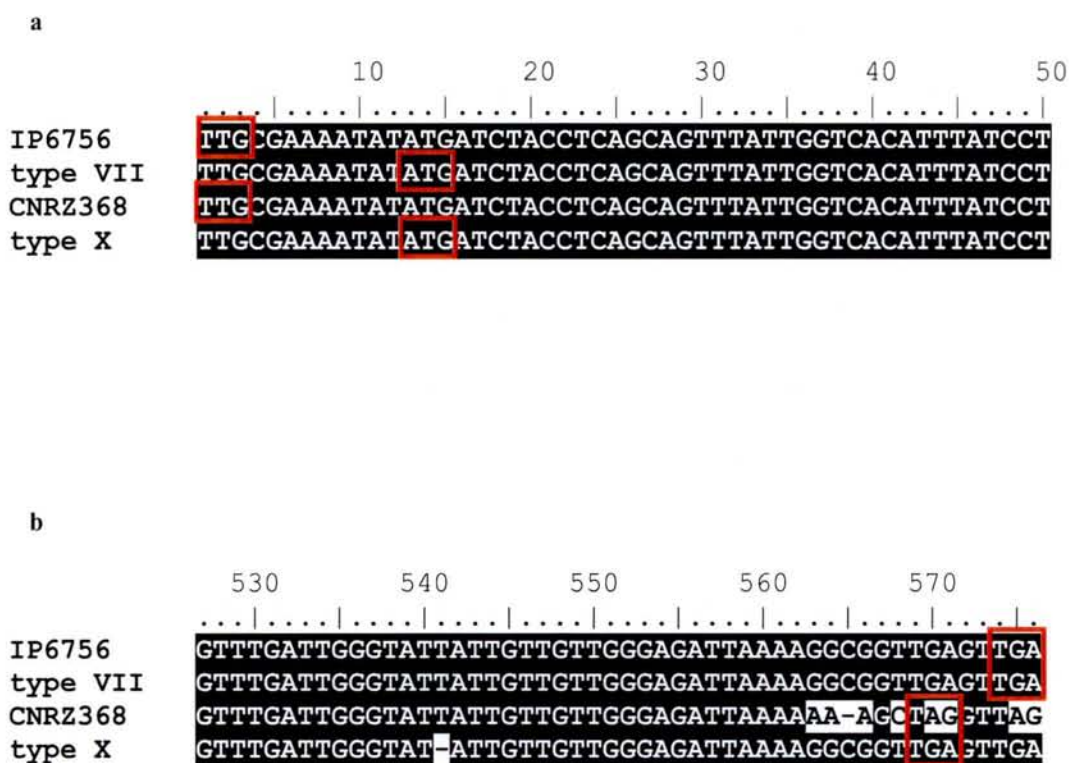


Figure 38 : Alignement des extrémités 5' (a) et 3' (b) des allèles *orf14.9* de IP6756 (576 bp), CNRZ368 (570 bp), de type VII (564 bp) et de type X (558 bp).

Le codon start (a) et stop (b) de traduction sont encadrés. Les nucléotides partagés par au moins la moitié des séquences sont écrits sur fond noir. Les tirets correspondent à des décalages introduits lors de l'alignement.

est à ce jour indéterminé (Walsh & Amyes, 2004) et en particulier le rôle du domaine VanZ identifié.

3.2 Activité transcriptionnelle d'*orf14.9*

La transcription d'*orf14.9* a été recherchée chez *S. thermophilus* dans des cultures réalisées en milieu M17 et au début de la phase stationnaire. Les ARN totaux de trois souches (IP6756, A054 et NST2280) sont extraits, puis des ADN complémentaires sont générés et criblés par RT-PCR. Elle a permis d'identifier un produit de RT-PCR de 535 pb correspondant à un transcrit spécifique à *orf14.9* dans ces trois souches (Figure 39).

Rappelons que la séquence promotrice entre les différents allèles présente une très forte conservation et que la séquence de l'ORF d'*orf14.9* est également très conservée. L'ensemble de ces observations ainsi que les résultats obtenus lors de l'étude de la transcription d'*orf14.9* chez les souches IP6756, NST2280 et A054 suggèrent qu'*orf14.9* serait transcrite chez les autres souches de *S. thermophilus*.

3.3 Etude de la fonction d'*orf14.9* chez *S. thermophilus* A054 et NST2280

L'ensemble de ces résultats suggère que cette ORF est transcrite au sein des clusters *eps* de *S. thermophilus*, et que sa fonction est importante, mais inconnue. L'approche expérimentale suivie pour étudier cette fonction repose sur l'interruption d'*orf14.9*, puis la caractérisation phénotypique des mutants obtenus. La souche IP6756 étant difficilement transformable, les études ont été effectuées chez *S. thermophilus* A054 et NST2280, deux souches également productrices d'EPS puisqu'elles produisent respectivement 118 mg/l (± 10 mg/l) et 170 mg/l (± 36 mg/l) d'EPS.

Le plasmide pGHost9*orf14.9* qui contient un fragment interne d'*orf14.9* de 451 pb a permis d'interrompre *orf14.9* du cluster *eps* des souches NST2280 et A054. Les clones

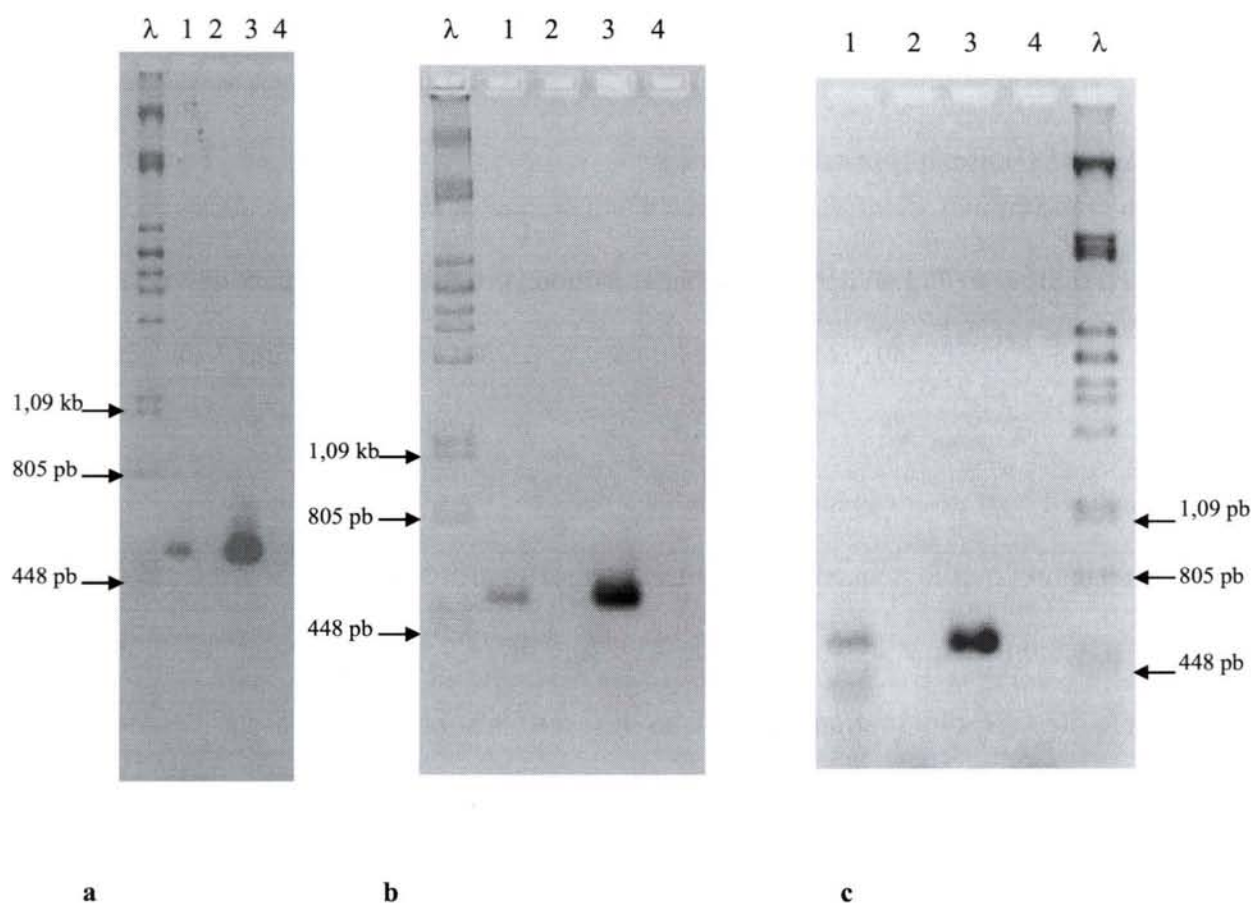


Figure 39 : Détection d'un transcrit spécifique à *orf14.9* chez *S. thermophilus* IP6756 (a), A054 (b), et NST2280 (c) PCR.

1 : ADNc obtenus par action de la transcriptase inverse

2 : ARN totaux extraits de *S. thermophilus* IP6756 (a), A054 (b) ou NST2280 (c) ayant subi le traitement de la DNase

3 : ADN génomique de *S. thermophilus* IP6756 (a), A054 (b) ou NST2280 (c)

4 : sans matrice d'acides nucléiques

λ : ADN du phage λ digéré par *Pst*I

Electrophorèse réalisée sur un gel contenant 2% (IP6756 et A054) ou 1,5% d'agarose (NST2280).

obtenus NST2280-O7 (Figure 40) et A054-O12 présentent une intégration multiple du plasmide et sont retenus pour les analyses phénotypiques ultérieures.

3.3.1 Etude de l'implication d'*orf14.9* dans la biosynthèse des EPS

Les EPS produits par les mutants NST2280-O7 et A054-O12 sont extraits puis dosés afin de comparer leur taux de production à celui des souches sauvages. Chez ces mutants, le taux d'EPS produit est de 148 mg/l ($\pm$ 34 mg/l) pour NST2280-O7 et 124 mg/l ($\pm$ 7 mg/l) pour A054-O12 (Figure 41). Ces valeurs ne présentent pas de différence statistiquement significative par rapport à leurs souches sauvages respectives. Les mutants réalisés ne présentent donc pas de modification de la quantité d'EPS produits. Ce résultat suggère qu'*orf14.9* muté ne modifie pas la quantité d'EPS produit chez *S. thermophilus* dans les souches mutantes réalisées.

3.3.2 Implication d'*orf14.9* dans la croissance cellulaire de *S. thermophilus*

La croissance cellulaire des mutants NST2280-O7 et A054-O12 et des souches sauvages respectives a été suivie et a permis d'établir des courbes de croissance. Les résultats présentés ont été reproduits trois fois. Ces courbes de croissance permettent de déterminer les paramètres de croissance de ces souches, soit le taux de croissance spécifique (μ) et du temps de génération (g). La comparaison des valeurs obtenues chez les mutants et les souches sauvages respectives montre que les mutants présentent un taux de croissance spécifique significativement plus réduit que les souches sauvages (Tableau 3). A l'inverse, le temps de génération des mutants est significativement plus élevé que celui des souches sauvages respectives (Tableau 3).

Par ailleurs, les courbes obtenues pour les souches A054 et A054-O12 sont similaires à celles obtenues pour les souches NST2280 et NST2280-O7 qui sont présentées dans la

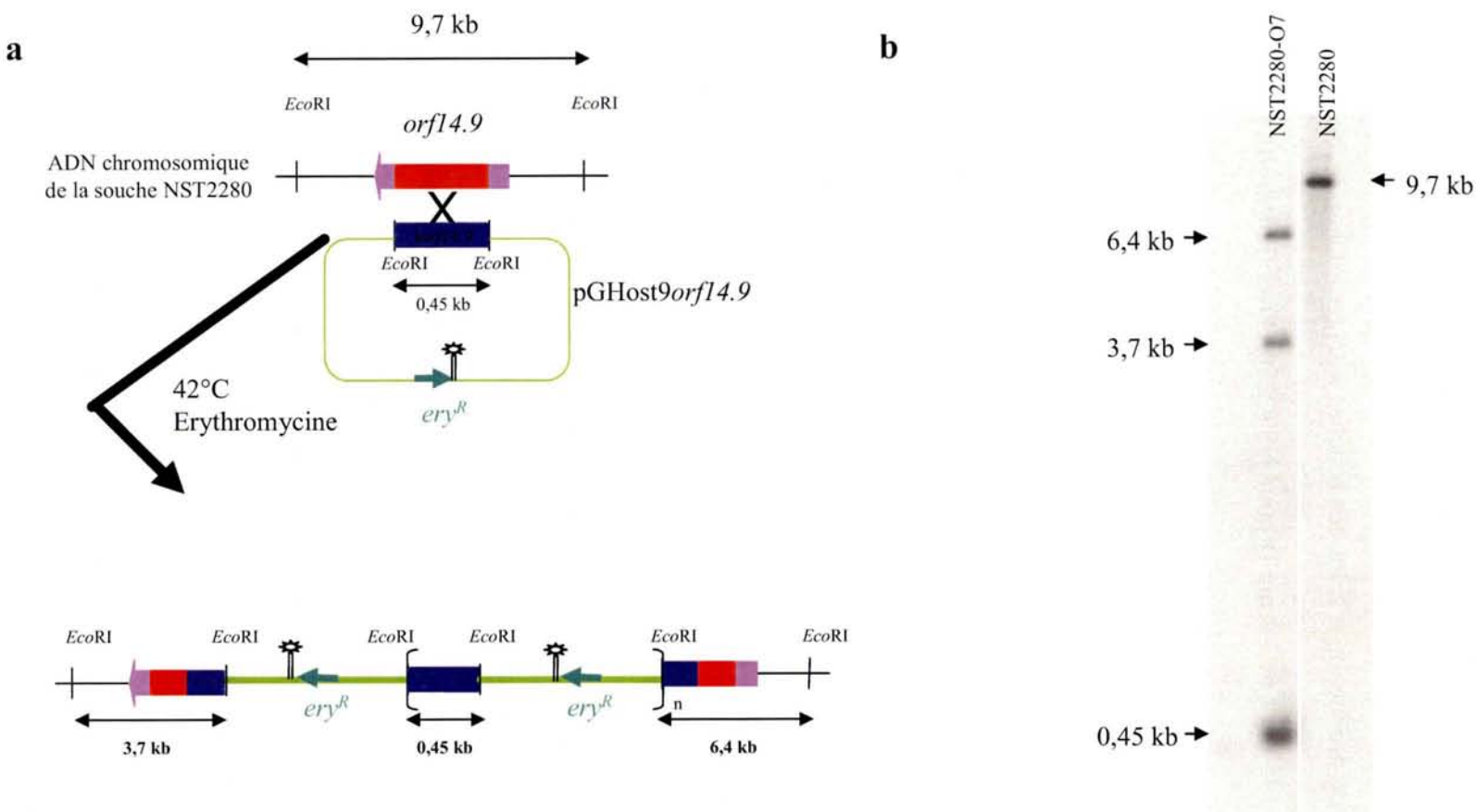


Figure 40 : Caractérisation du clone NST2280-O7 obtenu par intégration multiple du plasmide pGHost9*orf14.9* dans le chromosome de *S. thermophilus* NST2280 par recombinaison homologue dans *orf14.9*.

a: Mécanisme d'intégration multiple du plasmide pGHost9*orf14.9* dans le chromosome de *S. thermophilus* NST2280 par recombinaison homologue dans *orf14.9*.

EcoRI : site de restriction *EcoRI*. *ery^R* : gène de résistance à l'érythromycine.

orf14.9: fragment interne à *orf14.9*. : terminateur de transcription du gène *ery^R*

[]n: n copies du plasmide

b: Hybridation du fragment interne à *orf14.9* sur un profil de restriction *EcoRI* de l'ADN génomique de *S. thermophilus* NST2280 et de l'intégrant NST2280-O7.

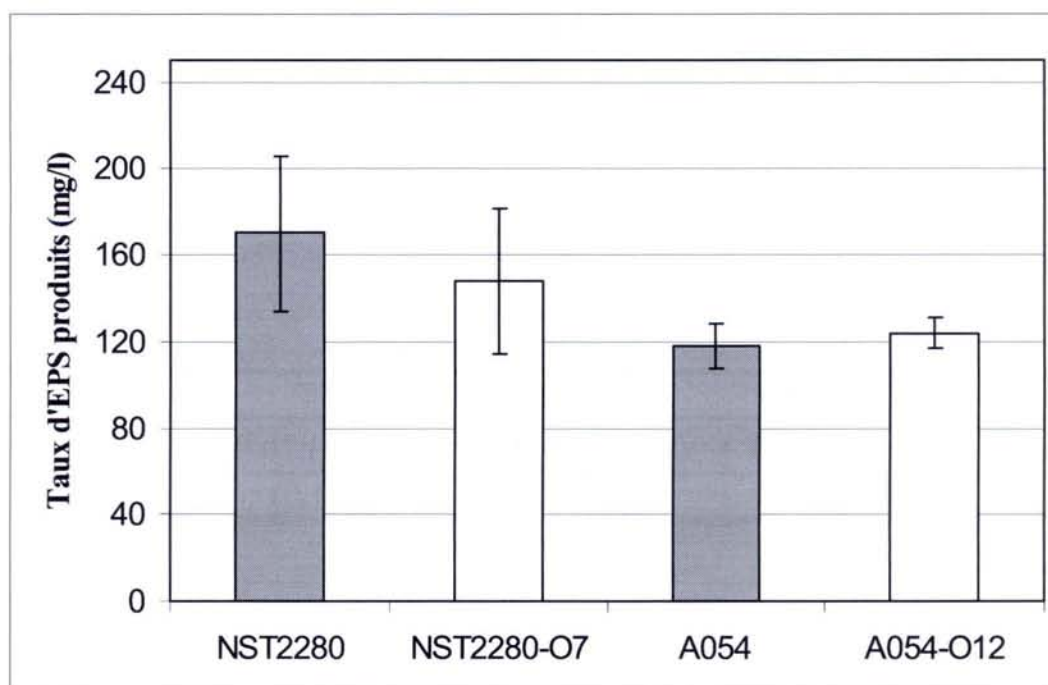


Figure 41 : Détermination du taux d'EPS produit par *S. thermophilus* NST2280 et A054 (barres grises) et les mutants NST2280-O7 et A054-O12 (barres blanches).

Le taux de production d'EPS a été déterminé quatre fois pour les souches A054, A054-O12 ainsi que NST2280-O7 et neuf fois pour la souche NST2280. Les écarts-types de chaque valeur sont indiqués.

Tableau 3 : Paramètres de croissance des souches NST2280 et A054 et de leurs mutants insertionnels d'*orf14.9* respectifs NST2280-O7 et A054-O12.

	NST2280	NST2280-O7	A054	A054-O12
μ	0.016 ($\pm 2.10^{-3}$)	0.012 ($\pm 1.10^{-3}$)	0.026 ($\pm 1.10^{-3}$)	0.021 ($\pm 1.10^{-3}$)
g	43 (± 7.5)	58 (± 5.2)	27 (± 1.0)	33 (± 0.8)

μ : taux de croissance spécifique (/min)

g : temps de génération (min)

L'erreur standard est indiquée pour chaque valeur.

figure 42. Ces courbes permettent aussi d'observer une entrée en phase stationnaire plus tardive chez les mutants NST2280-O7 et A054-O12 que chez leurs souches sauvages respectives. L'ensemble de ces résultats suggère qu'*orf14.9* est impliquée dans la croissance cellulaire de *S. thermophilus*.

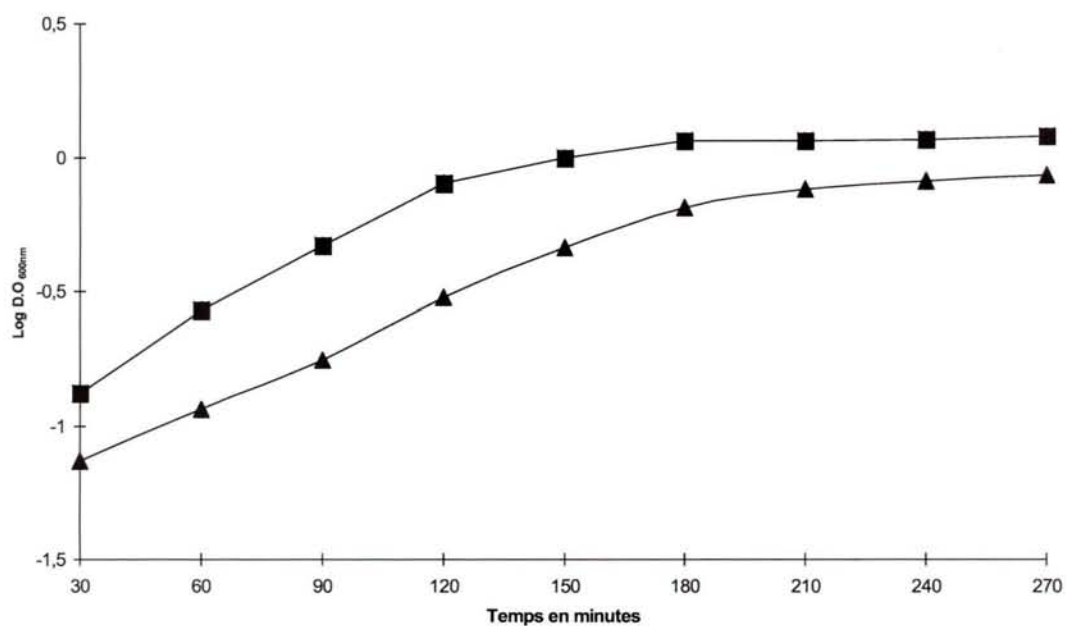


Figure 42 : Courbe de croissance de *S. thermophilus* NST2280 (carré) et de son mutant insertionnel d'*orf14.9* NST2280-O7 (triangle).

Les résultats présentés ont été reproduits trois fois.

DISCUSSION-PERSPECTIVES

Le cluster *eps* des souches de *S. thermophilus* peut être divisé en deux régions. La première région présente une organisation opéronique et modulaire : un premier module regroupe les gènes impliqués dans différents niveaux de régulation de la production d'EPS et un second module contient les gènes impliqués dans la biosynthèse des EPS, dont ceux codant les glycosyltransférases, enzymes nécessaires à la formation des unités répétées (Jolly & Stingle, 2001). Le premier module est fortement conservé dans tous les clusters *eps*, à l'exception de celui de la souche IP6756, alors que la variabilité du second module est associée à la variabilité des EPS produits par *S. thermophilus*.

La première région du cluster *eps* de *S. thermophilus* IP6756 présente une organisation modulaire se différenciant nettement du modèle décrit précédemment. Ainsi, le module de régulation n'est constitué que des deux premières ORF habituellement rencontrées dont l'une, *epsB*, est tronquée en 3'. De plus, aucun gène codant de glycosyltransférase ne peut être identifié dans le second module alors que la souche est productrice d'EPS. L'absence de gènes codant des glycosyltransférases au sein du cluster *eps* ainsi que l'étude fonctionnelle de ce cluster amènent à s'interroger sur l'implication de gènes localisés hors du cluster *eps* identifié dans la biosynthèse des EPS de la souche IP6756.

La seconde région des clusters *eps* de *S. thermophilus* est constituée des ORF situées immédiatement en aval de la première région : d'*orf14.9* à *pgm* (Pluvinet *et al.*, 2004). L'étude de cette seconde région a permis de montrer que, chez la souche NST2280, une de ces ORF (*epsN*) est impliquée dans la biosynthèse des EPS puisque son interruption s'accompagne d'une diminution drastique de la production d'EPS (Pluvinet *et al.*, 2004). Ce résultat amène à s'interroger sur la fonction des ORF de cette seconde région et en particulier

d'*orf14.9*, qui est présente dans tous les clusters *eps* séquencés et de *pgm*, qui constituerait une frontière du cluster *eps*.

Le cluster *eps* de *S. thermophilus* serait ainsi constitué d'un opéron regroupant l'essentiel des gènes de biosynthèse des EPS et d'autres gènes organisés différemment mais regroupés à proximité immédiate de celui-ci. Cette organisation composite du cluster *eps* chez *S. thermophilus* nous a amené à privilégier le terme de cluster (littéralement rassemblement) plutôt que celui de locus, afin d'insister sur le caractère dynamique du rassemblement de gènes.

Les gènes de la première région présentent un rapprochement physique et fonctionnel puisqu'ils sont organisés en opéron. Ce rapprochement, en rendant possible le transfert horizontal, tend à augmenter la probabilité d'acquisition par une autre bactérie présente dans la même niche écologique, de l'ensemble de l'information génétique nécessaire à la biosynthèse de l'EPS produit par la souche donneuse. La sélection naturelle s'exerçant sur le déterminisme génétique de la biosynthèse des EPS pourrait également être à l'origine du regroupement des ORF constituant la seconde région du cluster *eps* de *S. thermophilus*.

L'intervention d'événements de transfert horizontaux a été avancée pour expliquer que le cluster *eps* de *S. thermophilus* constitue une des régions extrêmement variables du chromosome de cette bactérie. Ainsi, le pourcentage en G+C du cluster *eps* de *S. thermophilus* CNRZ368 varie de 26,2 à 41,9% selon les ORF (Bourgoin *et al.*, 1999), alors que celui du génome total est compris entre 37,2 et 39,8% (Farrow & Collins, 1984). Ceci suggère que plusieurs gènes de ce cluster auraient été acquis par transfert horizontal à partir d'autres espèces bactériennes. En particulier, une région de 13,6 kb contenant plusieurs copies d'*IS981* et d'*ISS1* ainsi que deux ORF quasiment identiques à *epsL* et *orfY* du cluster *eps* de *Lactococcus lactis* NIZOB40 aurait été transférée entre *L. lactis* et *S. thermophilus* (Bourgoin *et al.*, 1999). Ces constatations amènent à s'interroger sur l'origine des séquences constituant

le cluster *eps* de la souche IP6756 et sur le mode d'acquisition de séquence d'ADN exogène impliquée dans la biosynthèse des EPS. Cette discussion présentera notamment des résultats issus du séquençage du génome de *S. thermophilus* qui permettent de proposer des hypothèses.

L'identification de la ou des pression(s) de sélection s'exerçant au niveau du cluster *eps* de *S. thermophilus* constitue une des questions les plus fondamentales concernant les EPS de *S. thermophilus*. Une autre bactérie du genre *Streptococcus*, *S. pneumoniae*, présente des clusters *cps* fortement homologues à ceux de *S. thermophilus* et une variabilité comparable. Chez cette bactérie pathogène, les polysaccharides de capsule synthétisés par ces clusters jouent un rôle fondamental dans la virulence, en conférant à la bactérie une protection face au système immunitaire de l'hôte. La variabilité structurale des capsules résultant de la variabilité du cluster *cps* constitue une adaptation à la production de nouvelles défenses immunitaires par l'hôte infecté. L'existence d'une telle variabilité au niveau des clusters *cps* représente donc un avantage sélectif pour cette bactérie. L'observation d'une variabilité comparable au niveau des clusters *eps* de *S. thermophilus* suggère qu'elle est également essentielle à cette bactérie. Cette discussion présentera les hypothèses qui peuvent être avancées chez cette bactérie qui n'est pas pathogène.

1 DES GENES EXTERIEURS AU CLUSTER *EPS* DE LA SOUCHE IP6756 SONT NECESSAIRES A LA BIOSYNTHESE DES EPS DE CETTE SOUCHE

1.1 Localisation et frontières du cluster *eps* de *S. thermophilus* IP6756

L'hybridation d'une sonde spécifique de *deoD* et d'*epsA* de *S. thermophilus* IP6757 révèle un fragment *EcoRI* de 6,4 kb de l'ADN génomique de la souche IP6756 et de 25 autres souches de *S. thermophilus* (Pluvinet, 2003), indiquant que les gènes *deoD* et *epsA* sont localisés dans la même région du chromosome chez l'ensemble des souches testées. Ainsi, le cluster *eps* de la souche IP6756 est localisé au même endroit que les clusters *eps* des autres souches, c'est-à-dire en 3' de *deoD*.

L'ORF *deoD*, qui code une phosphorylase de nucléosides puriques, intervient potentiellement dans le métabolisme des nucléotides et non dans la biosynthèse des EPS. Elle constitue ainsi la frontière gauche du cluster *eps* de la souche IP6756, comme dans les autres clusters *eps* de *S. thermophilus* (Broadbent *et al.*, 2003).

La frontière droite du cluster *eps* de la souche IP6756 est constituée par la pseudo-ORF *pgm*, comme pour les autres clusters *eps* de *S. thermophilus* (Pluvinet *et al.*, 2004). L'interrogation des séquences protéiques des banques de données révèle que le produit virtuel de cette ORF, retrouvée à l'état de pseudogène dans tous les clusters *eps* analysés chez cette espèce, présente des pourcentages d'identités significatifs avec des phosphoglycérate mutases. La phosphoglycérate mutase catalyse la conversion (réversible) du 3-phosphoglycérate en 2-phosphoglycérate lors des dernières étapes de la glycolyse et est associée directement à la production d'ATP *via* la glycolyse. L'analyse de la séquence du chromosome de *S. thermophilus* CNRZ1066 et LMG18311 permet d'identifier d'autres gènes codant potentiellement une phosphoglycérate mutase fonctionnelle : *gpmA* (693 pb) et *gpmC* (639 pb) pour CNRZ1066 et LMG18311 et *gpmB* (705 pb) pour CNRZ1066. L'activité de phosphoglycérate mutase nécessaire à la glycolyse serait ainsi codée par un ou plusieurs

gènes différents de l'ORF *pgm*. Cette ORF ne serait donc pas impliquée dans la biosynthèse des EPS et constituerait la frontière droite du cluster *eps* de la souche IP6756.

L'observation en aval de *pgm* d'une ORF (*epsX*) potentiellement impliquée dans la translocation des EPS, dans les clusters *eps* des souches LMG18311, MTC360, MR-1C, MR-2C, MTC330 et CNRZ1066, nuance cette affirmation. Les séquences des différents allèles de ce gène présentent au moins 99% d'identité. Immédiatement en 3' d'*epsX* sont localisés des ORF codant une glucosidase tronquée et un transporteur ABC, qui ne sont pas impliqués dans la biosynthèse des EPS. L'ORF *epsX*, si elle est effectivement impliquée dans la translocation des EPS chez les souches correspondantes, pourrait être également affectée par le mécanisme général conduisant au regroupement des gènes impliqués dans la biosynthèse des EPS.

1.2 Fonctionnalité de la région opéronique du cluster *eps* de *S. thermophilus* IP6756

1.2.1 Activité transcriptionnelle et traductionnelle

L'analyse de la séquence de la région intergénique *deoD-epsA* et d'*epsA* a permis d'identifier trois promoteurs de transcription potentiels : P1, P2 et P3. Le promoteur P1 est homologue au promoteur identifié en amont d'*epsA* dans les autres clusters *eps* de *S. thermophilus*. La séquence des boîtes -35 et -10 est en effet très conservée entre les différents clusters et proche de leurs séquences consensus respectives. Ces deux boîtes sont situées en amont d'un RBS dont la séquence est très proche de la séquence consensus et qui se trouve à 7 nucléotides du codon start de traduction d'*epsA* (CF paragraphe 3.1 des résultats).

Les promoteurs P2 et P3, internes à *epsA*, se distinguent du promoteur P1 par l'absence d'un RBS en aval de la boîte -10 dans le premier cas et par un RBS de séquence très éloignée de la séquence consensus (Rocha *et al.*, 1999) dans le second cas.

Le produit de PCR obtenu lors d'expériences de RT-PCR avec des amorces localisées en 5' du promoteur P3 suggère que P1 est fonctionnel, comme cela a été observé dans le cluster

eps de la souche Sfi6 (Stingele *et al.*, 1999b). Il reste toutefois possible que P2 et P3 soient fonctionnels et assurent également la transcription du cluster *eps* de la souche IP6756 en tronquant la transcription de l'ORF *epsA*. Afin de le vérifier, des expériences de Northern Blot couplées à une analyse par extension d'amorce et/ou nucléase S1 pourraient être envisagées.

Une analyse portant sur les RBS de *Bacillus subtilis* (Rocha *et al.*, 1999), une bactérie appartenant également à la division des Firmicutes, permet de discuter de la réalité fonctionnelle du promoteur potentiel P3. L'analyse des gènes de cette bactérie révèle que 90% d'entre eux présentent un RBS localisé entre 5 et 11 nucléotides en amont du codon start de traduction, mais l'espacement idéal serait de 7 à 9 nucléotides (Rocha *et al.*, 1999). Dans le cas du promoteur P1, un espacement de 7 nucléotides est observé entre le RBS et le codon start de traduction d'*epsA*, alors qu'il est de 13 nucléotides dans le cas du promoteur P3.

Par ailleurs, l'extrémité 3' des ARNr16S interagit avec le RBS lors de la traduction. Cette interaction par appariement de bases constitue une étape critique dans l'initiation de la synthèse protéique. Chez *E. coli*, la protéine S1 joue un rôle crucial dans l'appariement des ARN messagers à la sous-unité 16S du complexe ribosomique. Chez un grand nombre d'eubactéries à coloration de Gram positive, dont *B. subtilis*, la protéine S1 est absente, d'où la nécessité d'une grande conservation de la séquence consensus du RBS (AAAGGAGG), qui permet de faciliter l'appariement des ARN messagers à la sous-unité ribosomique 16S. Chez *B. subtilis*, seuls 55 gènes présentent un RBS éloigné de la séquence consensus, dont vingt codent des protéines potentielles homologues à des protéines présentes dans la banque de données Swissprot alors que les autres sont des pseudo-gènes (Rocha *et al.*, 1999). L'ensemble de ces données suggère que les promoteurs de transcription potentiels P2 et P3 ne soient pas fonctionnels. Une démonstration expérimentale devra être envisagée.

Un terminateur de transcription potentiel a été identifié en aval d'*epsJ*, dernière ORF de la première région du cluster *eps* d'IP6756. Il présente deux structures tiges boucles, la seconde correspondant potentiellement à un terminateur de transcription rho-indépendant (CF paragraphe 3.1 des résultats).

L'étude de la transcription des 4 ORF (*epsA*, *epsB*, *epsZ* et *epsJ*) appartenant à la première région du cluster *eps* de la souche IP6756 a permis d'identifier un transcrit spécifique de chacune de ces ORF. De plus, l'identification de transcrits de jonction entre *epsA* et *epsB*, *epsB* et *epsZ* ainsi qu'entre *epsZ* et *epsJ* suggère que ces 4 ORF présentent une organisation opéronique. Une telle organisation avait été décrite chez *S. thermophilus* Sfi6, autre souche productrice d'EPS de l'espèce (Stingele *et al.*, 1999b).

Les quatre gènes constituant cet opéron seraient potentiellement traduits. Ils sont en effet séparés par de courtes séquences intergéniques, voir juxtaposés dans le cas d'*epsA* et d'*epsB*, ce qui suggère un couplage traductionnel. Cette hypothèse est notamment avancée lors de l'étude du cluster *eps* de la souche NCFB2393 (Almiron-Roig *et al.*, 2000).

1.2.2 Régulation de la transcription de l'opéron du cluster *eps* de *S. thermophilus* IP6756

Au sein de la séquence promotrice identifiée en amont d'*epsA* de *S. thermophilus* IP6756, deux éléments CRE potentiels peuvent être identifiés. L'élément CRE potentiel qui chevauche la boîte -35 du promoteur P1 du cluster *eps* de la souche IP6756 est retrouvé dans la région promotrice du cluster *eps* de neuf autres souches analysées parmi les souches de *S. thermophilus* dont la séquence du cluster *eps* est disponible.

Chez deux autres bactéries appartenant au genre *Streptococcus* (*S. pneumoniae* et *S. mutans*), de tels éléments sont observés en amont des gènes subissant une régulation catabolique par le carbone. Chez *S. pneumoniae* et *S. mutans*, la protéine RegM et son homologue CcpA se fixent sur les éléments CRE et agissent comme un régulateur global du

métabolisme glucidique et comme un effecteur primaire de la répression catabolique par le carbone (Giammarinaro & Paton, 2002), (Browngardt *et al.*, 2004). L'inactivation du gène codant ces protéines s'accompagne d'une réduction drastique de l'expression du cluster *cps* chez *S. pneumoniae* et du cluster *gtfBC* et du gène *ftf* chez *S. mutans*, qui présentent des éléments CRE localisés en amont. Ces résultats suggèrent fortement que ces protéines agissent comme un activateur de la transcription de *gtfBC* et *ftf* chez *S. mutans* ainsi que celle du cluster *cps* chez *S. pneumoniae*. Les travaux réalisés sur les protéines CcpA et RegM suggèrent qu'elles exercent un rôle d'interface entre le métabolisme glucidique et la production d'EPS/CPS. Il a ainsi été montré que la transcription de *gtfBC* et de *ftf* peut être augmentée de 300% et 50% respectivement quand *S. mutans* est cultivé en chimostat en présence de saccharose (Wexler *et al.*, 1993).

Chez *S. thermophilus* CNRZ302, un gène codant une protéine potentielle CcpA présentant 66% d'identité en acides aminés avec la protéine RegM de *S. pneumoniae* a été identifié. Il est également impliqué dans la répression catabolique par le carbone (van den Bogaard *et al.*, 2000). A ce jour, l'intervention de cette protéine dans la production d'EPS chez *S. thermophilus* n'a pas été montrée. L'identification d'éléments CRE potentiels dans la région promotrice du cluster *eps* de *S. thermophilus* suggère que la protéine CcpA pourrait être nécessaire à une expression optimale du cluster comme observé chez *S. pneumoniae* et *S. mutans*. Une telle étude permettrait d'acquérir une vision plus globale des relations existant dans la cellule entre le métabolisme glucidique et la production d'EPS. La figure 43 présente les voies métaboliques impliquées dans le catabolisme du lactose chez *S. thermophilus*, dont les liens entre la glycolyse et la biosynthèse des EPS. La compréhension des mécanismes de régulation des gènes impliqués dans la biosynthèse des EPS et des gènes soumis à la répression catabolique par le carbone constituerait un approfondissement intéressant. Il serait donc également intéressant de montrer expérimentalement si la protéine EpsA de

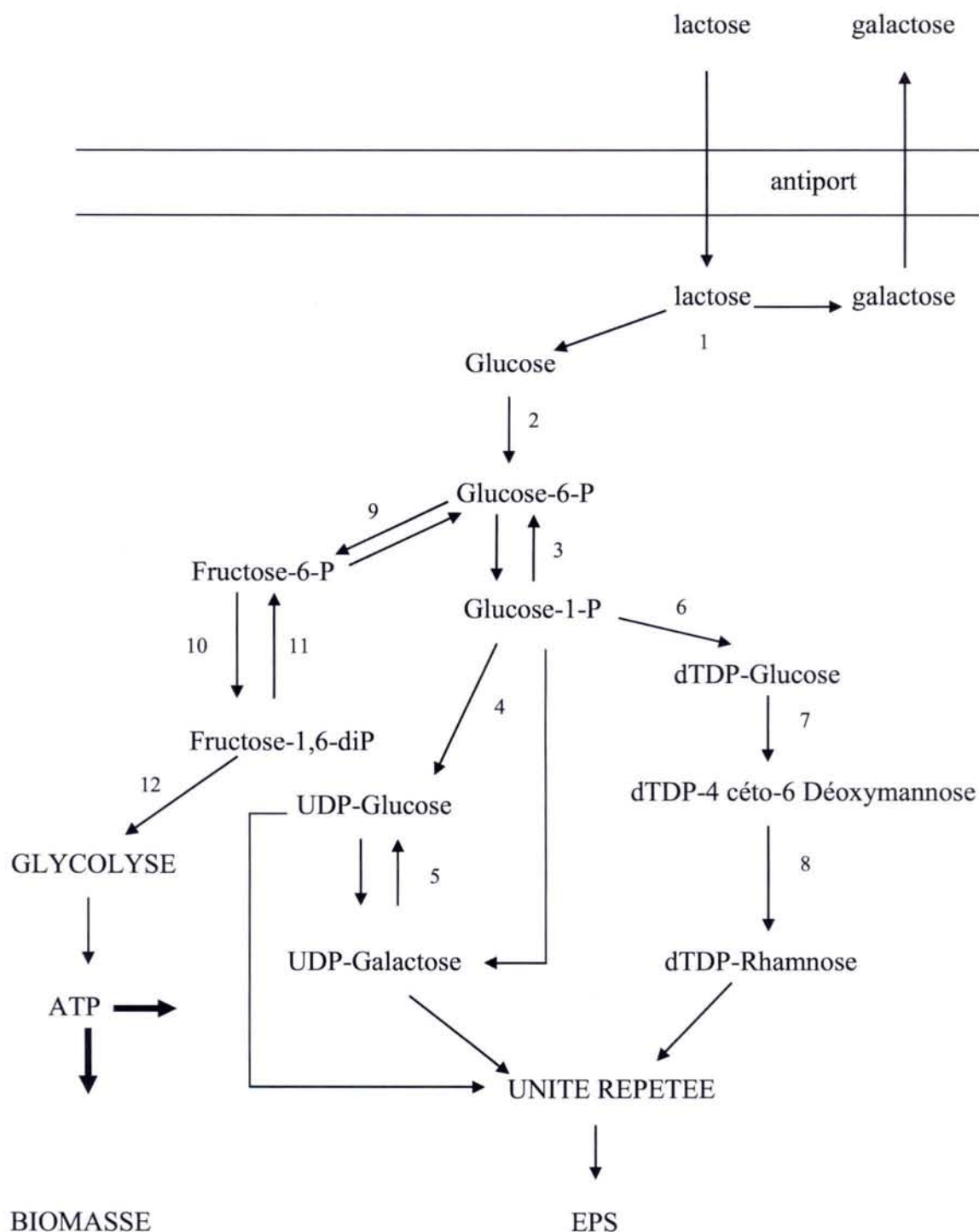


Figure 43 : Voies métaboliques impliquées dans le catabolisme du lactose chez *S. thermophilus*, d'après De Vuyst et Degeest (1999).

1: β-galactosidase; 2: glucokinase; 3: phosphoglucomutase; 4: UDP-glucose pyrophosphorylase; 5: UDP-galactose-4-épimérase; 6: dTDP-glucose pyrophosphorylase; 7: déhydratase; 8: épimérase réductase; 9: phosphoglucose isomérase; 10: 6-phosphofructokinase; 11: fructose-1,6-biphosphatase; 12: fructose-1,6-diphosphate aldolase.

S. thermophilus présente une fonction d'activateur transcriptionnel, comme cela a été montré pour les protéines CpsA de *S. pneumoniae* et de *S. agalactiae*, homologues de la protéine EpsA (Morona *et al.*, 2000) (Cieslewicz *et al.*, 2001).

1.3 Quels sont les gènes impliqués dans la biosynthèse des EPS chez *S. thermophilus* IP6756 ?

Le module de régulation de la première région du cluster *eps* de IP6756 se différencie nettement de ceux décrits dans les autres clusters de *S. thermophilus*. Seuls deux des quatre ORF de ce module (*epsA* et *epsB*) sont présentes. De plus, la séquence d'*epsB* est partiellement délétée par rapport à la séquence de tous les autres allèles d'*epsB*. Ces constatations amènent à considérer que le fonctionnement du module de régulation de la souche IP6756 puisse différer nettement de celui de *S. pneumoniae* qui sert de modèle pour les autres bactéries lactiques (Jolly & Stingeles, 2001). Chez cette bactérie, le module de régulation est également constitué de 4 ORF : *cpsA*, *cpsB*, *cpsC* et *cpsD*, qui sont respectivement homologues à *epsA*, *epsB*, *epsC* et *epsD* de *S. thermophilus*. Chez *S. pneumoniae*, le complexe constitué des protéines CpsC et CpsD, capable d'autophosphorylation, contrôle le taux de polymérisation des CPS selon l'état de phosphorylation de la protéine CpsD. La protéine CpsB possède une activité de phosphatase et déphosphoryle la protéine CpsD (Morona *et al.*, 2000) (Morona *et al.*, 2002), modifiant les interactions entre les protéines CpsC et CpsD (CF paragraphe 1.6.1 de l'introduction). Dans la souche IP6756, un tel modèle ne peut pas se concevoir du fait de l'absence d'*epsC* et d'*epsD*. De plus, l'activité phosphatase de la protéine potentielle EpsB n'est pas forcément conservée suite à la réduction importante de la taille de la séquence d'*epsB*.

La mutagénèse insertionnelle réalisée dans *epsA* provoque un effet polaire sur l'expression des ORF situées en aval (*epsB*, *epsZ* et *epsJ*) et par conséquent l'inactivation du cluster *eps* de *S. thermophilus* IP6756. Elle ne s'accompagne pas d'une diminution

statistiquement significative de la production d'EPS dans la souche mutante par rapport à la souche sauvage. Ce résultat montre que le cluster *eps* identifié dans la souche IP6756 n'est pas impliqué en terme quantitatif dans la production d'EPS de la souche, alors que les ORF constituant l'opéron sont transcrites et potentiellement traduites. Ce résultat suggère fortement que un ou plusieurs gène(s) extérieur(s) au cluster *eps* est (sont) nécessaire(s) à la biosynthèse d'EPS dans cette souche. Cependant, les produits potentiels de certaines ORF du cluster identifié pourraient intervenir de façon temporaire ou minoritaire dans la production d'EPS de la souche. Parmi les quatre ORF identifiées dans la première région, les ORF du module de régulation *epsA* et *epsB* ne sont vraisemblablement pas impliquées dans la diminution du taux moyen de production d'EPS chez le mutant. Par contre, l'implication d'*epsZ* pourrait être associée à cette observation. En effet, cette ORF coderait un transporteur potentiel d'unités répétées pouvant intervenir de façon mineure dans l'export des EPS de la souche. De plus, cette comparaison quantitative entre les souches sauvage et mutante ne donne pas d'indication sur une implication potentielle du cluster *eps* identifié dans la structure de l'EPS. Ainsi, la protéine potentielle EpsJ qui présenterait une activité de mutase pourrait catalyser la conversion d'oses d'une forme pyranose à une forme furanose.

L'absence de modification significative du taux de production d'EPS suite à l'inactivation du cluster *eps* de la souche est en accord avec l'observation qu'aucun gène codant une glycosyltransférase n'ait été identifié dans le cluster *eps* étudié. Ces constatations suggèrent l'intervention d'un ou plusieurs gène(s) extérieur(s) au cluster *eps* de la souche IP6756, et notamment de gènes codant des glycosyltransférases, enzymes indispensables à la biosynthèse des unités répétées constituant les EPS. Les molécules d'EPS produites par la souche IP6756 sont constituées de glucose, de galactose et de traces d'une autre molécule glucidique, suggérant que la synthèse de cet hétéropolysaccharide impliquerait plusieurs gènes extérieurs au cluster *eps* identifié.

Cette situation est à rapprocher de celle décrite chez *S. pneumoniae* type 37, qui contient un cluster *cps* identique à celui des souches de sérotype 33f mais pour lequel la phase de lecture de plusieurs ORF est interrompue par des mutations. Il a été montré expérimentalement que ce cluster n'intervient pas dans la biosynthèse de l'homopolysaccharide constituant la capsule de type 37. Celle-ci est assurée par le gène *tts* qui est situé à l'extérieur du cluster *cps* et qui code une synthase (Llull *et al.*, 1999) (Llull *et al.*, 2001). L'hybridation d'une sonde spécifique de *tts* sur l'ADN génomique de la souche IP6756 montre qu'aucune séquence homologue à *tts* n'est présente dans le génome de *S. thermophilus* IP6756. Le mécanisme de biosynthèse des EPS dans la souche IP6756 serait donc différent de celui décrit chez *S. pneumoniae* type 37.

Les résultats d'hybridation de sondes spécifiques des ORF des clusters *eps* de *S. thermophilus* CNRZ368 (Bourgoin *et al.*, 1999), NST2280 (Pluvinet, 2003) et IP6757 (Pluvinet *et al.*, 2004) codant des glycosyltransférases sur l'ADN génomique de la souche IP6756 montrent que le ou les gène(s) codant la ou les glycosyltransférase(s) présente(nt) une séquence éloignée de celles des trois clusters testés.

Les gènes impliqués pourraient constituer un cluster situé ailleurs sur le chromosome de la souche IP6756, mais celui-ci présenterait une structure très différente de celle de tous les autres clusters *eps* connus chez *S. thermophilus*. Il serait en effet dépourvu du module de régulation. Les gènes pourraient également ne pas être liés génétiquement et leur association fonctionnelle pourrait n'intervenir qu'au niveau post-transcriptionnel. Quelque soit le nombre de gènes impliqués et leur nature, il serait intéressant de préciser leurs interactions éventuelles avec le cluster *eps* identifié. En effet, le transporteur potentiel EpsZ pourrait constituer un transporteur alternatif.

L'identification de ces gènes constitue l'une des perspectives de cette thèse. Une approche expérimentale serait de réaliser une mutagenèse insertionnelle aléatoire chez *S. thermophilus*

IP6756 à l'aide du plasmide à réplication thermosensible pGhost9ISSI construit par E. Maguin. Ce type d'approche a été utilisé chez *S. thermophilus* CNRZ368, où l'intégration du plasmide au sein du chromosome résulte d'événements de transposition répllicative aléatoires de la séquence d'insertion ISSI (Thibessard *et al.*, 2002). Les mutants insertionnels dont la production d'EPS serait abolie ou réduite peuvent être sélectionnés par un crible rapide comme l'étalement sur milieu rouge de ruthénium, utilisé par Stingele *et al.* (1996a) lors de la caractérisation du cluster *eps* de *S. thermophilus* Sfi6. Dans ce crible, les souches non productrices d'EPS apparaissent rouges alors que les souches productrices sont blanches. L'identification des loci interrompus des mutants sélectionnés pourrait alors être entreprise par clonage puis séquençage des régions flanquant le site d'insertion du plasmide.

2 FONCTIONNALITE D'*ORF14.9*, APPARTENANT A LA SECONDE REGION DU CLUSTER *EPS* DE LA SOUCHE IP6756

2.1 Cette ORF est très conservée chez *S. thermophilus*

orf14.9 est retrouvée dans tous les clusters *eps* séquencés de *S. thermophilus*, sauf dans le cluster *eps* de type XI, dont la seconde région présente une organisation particulière. Elle contient un élément IS1194 potentiellement fonctionnel puis d'un élément IS905 dont l'ORF codant la transposase présente un caractère de pseudogène. La troisième et dernière ORF identifiée est également pseudogénique mais son produit présenterait des homologies avec la famille des protéines Cof. De telles ORF (pseudogène ou codant une protéine tronquée) sont annotées immédiatement en amont de *pgm* dans les clusters *eps* des souches MR-1C, MR-2C, MTC310, MTC330 et MTC360, CNRZ1066 et LMG18311. Il est donc possible qu'*orf14.9* soit absente du cluster *eps* de type XI. Cependant sans analyse supplémentaire, il ne peut être exclu qu'*orf14.9* soit néanmoins présente dans le génome de la souche qui présente ce cluster. Dans la suite de ce propos, nous ne prendrons pas en compte le cluster *eps* de type XI. Par ailleurs, *orf14.9* est détectée par hybridation dans 7 autres souches de *S. thermophilus*. Cette expérience d'hybridation n'a pas été possible sur des souches présentant un cluster *eps* de type XI, faute de souches disponibles.

La comparaison des séquences des différents allèles connus d'*orf14.9* permet de relever une très forte conservation de la phase codante de cette ORF (96 à 100% d'identité). Les comparaisons des séquences des gènes *lacS* et *lacZ* de 29 souches de *S. thermophilus* révèlent une variabilité comparable (Ercolini *et al.*, 2005). Ces gènes qui codent respectivement la perméase permettant l'entrée du lactose dans la cellule et la β -galactosidase permettant l'hydrolyse du lactose peuvent être considérés comme des gènes de ménage. Ces résultats suggèrent qu'*orf14.9* est aussi conservée que d'autres gènes du chromosome de *S. thermophilus* et que cette ORF conférerait un avantage sélectif à *S. thermophilus*.

Par ailleurs, *orf14.9* présente un contenu en bases G et C comparable à celui du génome de *L. lactis* (de 34 à 36%) (Bridge & Sneath, 1983). Cette ORF pourrait donc provenir du génome d'une autre bactérie présentant un pourcentage en bases G et C comparable, dont *L. lactis*. L'interrogation de la banque de donnée établie à l'issue du séquençage du chromosome de *L. lactis* IL1403 (Bolotin *et al.*, 2001) révèle qu'aucune protéine ne présente d'homologie significative avec la protéine potentielle Orf14.9. Comme *orf14.9* est rencontrée dans tous les clusters *eps* séquencés de *S. thermophilus* et détectée par hybridation dans toutes les autres souches de *S. thermophilus* analysées, cette ORF était probablement présente dans le cluster *eps* de l'ancêtre commun à toutes les souches de *S. thermophilus*.

Un dendrogramme des streptocoques a été réalisé (Figure 44) en utilisant la séquence de la superoxyde dismutase (Poyart *et al.*, 1998). *S. thermophilus* se révèle très proche de *S. vestibularis* et *S. salivarius*. L'espèce *S. thermophilus*, observée uniquement dans le lait et les produits laitiers, dériverait d'un clone de *S. salivarius* qui se serait adapté au milieu laitier (Hols *et al.*, 2005). Aucune donnée n'est disponible concernant la présence ou non d'*orf14.9* dans le génome de *S. salivarius*. Il est actuellement impossible de savoir si l'acquisition éventuelle d'*orf14.9* à partir de *L. lactis* ou d'une autre bactérie a eu lieu avant ou au moment de l'émergence de l'espèce *S. thermophilus*.

2.2 L'inactivation d'*orf14.9* ne s'accompagne pas d'une modification quantitative de la production d'EPS

Cette thèse a permis d'aborder la fonctionnalité de cette ORF et d'étudier son implication dans la biosynthèse des EPS. Les analyses *in silico* effectuées ne permettent pas de proposer une fonction pour cette ORF, puisqu'aucune protéine référencée dans les banques de données ne présente d'homologie avec la protéine potentielle Orf14.9. Ces analyses permettent néanmoins de prédire l'existence de 6 hélices transmembranaires au sein de la

S. thermophilus est indiquée en rouge.

protéine potentielle, ce qui suggère que la protéine Orf14.9 aurait une localisation membranaire. Cependant, aucun peptide signal n'a été détecté dans sa séquence protéique.

La première étape de l'étude de la fonctionnalité de cette ORF a consisté à mettre en évidence un transcrit spécifique d'*orf14.9* par RT-PCR chez *S. thermophilus* IP6756 ainsi que chez les souches NST2280 et A054. La transcription de cette ORF dépend d'un promoteur potentiel spécifique à *orf14.9* dont la structure est très conservée dans chaque allèle d'*orf14.9*. En 3' d'*orf14.9*, un terminateur de transcription rho-indépendant potentiel a également été identifié. Ces résultats démontrent le caractère fonctionnel de l'ORF au moins sur le plan transcriptionnel.

Des mutants insertionnels de cette ORF ont été construits dans les souches A054 et NST2280 puis caractérisés phénotypiquement. L'étude n'a pas été menée chez *S. thermophilus* IP6756 du fait des difficultés expérimentales rencontrées dans les expériences d'électrotransformation de cette souche. Ces difficultés pourraient être associées à la présence dans le génome de la souche IP6756 d'un élément mobilisable CIME302 qui porte un système de Restriction-Modification (Pavlovic *et al.*, 2004).

La présence d'*orf14.9* au sein de la seconde région de tous les clusters *eps* séquencés de *S. thermophilus* suggère que cette ORF pourrait être liée à la biosynthèse des EPS. Cependant, les taux d'EPS produits par chacun des mutants ne diffèrent pas significativement de ceux de leurs souches sauvages respectives.

Cependant, *orf14.9* pourrait être impliquée dans cette biosynthèse sur un plan qualitatif. En effet, de par sa localisation probablement membranaire, la protéine potentielle Orf14.9 pourrait être impliquée dans l'ancrage des EPS à la surface de la cellule ou dans l'export des EPS. Dans le premier cas, une observation au microscope électronique des cellules des souches sauvages et de leurs mutants respectifs apporterait des informations. Dans le second cas, les mécanismes d'export des EPS sont particulièrement complexes chez *S. thermophilus*.

Ainsi, les travaux sur la biosynthèse des EPS de *S. thermophilus* Sfi6 suggèrent qu'une molécule lipidique d'undécaprényl-phosphate joue le rôle d'accepteur pour l'unité répétée en voie de biosynthèse (Stingele *et al.*, 1999a). Cependant, les auteurs ont observé qu'un ajout de bacitracine, une molécule bloquant le recyclage des molécules d'undécaprényl-phosphate, n'affecte pas la production d'EPS (Stingele *et al.*, 1999a). Ces résultats suggèrent que l'export des EPS de *S. thermophilus* pourrait impliquer une molécule acceptrice alternative, rôle que pourrait jouer la protéine Orf14.9.

2.3 Le caractère fonctionnel d'*orf14.9* est nécessaire à une croissance cellulaire optimale

L'étude de la croissance des mutants insertionnels d'*orf14.9* réalisés a montré une entrée en phase stationnaire retardée par rapport aux souches sauvages et des paramètres de croissance significativement différents de ceux des souches sauvages correspondantes : un temps de génération significativement augmenté et un taux de croissance spécifique sensiblement réduit. Ces résultats permettent de valider le caractère fonctionnel d'*orf14.9* ainsi que l'efficacité des interruptions réalisées puisque les mutants réalisés présentent un phénotype modifié par rapport aux souches sauvages. Ces résultats suggèrent qu'*orf14.9* serait impliquée dans la croissance cellulaire. Le retard de croissance observé pourrait être dû à une perturbation du métabolisme général de la cellule ou à une perturbation de la mise en place de nouvelles structures cellulaires. Par exemple, la protéine Orf14.9 pourrait constituer un pore membranaire qui faciliterait l'entrée de nutriments. Une telle organisation en pore est proposée pour la protéine Tts (de *S. pneumoniae* de type 37) qui présente également 6 hélices transmembranaires (Llull *et al.*, 2001). La constitution d'un pore membranaire pourrait également nécessiter l'association de plusieurs monomères de la protéine Orf14.9. Cependant, cette hypothèse ne permet pas de comprendre la présence constante d'*orf14.9* dans le cluster *eps* de *S. thermophilus*. La protéine Orf14.9 pourrait également être liée à la biosynthèse du

peptidoglycane et constituer une interface entre la production de peptidoglycane et la biosynthèse d'EPS. Dans ce cadre, il serait intéressant d'étudier une éventuelle corrélation entre le taux de transcription d'*orf14.9* et le taux de production des EPS.

3 LE CLUSTER *EPS* DE *S. THERMOPHILUS* RESULTERAIT D'UN REGROUPEMENT DE GENES

3.1 L'organisation en cluster s'accompagne d'une possibilité de transfert de celui-ci à une souche receveuse présente dans la même niche écologique

Selon le schéma général décrit, la première région du cluster *eps* de *S. thermophilus* serait organisée en opéron et contient l'essentiel des gènes directement impliqués dans la biosynthèse des EPS, notamment ceux codant les glycosyltransférases. L'ensemble de ces gènes présente donc un regroupement tant physique que fonctionnel. Les gènes localisés dans la seconde région du cluster *eps* de *S. thermophilus* pourraient également participer à ce rapprochement physique.

Un tel regroupement s'accompagne d'une augmentation de la probabilité de transfert de l'ensemble du déterminisme génétique de la biosynthèse des EPS à une autre bactérie présente dans la même niche écologique et donc de la valeur sélective des gènes concernés. Ce regroupement s'accompagne donc également de la possibilité d'évolution rapide des clusters *eps* de cette espèce par remplacement de séquence. L'intervention de transferts horizontaux a été avancée dans la littérature pour expliquer le caractère extrêmement variable du cluster *eps* dans cette espèce. Les différentes questions soulevées seront discutées en fonction des résultats obtenus lors de cette étude du cas particulier du cluster *eps* de la souche IP6756. Quels sont les mécanismes impliqués dans les transferts d'ADN exogène chez *S. thermophilus* et en particulier de séquences appartenant aux clusters *eps* ? D'autres hypothèses ne pourraient-elles pas être formulées quant au rôle d'*orf14.9* dans les événements de transferts horizontaux ? Par ailleurs, le regroupement de gènes observé dans le cluster *eps* de cette espèce résulte très probablement d'une pression de sélection particulière s'exerçant sur le déterminisme génétique de la biosynthèse des EPS. Cette discussion terminera donc par les différentes hypothèses qui peuvent être formulées à ce sujet.

3.2 La constitution du cluster *eps* de la souche IP6756 implique l'intervention de transferts horizontaux

3.2.1 Un remplacement de séquence serait à l'origine de l'organisation du cluster *eps* de la souche IP6756

L'analyse de la séquence de chacune des ORF constituant ce cluster permet d'observer qu'*epsB* présente une taille de 489 pb, suite à une délétion de 243 pb de l'extrémité 3' par rapport aux allèles des autres clusters *eps* de *S. thermophilus*. De plus, les 24 derniers nucléotides de la séquence d'*epsB* ne présentent pas d'homologie avec les séquences référencées dans les banques de données. Un autre exemple de variabilité observée au niveau de la séquence d'ORF est fourni par *epsZ* qui présente une structure mosaïque. La comparaison de sa séquence avec les séquences des ORF du cluster *eps* des souches Sfi39 et MTC310 a en effet permis de constater qu'*epsZ* ne présente des homologies avec ces clusters *eps* qu'à partir du nucléotide 87 de sa séquence.

La comparaison de l'ensemble du cluster *eps* de *S. thermophilus* IP6756, localisé au même endroit sur le chromosome que les autres clusters *eps* identifiés chez *S. thermophilus*, avec celui des souches Sfi39 et MTC310 permet de définir trois grandes régions. La première et la troisième présentent des homologies significatives (pourcentage d'identité nucléotidique supérieur à 98%) avec les régions correspondantes des clusters *eps* de MTC310 et Sfi39, alors que la seconde région de 125 pb du cluster *eps* d'IP6756 ne présente aucune homologie avec les clusters *eps* de ces deux souches. Ces observations suggèrent très fortement que le cluster *eps* de la souche IP6756 résulterait d'un remplacement de séquence à partir d'un cluster ancestral éventuellement comparable à celui des souches Sfi39 et MTC310. Le remplacement de la région centrale initiale de plus de 11 kb par la région de 125 pb identifiée dans IP6756 s'accompagne de la création de l'ORF mosaïque *epsZ*, dont les 86 premiers nucléotides ne présentent pas d'homologie avec des séquences des souches Sfi39 et MTC310. Ce remplacement de séquence pourrait faire intervenir des événements de recombinaison homologue au niveau des régions proximales (*epsA* et *epsB*) et distales (*epsZ-epsJ*) du cluster

eps initial. Un tel mécanisme a déjà été proposé pour expliquer la constitution du cluster *eps* de *S. thermophilus* IP6757 (Pluvinet *et al.*, 2004). Ainsi, les souches NST2280 et IP6757 sont très proches phylogénétiquement alors qu'elles présentent un cluster *eps* très différent. Toutes les souches les plus proches de NST2280 et d'IP6757 présentent le même cluster que celui de la souche NST2280, montrant que le cluster *eps* de la souche *eps* d'IP6757 a évolué rapidement par remplacement de séquences. L'intervention d'un phénomène comparable a également été proposée pour expliquer l'observation de variants de la souche *S. pneumoniae* type 23F dont la seule différence avec la souche parentale réside dans leur cluster *cps* qui confère le sérotype 19F (Coffey *et al.*, 1998).

Ces deux derniers exemples, le remplacement de séquence s'accompagne du maintien de la présence d'ORF codant des glycosyltransférases, et par conséquent de la fonctionnalité du cluster *eps* et de la synthèse d'EPS. Dans le cas du cluster *eps* de la souche IP6756, le remplacement de séquence provoquerait la perte de telles ORF.

3.2.2 Origine de la région centrale du cluster *eps* chez la souche IP6756 et des gènes extérieurs au cluster *eps* nécessaires à la biosynthèse des EPS

La détermination du pourcentage en bases G et C de chacune des trois régions montre que celui des régions 1 et 3 (respectivement 36,2 et 37,5%) est comparable à celui du chromosome de *S. thermophilus* (Farrow & Collins, 1984). Par contre, le pourcentage en bases G et C de la région centrale (région 2) est très différent (32%). Cette constatation est cohérente avec l'intervention d'un remplacement de séquence relativement récent au niveau de la région centrale du cluster *eps* de la souche IP6756. En effet, ce pourcentage se différencie nettement du pourcentage en bases G et C du chromosome de *S. thermophilus*. La séquence de la région centrale proviendrait d'une autre espèce bactérienne, probablement présente dans la même niche écologique.

Par ailleurs, les gènes situés à l'extérieur du cluster *eps* de la souche IP6756 et impliqués dans la biosynthèse des EPS (notamment les gènes codant des glycosyltransférases) de cette souche ne sont pas identifiés. Il pourrait s'agir d'une part de gènes acquis par transfert horizontal au sein de la niche écologique de *S. thermophilus* ou d'autre part d'autres gènes génomiques complétant l'absence de glycosyltransférase. Leur origine est donc indéterminée.

L'ensemble de ces observations ne permet pas de formuler des hypothèses quant à la chronologie des événements ayant conduit à l'organisation observée dans la souche IP6756 pour le déterminisme génétique de la biosynthèse des EPS. Cependant, l'hypothèse la plus simple est que le remplacement de séquence soit concomitant à l'acquisition des gènes permettant un nouveau type de biosynthèse d'EPS.

3.3 Mode d'acquisition de l'ADN exogène : cas des séquences appartenant aux clusters *eps*

L'acquisition de molécules d'ADN dans une cellule peut se produire par trois mécanismes : la transduction, la transformation et la conjugaison. Au sein du cluster *eps* de la souche IP6756, aucun élément ne permet de privilégier l'un de ces mécanismes. Cette partie traitera donc uniquement du lien entre ces mécanismes et la constitution du cluster *eps* chez *S. thermophilus*.

La transduction fait intervenir une particule phagique. Des séquences particulières sont en général nécessaires à l'intégration de l'ADN phagique dans le chromosome : les sites *att* et les gènes *int* et *xis* qui codent les systèmes d'intégration/excision du bactériophage (Cho *et al.*, 2002). L'intervention de la transduction spécialisée semble ne pas affecter les clusters *eps* séquencés de *S. thermophilus* puisque aucune de ces séquences n'y a été identifiée. Par contre, aucun élément ne permet d'écarter l'intervention de la transduction généralisée qui s'accompagne du transfert de n'importe quelle région du génome d'une cellule lysée par un

phage. L'intégration de cet ADN exogène dans le chromosome se produit par recombinaison homologue et ne nécessite donc pas de séquences spécifiques.

Le second mécanisme pouvant être impliqué dans le transfert de séquences exogènes appartenant à des clusters *eps* est la transformation. Cette propriété avait déjà été observée par Griffith à la fin des années 1920 sur *S. pneumoniae*, les bactéries de phénotype rugueux présentant après ajout d'ADN dans le milieu un phénotype lisse qui traduit une production de capsule. Des analyses plus récentes au sein du genre *Streptococcus* ont montré que la capacité à incorporer de l'ADN étranger est régulée par des systèmes de *quorum sensing*. Ces systèmes, qui régulent notamment les gènes impliqués dans l'induction de la compétence chez *S. pneumoniae*, sont généralement constitués de trois composantes : un peptide signal et un système de régulation à deux composantes. Chez cette bactérie, l'initiation de la compétence est assurée par l'activité du modulateur de transcription global ComX, qui agit comme un facteur sigma alternatif. La transcription du gène *comX* est régulée par la protéine ComE, régulateur du système à deux composants activé par le système de *quorum sensing*. Le facteur sigma ComX initie ensuite la transcription d'opérons spécifiques de la compétence, qui sont impliqués dans les échanges d'ADN et la recombinaison. Ces gènes cibles présentent dans leurs régions promotrices une boîte *com* dont la séquence consensus est TACGAATA chez *S. pneumoniae* (Cvitkovitch *et al.*, 2003). Des premiers résultats obtenus chez *S. mutans* suggèrent que l'activité de *comX* serait corrélée à la fréquence de transformation (Aspiras *et al.*, 2004). Chez *S. thermophilus*, aucune étude n'a été menée à ce jour sur ce sujet. Cependant, l'analyse de la séquence du génome des souches LMG18311 et CNRZ1066 permet d'identifier un homologue du gène *comX*. De même, des homologues des gènes activés par ComX peuvent être identifiés (V. Libante, com. pers.). Chez *S. mutans*, la formation de biofilm qui implique les EPS synthétisés par la bactérie est liée à l'acquisition de

l'état de compétence, ce qui suggère que les clusters *eps* pourraient intervenir dans le développement de la compétence chez *S. thermophilus*.

La conjugaison chez *S. thermophilus* pourrait impliquer les éléments ICE, tel que ICES_{St1} (Burrus *et al.*, 2000), qui possède toutes les fonctions nécessaires à l'établissement de la conjugaison. Le caractère conjugatif de ces éléments reste cependant à démontrer. Ces éléments pourraient contenir des ORF homologues avec les séquences constituant le clusters *eps*. Leur transfert par conjugaison permettrait de transférer en même temps ces séquences homologues qui pourraient alors recombiner avec les séquences du cluster *eps* de la souche receveuse. Cette hypothèse est soutenue par l'observation que les ORF 308A d'une part et 308B et 308C d'autre part, situées de part et d'autre du module de réplication de l'élément mobilisable en cis CIME308 identifié dans le génome de *S. thermophilus* CNRZ308, présentent respectivement 94% et 98% d'identité nucléotidique avec *orf*c2 d'une part et *orf*-98 et *orf*-154 d'autre part appartenant au plasmide pNZ4000 de *L. lactis* NIZOB40. Ce plasmide contient le cluster *eps* de cette souche et *orf*-98 et *orf*-145 sont immédiatement adjacents à *orf*Y, une des frontières du cluster *eps* chez *L. lactis* (G. Guédon, com. pers.).

3.4 Hypothèses concernant le rôle d'*orf*14.9 dans l'acquisition de nouvelles séquences

La séquence d'*orf*14.9 présente une forte conservation et un pourcentage en bases G et C différent du génome. Cette double observation amène à s'interroger sur la fonction de cette ORF. Plusieurs hypothèses peuvent être envisagées. La première est celle d'une forte pression de sélection s'exerçant au niveau de la fonction codée par cette ORF empêchant le maintien de formes mutantes. La seconde est que la séquence de cette ORF pourrait servir de site de recombinaison homologue, processus exerçant une pression de sélection au niveau de cette ORF et qui serait à l'origine du maintien de l'homologie de cette ORF avec celle issue de *L. lactis*.

Une autre hypothèse prend en compte les résultats obtenus lors de cette thèse. Puisque l'interruption d'*orf14.9* s'accompagne d'une réduction significative de la croissance des souches de *S. thermophilus* testées, la fonctionnalité de cette ORF conférerait un avantage sélectif aux cellules qui la possèdent. Le fait qu'*orf14.9* soit associée au cluster *eps* de *S. thermophilus* pourrait conférer une valeur sélective supérieure à l'ensemble de ces gènes. Ce paramètre a probablement joué et pourrait jouer encore un rôle important dans les transferts de séquences impliquant des clusters *eps* et explique également le maintien du caractère fonctionnel de cette ORF au sein des clusters *eps*.

3.5 Quelles sont la ou les pression(s) de sélection s'exerçant sur le cluster *eps* chez *S. thermophilus* ?

Chez *S. thermophilus*, les clusters *eps* présentent une très grande variabilité. Cette variabilité repose sur des remplacements de séquence, pouvant résulter d'événements de transferts horizontaux avec des bactéries présentes dans la même niche écologique, notamment *L. lactis*. Des hypothèses de transferts intraspécifiques chez *S. thermophilus* ont également été émises mais restent à démontrer. Cette variabilité génétique est corrélée à la très grande diversité structurale présentée par les EPS de *S. thermophilus*.

Cette variabilité est comparable à celle observée au niveau des clusters *cps* de *S. pneumoniae*. Chez cette bactérie, ces clusters codent la biosynthèse de la capsule qui joue un rôle très important dans la virulence de cette bactérie pathogène puisqu'elle confère une protection face à la phagocytose. L'intervention de remplacements de séquence au niveau des clusters *cps* est également proposée chez cette bactérie (Coffey *et al.*, 1998) et permet une évolution rapide du cluster *cps* de cette bactérie. Chez cette bactérie pathogène, la variabilité des clusters *cps* résulte d'une sélection de nouveaux variants en réponse au système immunitaire de l'hôte, lui aussi capable de générer de la variabilité dans ses mécanismes de défense.

La variabilité des clusters *eps* observée chez *S. thermophilus* est probablement essentielle chez cette bactérie. Il est ainsi possible de formuler différentes hypothèses sur la fonction de cette variabilité. Cette question est intimement liée à celle du rôle des EPS chez *S. thermophilus*, qui à ce jour n'est pas déterminé avec précision. *S. thermophilus* dérivant probablement de *S. salivarius*, une telle variabilité serait-elle également présente chez cette bactérie, hôte commensal de la cavité buccale des Mammifères ? Si tel était le cas, cela suggérerait que *S. thermophilus* a conservé la très grande variabilité de ses clusters *eps* ce qui lui conférerait un avantage sélectif dans sa niche écologique différente : le lait et les produits laitiers.

Les EPS de *S. thermophilus* confèrent au milieu des propriétés rhéologiques intéressant l'industrie agro-alimentaire (CF paragraphe 1.3.2 de l'introduction). Une des pressions de sélection s'exerçant sur le cluster *eps* correspond à la sélection par les fabricants artisanaux ou industriels de souches conférant les propriétés rhéologiques les plus intéressantes. Une telle pression de sélection ne peut expliquer la très grande variabilité observée au niveau des clusters *eps* de *S. thermophilus*. Cette constatation amène à s'intéresser à l'écologie de la bactérie, afin d'affiner les hypothèses qui peuvent être formulées.

S. thermophilus présente comme seule niche écologique connue le lait ou les produits laitiers. De plus, elle est une des rares espèces du genre *Streptococcus* à ne pas être pathogène. L'interprétation de la variabilité observée au niveau des clusters *cps* de *S. pneumoniae* ne convient donc pas à la variabilité observée chez *S. thermophilus*. L'hypothèse qui ferait intervenir des relations de commensalité avec les différentes espèces de Mammifère peut être avancée. Il paraît possible d'envisager des interactions entre cellules eucaryotes et *S. thermophilus* dans le milieu lait. Ainsi, la teneur moyenne en cellules somatiques du lait de vache produit en Suède est de $233 \cdot 10^3$ cellules par ml (Lindmark-Mansson *et al.*, 2003).

L'hypothèse de la commensalité pourrait également expliquer l'éventuelle variabilité des clusters *eps* de *S. salivarius*.

Il est concevable que la variabilité observée au niveau des clusters *eps* de *S. thermophilus* soit associée à un ou plusieurs facteurs eux-mêmes variables. Quels peuvent-ils être ? Dans le lait ou les produits laitiers, *S. thermophilus* fait très souvent partie d'une flore lactique complexe. Si cette bactérie n'est associée qu'à *Lactobacillus delbrueckii* subsp *bulgaricus* lors de la fabrication de yaourt, la flore lactique de certains fromages est éminemment plus complexe. La variabilité observée pourrait être associée aux relations possibles entre ces bactéries. Les EPS pourraient en effet jouer un grand rôle dans l'établissement de biofilm lors de la fabrication de produits laitiers. La variabilité des clusters *eps* et la variabilité structurale des EPS en résultant pourraient contribuer à favoriser l'établissement de telles structures complexes. L'organisation d'un biofilm est particulièrement complexe et présente des conséquences physiologiques importantes sur les bactéries qui le constituent. Un biofilm peut ainsi avoir des conséquences sur la nutrition des différentes bactéries présentes dans ce biofilm. Certaines substances excrétées par une bactérie peuvent ainsi être métabolisées plus facilement par une autre bactérie présente à proximité. Par exemple, le galactose excrété par *S. thermophilus* lors du fonctionnement de la pompe LacS, qui permet l'entrée du lactose dans la cellule, pourrait être métabolisé par une bactérie voisine. Une situation où *S. thermophilus* tirerait avantage de la présence proche d'une autre bactérie pourrait aussi se concevoir. Les biofilms permettent également de favoriser les transferts génétiques entre les bactéries constituant ce biofilm (CF paragraphe 3.3 de la discussion). La variabilité présentée par les EPS de *S. thermophilus* permettrait de favoriser les interactions avec certaines bactéries présentes dans la même niche écologique et de favoriser ainsi les phénomènes de protocoopération entre celles-ci.

Les EPS de *S. thermophilus* pourraient également constituer une barrière de protection pour la cellule. A quels paramètres variables est soumis *S. thermophilus* ? Il pourrait s'agir de la présence de phages qui peuvent contaminer les cultures. Les bactériophages sont les entités les plus nombreuses de la biosphère (de l'ordre de 10^{31} (Lopez & Garcia, 2004)) et ils sont fréquemment responsables de problèmes de fermentation dans l'industrie laitière. Les EPS pourraient protéger *S. thermophilus* des attaques phagiques en perturbant l'adsorption des phages. L'infection phagique débutant par l'adsorption des phages sur des molécules osidiques telles que le galactose ou le rhamnose de la paroi cellulaire, il est possible de se demander si la présence de ces molécules osidiques au niveau des EPS ne pourrait pas entrer en compétition avec les sites d'adsorption. La présence d'EPS associés à la cellule confère ainsi une faible protection à la souche MG1614 de *L. lactis* (Looijesteijn *et al.*, 2001). De même, le plasmide pCI658, identifié chez *Lactococcus lactis* subsp *cremoris* HO2, code la biosynthèse d'un EPS qui diminue l'adsorption des phages $\phi 712$ et $\phi c2$ sur leurs récepteurs cellulaires (Forde *et al.*, 1999). Cependant, des travaux menés sur de nombreuses autres souches de *L. lactis* ne permettent pas d'observer de différence de sensibilité aux phages entre les souches productrices d'EPS et les variants non producteurs (Deveau *et al.*, 2002). Des résultats obtenus chez *S. thermophilus* MR-1C suggèrent que la présence d'EPS n'est pas suffisante pour protéger les cellules contre l'attaque des phages de types Q1, Q3 et Sf11 (Broadbent *et al.*, 2003). Il est également possible que certaines insensibilités aux phages soient dues à l'absence de récepteurs cellulaires spécifiques à ces phages. De plus, les systèmes de restriction-modification constituent un autre mécanisme de défense lors d'infection phagique. Les EPS ne constitueraient ainsi qu'une partie des paramètres impliqués dans la protection face au phage (Broadbent *et al.*, 2003).

Cette hypothèse de la relation entre la variabilité des clusters *eps* de *S. thermophilus* et la potentialité d'attaques phagiques semble particulièrement intéressante. Elle ne diffère pas

fondamentalement de celle communément admise pour la variabilité des clusters *cps* de *S. pneumoniae*. Il s'agirait là aussi d'une relation hôte-pathogène classique. Cette question constitue une des perspectives les plus importantes pour comprendre le rôle effectif des EPS chez *S. thermophilus*.

CONCLUSION

Le déterminisme génétique de la biosynthèse des EPS chez *S. thermophilus* IP6756 présente une organisation particulièrement originale. Le cluster *eps* de cette souche présente ainsi une taille réduite par rapport aux autres clusters *eps* connus chez *S. thermophilus* et son organisation modulaire diffère également de celle décrite dans les autres clusters *eps*. Alors que tous les clusters *eps* séquencés chez *S. thermophilus* contiennent des gènes codant des glycosyltransférases, aucune ORF de ce type ne peut être identifiée dans le cluster *eps* de la souche IP6756. De telles enzymes étant nécessaires à la biosynthèse des unités répétées constituant les EPS, cette situation suggère fortement qu'un ou plusieurs gènes extérieurs au cluster *eps* identifié sont impliqués dans la biosynthèse des EPS. Cette hypothèse a été confirmée lors de l'étude fonctionnelle du cluster *eps* de la souche IP6756. Il est ainsi transcrit et potentiellement traduit. Cependant, l'inactivation de ce cluster se traduit par aucune modification significative du taux d'EPS produit. L'élucidation complète du mécanisme de biosynthèse des EPS dans la souche IP6756 constitue une des perspectives principales de cette thèse. Il reste également à étudier l'implication qualitative du cluster *eps* identifié. Par exemple, le transporteur d'unités répétées potentiel codé par *epsZ* (la troisième ORF du cluster) pourrait constituer un transporteur alternatif dispensable. De même, la mutase potentielle EpsJ (codée par *epsJ*, la quatrième ORF du cluster) pourrait également modifier la nature des oses incorporés dans l'EPS de la souche IP6756.

Cette situation rappelle celle observée chez *S. pneumoniae* de type 37, qui contient un cluster *cps* dont plusieurs gènes sont inactivés. La biosynthèse de l'homopolysaccharide formant la capsule de type 37 est codée par le gène *tts* situé à l'extérieur du cluster *cps* de la souche. Les analyses structurales préliminaires des molécules d'EPS produites par la souche IP6756 révèlent qu'il s'agit d'un hétéropolysaccharide. Ce résultat indique que le mécanisme de biosynthèse des EPS dans la souche IP6756 diffère de celui décrit chez *S. pneumoniae* de

type 37. De plus, aucune séquence présentant une forte homologie avec celle de *tts* ne peut être détectée par hybridation dans le génome de *S. thermophilus* IP6756.

L'étude du cluster *eps* de la souche IP6756 a également permis de s'intéresser au concept de cluster. Alors que ce cluster est très particulier, il contient comme les autres *orf14.9*. Cette observation amène à s'interroger sur la fonctionnalité des ORF regroupées à droite de la première région. Ce regroupement pourrait permettre un transfert horizontal efficace de l'ensemble des ORF impliquées dans la biosynthèse des EPS. La situation présentée par le cluster *eps* de la souche IP6756 est une des plus simples observable chez *S. thermophilus* puisque cette région ne présente qu'une seule ORF fonctionnelle *orf14.9*. L'interruption de cette ORF ne modifie pas significativement le taux d'EPS produit. Une des perspectives de cette thèse consiste à étudier une implication qualitative de cette ORF dans la biosynthèse des EPS chez *S. thermophilus*. Cette ORF pourrait également jouer un rôle important dans les mécanismes de transferts des séquences contenant les clusters *eps* de par l'avantage sélectif qu'elle confère aux cellules receveuses et aux clusters *eps* qui la contiennent.

La présence dans le génome de la souche IP6756 d'un cluster *eps* localisé au même endroit que dans les autres souches ainsi que celle d'autre(s) gène(s) impliqué(s) dans la biosynthèse des EPS ailleurs dans le génome pourrait conférer un avantage évolutif à la souche. Si les EPS jouent effectivement un rôle important dans la physiologie de *S. thermophilus*, ce qui est fortement suggéré par la variabilité observée, le cluster *eps* identifié pourrait constituer un site de recombinaison permettant à la bactérie d'acquérir par remplacement de séquence un autre type de cluster *eps*. De tels remplacements de séquence nécessitent le transfert d'ADN exogène dont le ou les mécanismes ne sont pas encore déterminés.

A plus long terme, la régulation des clusters *eps* chez *S. thermophilus* constitue une perspective importante. Il serait ainsi particulièrement intéressant d'étudier l'interaction

éventuelle entre la protéine CcpA et la transcription du cluster *eps*. Des éléments CRE potentiels peuvent en effet être identifiés dans la région promotrice du cluster *eps* chez *S. thermophilus*. L'intervention d'une protéine homologue à CcpA sur l'activité des gènes de biosynthèse des EPS a été mise en évidence chez *S. pneumoniae* et *S. mutans*. Chez ces bactéries, cette protéine jouerait un rôle d'interface entre la production d'EPS et la régulation catabolique de gènes (autre fonction de la protéine CcpA). De même, afin d'avancer dans l'identification du ou des rôles joués par les EPS chez *S. thermophilus*, il serait intéressant d'étudier le niveau de transcription du cluster *eps* en réponse à différents facteurs.

MATERIEL ET METHODES

1 SOUCHES BACTERIENNES ET CONDITIONS DE CULTURE

Les souches de *S. thermophilus* utilisées lors de ce travail sont présentées dans le tableau 4. En particulier, *S. thermophilus* IP6756 a été isolé en 1959 à partir d'une culture servant au démarrage de la production de gruyère dans le Jura français. Les précultures et cultures de *S. thermophilus* sont réalisées à 42°C, à l'exception des transformants de *S. thermophilus* obtenus, présentant un plasmide dérivé de pGHost9, qui sont cultivés à 30°C, sans agitation en aérobiose ou en anaérobiose. L'anaérobiose est obtenue par incubation des cellules en jarres munies de système GasPak (BioMérieux). Les précultures et cultures sont effectuées dans du lait reconstitué à 10% (p/v), dans du milieu M17 (Terzaghi & Sandine, 1975), dans du milieu Belliker (milieu Elliker additionné d'extrait de bœuf 10g/l) (O'Sullivan & Fitzgerald, 1999) ou dans du milieu TPPY (Bracquart, 1981). Ces milieux sont autoclavés 20 minutes à 110°C (lait reconstitué, TPPY et M17) ou à 120°C (Belliker). Les transformants de *S. thermophilus* sont cultivés en présence d'érythromycine à une concentration de 5 µg/ml. Les intégrants obtenus à partir des transformants sont cultivés en présence d'érythromycine à une concentration de 2 µg/ml.

L'ADN total d'une souche de *S. pneumoniae* de type 37 nous a été fourni par l'équipe du Dr Garcia (Centro de Investigaciones Biologicas, Consejo Superior de Investigaciones Cientificas, Madrid, Espagne).

La souche *Escherichia coli* TG1 (Stratagene) est utilisée comme hôte des plasmides recombinants dérivant de pGHost9. Son génotype est le suivant : *supE*, *thi-1*, $\Delta(lac-proAB)$, $\Delta(mcrB-hsdSM)$, $5(r_k^- m_k^-)$, [F' *traD36*, *proAB*, *lacIqZAM15*]. Les cellules d'*E. coli* TG1 porteuses de plasmides recombinants sont cultivées à 37°C sous agitation (250 rpm) sur milieu LB (Luria-Bertani) en présence d'érythromycine (150 µg/ml).

Tableau 4 : Souches de *S. thermophilus* utilisées.

Souche (a)	Collection (b)	Origine	Type de cluster <i>eps</i> (c)
ATCC19258 ^T	Collection ATCC	isolée du lait pasteurisé	5
ATCC19987		fromage	4
CNRZ7		yaourt	2
CNRZ308	Collection CNRZ	gruyère	4
CNRZ368		yaourt	5
CNRZ385		yaourt	2
CNRZ388		yaourt	5
CNRZ391		yaourt	5
CNRZ404		yaourt	5
CNRZ445		emmental	3
CNRZ464		yaourt	5
CNRZ702		yaourt	4
CNRZ1066		yaourt	5
CNRZ1068		yaourt	5
CNRZ1402		yaourt	3
IP6756	Collection IP	gruyère	6
IP6757		yaourt	1
LMG18311	Collection Université de Gand	yaourt	1
A054	souche industrielle	yaourt	5
NST1	Collection du laboratoire	yaourt	5
NST3		yaourt	ND
NST5		yaourt	5
NST10		yaourt	ND
NST11		yaourt	5
NST12		yaourt	ND
NST2280		yaourt	5

(a) ATCC19258^T souche type de *S. thermophilus*

(b) ATCC : American Type Culture Collection

CNRZ : Centre National de la Recherche Zootechnique

IP : Institut Pasteur

NST : Nancy *Streptococcus thermophilus*

(c) D'après Pluvinet *et al.*, 2004

2 PLASMIDE

Le plasmide pGHost9 (Maguin *et al.*, 1992) a été utilisé pour cloner des fragments internes aux ORF *epsA*, *orf5* et *orf14.9*, générant respectivement les plasmides pGHost9*epsAorf5* et pGHost9*orf14.9*. Une carte génétique simplifiée du plasmide pGHost9 est présentée figure 45. Il porte le gène *ery* qui confère une résistance à 150 µg/ml d'érythromycine chez *E. coli* et à 5 µg/ml chez *S. thermophilus*. Il possède une origine de réplication lui permettant de se répliquer selon le mécanisme du cercle roulant à la fois chez *E. coli* et chez les bactéries lactiques. La réplication de pGHost9 est thermosensible. Chez *E. coli* TG1, la protéine RepA codée par le génome n'est pas thermosensible et complémente la protéine thermosensible produite par pGHost9, ce qui permet la réplication du plasmide à 37°C. Par contre, chez *S. thermophilus* aucune réplication n'est possible à 42°C (température restrictive) alors qu'elle l'est lorsque la culture est effectuée à 30°C (température permissive). Le caractère thermosensible de la réplication de ce plasmide chez *S. thermophilus* a été utilisé lors des expériences d'interruption des ORF *epsA* et *orf14.9* réalisées par mutagenèse insertionnelle avec les deux plasmides recombinants obtenus.

3 PREPARATION D'ACIDES NUCLEIQUES

3.1 Extraction d'ADN plasmidique et génomique

Les minipréparations d'ADN plasmidique d'*E. coli* sont effectuées selon la méthode de lyse alcaline (Sambrook *et al.*, 1989), puis purifiées par extraction au phénol/chloroforme.

L'ADN plasmidique de *S. thermophilus* est extrait selon le protocole de Frere (Frere, 1994). Le protocole suivi pour l'extraction de l'ADN génomique de *S. thermophilus* est celui décrit par Leblond *et al.* (Leblond *et al.*, 1996).

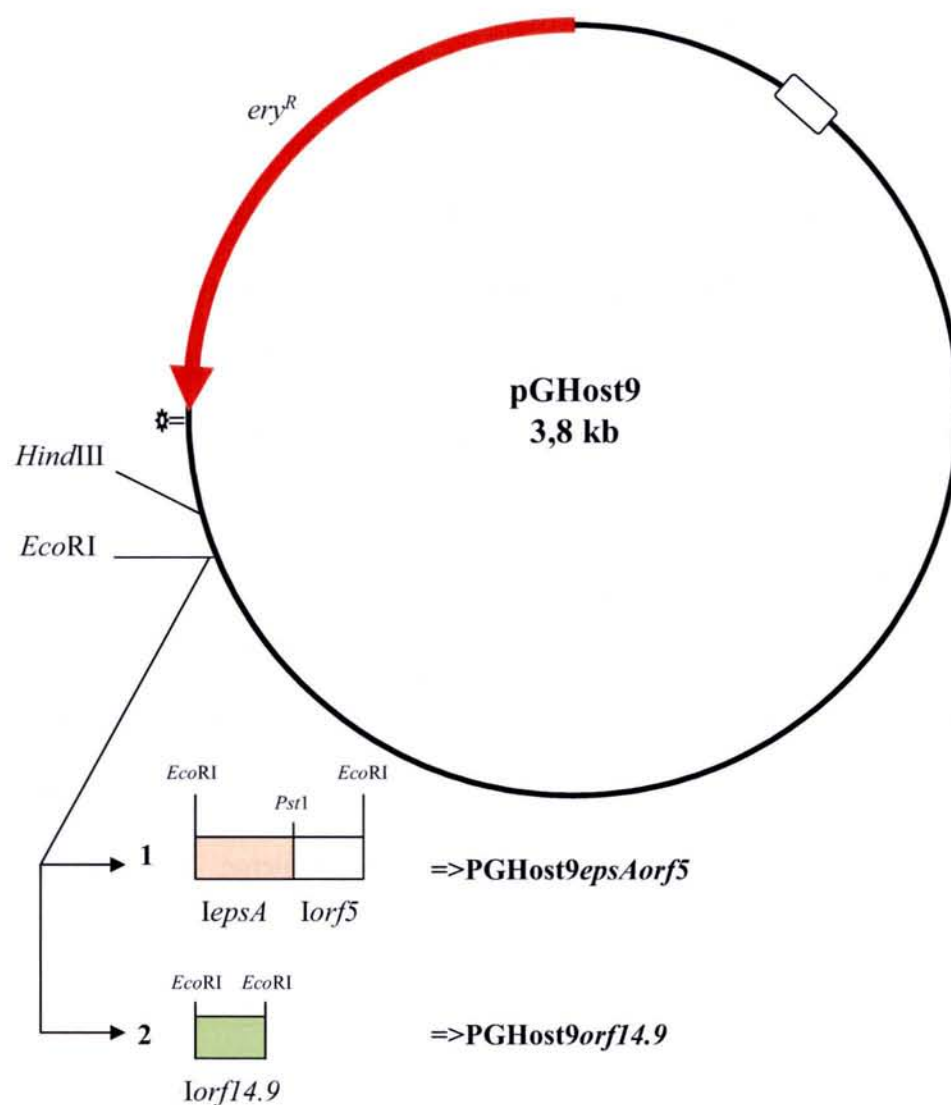


Figure 45 : Représentation simplifiée du plasmide pGHost9 (d'après Maguin *et al.*, 1992).

→ ery^R : gène de résistance à l'érythromycine

□ Origine de répllication thermosensible

Seuls les sites de restriction uniques (*HindIII* et *EcoRI*) utilisés lors de ce travail sont représentés.

1, 2 : représentation simplifiée des fragments clonés dans le site unique *EcoRI* du plasmide PGhost9 ce qui génère respectivement les plasmide pGHost9*epsAorf5* et pGHost9*orf14.9* (CF paragraphe 8).

lepsA, *lorf5* et *lorf14.9* : fragment interne à *epsA*, *orf5* et *orf14.9* respectivement.

⚙ : terminateur de transcription du gène *ery*.

3.2 Extraction des ARN totaux

Les ARN totaux sont extraits selon le protocole décrit par Fernandez *et al.* (Fernandez *et al.*, 2004) avec quelques modifications. Ainsi, les cellules sont cultivées en milieu M17 et sont récupérées au début de la phase stationnaire ($DO_{600}=1,2-1,4$). Les cellules sont lysées en vortexant les cultures pendant quatre-vingt secondes à vitesse maximum. Ce protocole dérive de celui de Chomczynski et Sacchi (Chomczynski & Sacchi, 1987) : utilisation du réactif Extract-All (Eurobio), un mélange de thiocyanate de guanidine et de phénol.

3.3 Amplification d'ADN par polymérisation en chaîne

La PCR permet d'amplifier de manière exponentielle un segment d'ADN en utilisant deux amorces complémentaires de deux régions flanquant l'ADN à amplifier (Saiki *et al.*, 1988). Les amorces utilisées sont définies à l'aide du logiciel primer3 (Rozen & Skaletsky, 2000) et présentent une température de fusion (T_m) équivalente. Ces amorces sont synthétisées par MWG-Biotech ou Invitrogen.

3.3.1 Amplification par PCR

Les 25 μ l du mélange réactionnel contiennent l'ADN ou l'ADNc (10 ng), les dNTP (200 μ M), les amorces (25 pmol), le tampon de réaction 1X et 2 UE de *Taq* polymérase (Biolabs). Une réaction de PCR ne contenant pas d'ADN matrice constitue le témoin négatif, ce qui permet de s'assurer de la spécificité des amplifiats obtenus.

Une première dénaturation de l'ADN est réalisée par incubation du mélange réactionnel pendant 5 minutes à 94°C. Le cycle de dénaturation-hybridation-élongation dont les paramètres sont précisés comme suit est répété 30 ou 35 fois : la dénaturation s'opère à 94°C pendant 30 secondes, l'hybridation des amorces se déroule pendant 30 secondes à la température ajustée en fonction du T_m des amorces utilisées, et enfin l'élongation des

amorces par la *Taq* polymérase se déroule à 72°C pendant 30 secondes. A l'issue de ces 30 ou 35 cycles, le mélange est incubé à 72°C pendant 7 min, ce qui permet l'élongation des brins d'ADN partiellement synthétisés. Les réactions de PCR sont réalisées dans le thermocycleur iCycler (Biorad). Les amorces utilisées pour la réalisation de différentes sondes, l'obtention d'insert interne aux ORF *epsA*, *orf5* et *orf14.9* et les réactions d'amplification sur l'ADN complémentaire obtenu par RT-PCR sont présentées dans le tableau 5.

3.3.2 Amplification d'ADN par LongRange PCR

La PCR réalisée avec la *Taq* polymérase permet d'amplifier des fragments d'une taille maximale de 4 kb. Cette enzyme ne présentant pas d'activité exonucléasique 3'→5' (activité lecture d'épreuve), l'efficacité de la réaction diminue à chaque incorporation de base incorrecte. La LongRange PCR permet d'amplifier des fragments d'ADN pouvant aller jusqu'à 35 kb. La polymérase utilisée (LA *Taq* -LA pour Long Accurate- commercialisée par Takara) est pourvue d'une activité lecture d'épreuve ce qui augmente considérablement l'efficacité de la réaction d'amplification. D'après les données du fabricant, cette enzyme est 7 fois plus fidèle que la *Taq* polymérase.

Le cluster *eps* de *S. thermophilus* IP6756 a été amplifié dans sa totalité en utilisant ce type de réaction de PCR. Les amorces utilisées sont précisées dans le tableau 5. Une réaction de LongRange PCR est réalisée en parallèle sans matrice pour constituer un témoin négatif.

Les 50 µl de mélange réactionnel contiennent l'ADN matrice (100 ng), les dNTP (500 µM), les amorces (50 pmol chacune), 5 µl du tampon de LR-PCR fourni, 5 µl de MgCl₂ (2,5 mM) et 0,5 µl d'enzyme LA *Taq*. Une première dénaturation de l'ADN est réalisée par incubation du mélange réactionnel pendant 1 minute à 94°C. Le cycle de dénaturation-hybridation-élongation, dont les paramètres sont précisés comme suit, est répété 35 fois :

Tableau 5 : Amorces utilisées pour les réactions de PCR.

Les sites de restriction introduits sont indiqués en noir.

Amorces	Sequence (5'-3')	Etude	Température de PCR/ nombre de cycle
o-epsA5	CAGGCACCTACTAAGGTTG	Transcription d' <i>epsA</i>	49°C/ 35
o-epsA6a	ACATCTGAACGTGACACTG		
o-epsB1	CATTAGTGAAAGTTACGCCC	Transcription d' <i>epsB</i>	49°C/ 35
o-epsA8	CTTGCAATTCCTAGTGTCG		
o-epsI*a	GGGTGCCTTGTCATCGAC	Transcription d' <i>epsZ</i>	49°C/ 35
o-epsI*b	GAGGCCCAACAAAGCCA		
o-epsJA'	ACAGACCTGCCTCCTTTC	Transcription d' <i>epsJ</i>	49°C/ 35
o-epsJ1	GCAAGACGTCCACAGAAG		
o-14.9a	CTCCCAACAACAATAATACC	Transcription d' <i>orf14.9</i>	49°C/ 35
o-14.9b	CCTCAGCAGTTTATTGGTC		
o-epsA7	GACCGTGGTAAAAACCAGG	Transcrit chevauchant <i>epsA-epsB</i>	49°C/ 35
o-epsBa	GGTCTAAACTTCTCTAAAG		
o-epsB3'	TCATAGTGGGCTTAGTAATG	Transcrit chevauchant <i>epsB-epsZ</i>	49°C/ 35
o-epsI	TGCATACCAGGAAAAGAACAA		
o-eps2	AGCAGTAGATGGTCTTGTC	Transcrit chevauchant <i>epsZ-epsJ</i>	49°C/ 35
o-epsJ5'	GAAAGGAGGCAGGTCTGT		
o-epsA4	GGAAGAAGAAGACTTATAATA	Amplification par LR-PCR du cluster	49°C/ 35
o-pgm	CCTGAATTCGACTGAAAGGCAC		
o-epsJA	ACAGACCTGCCTCCTTTC	Réalisation d'une sonde spécifique d' <i>epsJ</i>	50°C/ 30
o-epsJ1	GCAAGACGTCCACAGAAG		
o-epsZa	GGGTGCCTTGTCATCGAC	Réalisation d'une sonde spécifique d' <i>epsZ</i>	50°C/ 30
o-epsZb	GAGGCCCAACAAAGCCA		
o-tts1	TGAGGCCTTTGGTATTGG	Réalisation d'une sonde spécifique de <i>tts</i>	49°C/ 30
o-tts2	GCCTTAAAGCCATTTCGAT		
o-epsA2	CCTTCAGAATTCTAAGAAATTACC	Obtention de l'insert interne d' <i>epsA</i>	58°C/ 30
o-epsA3	GTCAATCAACTGCAGGAAAGAT		
o-orf14.9-2	GCGACGAATTCTTGCTGATTATGCCCG	Obtention de l'insert interne d' <i>orf14.9</i> et séquençage	67°C/ 30
o-14.9-4	TAAGAATTCTAATGGCAAATGTGCC		
o-orf5-1	CAAATGCTGCAGAAAAATTACTAA	Obtention de l'insert interne d' <i>orf5</i>	58°C/ 30
o-orf5-2	AAGAATGAATTCTCATACGCAGG		

dénaturation à 98°C pendant 5 secondes, hybridation des amorces pendant 1 minute à 55°C, puis élongation des amorces par la LA *Taq* polymérase à 68°C pendant 20 minutes. A l'issue de ces 35 cycles, le mélange est incubé à 72°C pendant 10 min, ce qui permet l'élongation des brins d'ADN partiellement synthétisés.

Les réactions de LongRange PCR sont réalisées dans le thermocycleur GeneAmp PCR System 2400 (Applied Biosystems/Perkin-Elmer). Les amorces utilisées pour ces réactions sont présentées dans le tableau 5.

3.3.3 Purification des fragments d'ADN

Les produits de PCR sont purifiés grâce au kit de purification « High Pure PCR Product Purification Kit » (Roche Molecular Biochemicals).

3.4 Dosage des acides nucléiques

Les solutions d'acide nucléique sont dosées au spectrophotomètre à 260 nm. Les différents échantillons d'ARN totaux extraits présentaient un rapport DO_{260}/DO_{280} compris entre 1,6 et 1,9, reflétant l'absence presque totale d'ADN génomique et de protéines contaminantes. De même, les solutions d'ADN totaux sont soumises à la mesure du rapport DO_{260}/DO_{280} . Ce rapport est compris entre 1,8 et 2 indiquant que la solution n'est pas contaminée par des protéines. Lorsque cette méthode n'est pas applicable, le dosage est effectué après électrophorèse en gel d'agarose et coloration au BET ($0,5 \mu\text{g.ml}^{-1}$). L'intensité de la fluorescence des bandes du complexe ADN/BET est observée sous UV (appareil GelDoc 1000 Biorad) et comparée à celle d'un marqueur de taille dont le poids moléculaire et la quantité sont connus.

4 ANALYSE DE RESTRICTION ET ELECTROPHORESE

Les enzymes de restriction sont utilisées dans les conditions de tampon et de température indiquées par le fournisseur (Roche Molecular Biochemicals ou New England Biolabs). Les produits de digestion sont ensuite séparés par électrophorèse en gel d'agarose en tampon TAE ou TBE (Sambrook *et al.*, 1989). La concentration en gel d'agarose, la tension appliquée et la durée de migration sont ajustées en fonction de la taille des fragments à séparer. L'ADN du bactériophage λ digéré par les enzymes *HindIII* ou *PstI* est utilisé comme marqueur de taille. Après coloration du gel dans une solution de BET ($0,5 \mu\text{g}.\text{ml}^{-1}$), une caméra couplée au système de traitement d'image GelDoc 1000 (Biorad) permet de visualiser les profils de restriction.

5 HYBRIDATION ADN-ADN (SOUTHERN BLOT)

5.1 Marquage de sonde

L'ADN est marqué par la méthode des amorces aléatoires "Random priming" à l'aide du kit "DIG DNA Labeling and Detection Kit Non radioactive" (Roche Molecular Biochemicals). Cette technique est basée sur la synthèse par la polymérase de Klenow d'un brin complémentaire au brin matrice en y incorporant aléatoirement des dUTP couplés à la digoxigénine à la place de dTTP. Chaque marquage est réalisé dans les conditions préconisées par le fournisseur pendant une durée minimale de 12 h. Le marquage est quantifié par détection en colorimétrie et comparaison avec une solution d'ADN marqué, fournie dans le kit.

5.2 Transfert d'ADN sur membrane

Après électrophorèse en gel d'agarose, l'ADN est transféré par capillarité sur membrane de nylon (Amersham) en appliquant une masse de 1 kg sur le gel et la membrane pendant 3 à 5 h. Au préalable, le gel est traité successivement par les solutions de dépurination (HCl 0,25 M) pendant 30 min, de dénaturation (NaCl 1,5 M/NaOH 0,5 M) pendant 30 min et de transfert (NH₄Ac 1 M/NaOH 0,02 M) pendant 1 h. Après transfert, la membrane est rincée dans du SSC 2X puis l'ADN est fixé à la membrane sous l'action des UV ($\lambda=254$ nm) pendant 30 s.

5.3 Hybridation et détection

La membrane est préhybridée à 68°C pendant 1 h dans 20 ml de solution de préhybridation (SSC 5x ; SDS 0,02% ; N-Lauroyl Sarcosine 0,1% ; agent bloquant 1%) selon le protocole du kit "DIG DNA Labeling and Detection Kit Non radioactive" (Roche Molecular Biochemicals) pour saturer les sites de fixation aspécifique. Elle est ensuite incubée pendant au moins 12 h à 68°C en présence de la solution d'hybridation (addition de 2,5 ml de solution de préhybridation, pour une membrane de 100 cm², contenant 50 µl de sonde marquée, dénaturée et 50 µl d'ADN du phage λ natif, marqué et dénaturé). Elle est alors lavée deux fois 5 min avec une solution de SSC 2X/SDS 0,1% sous agitation à température ambiante puis deux fois 15 min avec une solution de SSC 0,1X/SDS 0,1%. Compte tenu du faible contenu en G+C du génome de *S. thermophilus*, ces deux derniers lavages sont réalisés à 60°C, sous agitation, pour obtenir une forte stringence. Ces lavages permettent l'élimination de la solution d'hybridation, de la sonde non fixée à l'ADN et des hybrides non spécifiques. Les membranes sont ensuite équilibrées à un pH de 7,5 par un bain de 5 min dans du tampon de lavage (maléate 100 mM ; NaCl 150 mM pH7,5 ; Tween 20 0,3%) sous agitation à température ambiante puis incubées pendant 1 h dans la solution d'agent bloquant 1% (maléate 100 mM ; NaCl 150 mM pH7,5 ; agent bloquant 1% p/v) ce qui permet de saturer les sites aspécifiques de la membrane.

Les hybrides ADN-ADN sont détectés par la fixation spécifique d'un anticorps anti-digoxigénine (DIG) sur les nucléotides modifiés de la sonde marquée. Pour cela, les membranes sont incubées 30 min dans la solution d'anticorps anti-DIG conjugués à la phosphatase alcaline dilués au 1/20000 dans la solution d'agent bloquant. Deux lavages de 15 min à température ambiante dans du tampon de lavage permettent d'éliminer les anticorps fixés de manière aspécifique et un lavage de 5 min à température ambiante dans du tampon de détection (Tris-HCl 100 mM pH9,5 ; NaCl 100 mM) permet d'équilibrer les membranes à un pH de 9,5, optimal pour l'action de la phosphatase alcaline. L'anticorps utilisé possède en effet une activité phosphatase qui lui permet d'hydrolyser un substrat spécifique de la méthode de détection utilisée. En colorimétrie, les réactifs NBT et X-Phosphate (45 et 35 μ l ajoutés respectivement dans 10 ml de tampon de détection) forment, après réaction, un précipité violet à l'endroit où l'anticorps est fixé. En luminescence, l'hydrolyse du substrat CDPstar (Disodium 2-chloro-5-(4-methoxyspiro{1,2-dioxetane-3,2'-5'-chloro [3,3 1^{3,7}]decan}-4-yl)-1-phenyl Phosphate) provoque sa décomposition et l'émission de lumière est captée par l'appareil Fluor'S Max (Biorad). Chaque membrane est incubée 5 min à température ambiante dans du CDPstar dilué au 1/100 dans du tampon de détection.

6 TRANSCRIPTION INVERSE DES ARN EN ADNc

La transcription inverse ou synthèse d'ADN complémentaire (ADNc) à partir des ARN totaux nécessite deux étapes préliminaires avant l'action de la transcriptase inverse. La première étape consiste en l'élimination de l'ADN contaminant les échantillons d'ARN par un traitement à la DNase I (0,4 UE. μ l⁻¹) pendant 15 min à 37°C. Cette enzyme est ensuite dénaturée 5 min à 95°C. La seconde étape consiste en l'hybridation d'hexanucléotides aléatoires (pdN6, Pharmacia) sur les ARN, pour amorcer la transcription inverse. Cette hybridation est réalisée pendant 5 min à 70°C dans un mélange de 10 μ l contenant

l'échantillon d'ARN (10 µg) et les amorces aléatoires (100 ng.µl⁻¹). Puis le mélange réactionnel de 25 µl qui contient le tampon de transcription inverse 1X (Tris 10 mM pH 8,3/dithiothréitol 2 mM/MgCl₂ 1,6 mM), les dNTP (20 µM) et un inhibiteur de RNase d'origine porcine (RNA Guard®, Pharmacia) (1 UE.µl⁻¹) est ajouté. La transcription inverse est alors réalisée pendant 1 h à 37°C par ajout de la transcriptase inverse (Moloney Murine Leukemia Virus Reverse Transcriptase, Amersham) (8 UE.µl⁻¹). Cette dernière est ensuite inactivée par dénaturation thermique 5 min à 95°C.

7 SEQUENÇAGE D'ADN ET ANALYSE DES SEQUENCES

7.1 Réactions de séquençage

Le séquençage d'ADN simple brin est réalisé, à partir d'ADN double brin préalablement dénaturé, selon la méthode de Sanger (Sanger *et al.*, 1977) en utilisant le protocole BigDye Terminator (Applied Biosystems/Perkin-Elmer). Ce protocole permet de réaliser une réaction unique, chacun des 4 ddNTP étant lié à un fluorochrome différent. Les 20 µl de mélange réactionnel contiennent 0,25 pmol d'ADN matrice, 3,2 pmol d'amorce et 4 µl d'un mélange, fourni par Applied Biosystems/Perkin-Elmer (contenant la polymérase, les dNTP et les ddNTP liés aux fluorochromes). Le cycle de réaction (répété 25 fois) comprend trois étapes : dénaturation (10 secondes à 96°C), hybridation (5 secondes à 50°C) et élongation (4 minutes à 60°C). Les réactions sont réalisées dans le thermocycleur GeneAmp PCR System 2400 (Applied Biosystems/Perkin-Elmer). Une précipitation des produits de la réaction permet d'éliminer les fluorochromes libres. La migration et la détection des pics de fluorescence sont réalisées par le séquenceur automatique ABI prism 377 DNA sequencer. Les séquences sont ensuite obtenues grâce au logiciel ABI Prism Sequencing Analysis Version 3.0 et assemblées grâce au logiciel AutoAssembler 2.1.1.

La séquence du cluster eps de *S. thermophilus* IP6756 est référencée sous le numéro d'accèsion AY916787 dans la banque de données GenBank/EMBL.

7.2 Analyse des séquences

Le logiciel DNA Strider Version 1.2 (Marck, 1988) permet d'identifier les phases ouvertes de lecture, dont le pourcentage en G+C est déterminé grâce au logiciel BioEdit (Hall, 1999). Le programme BLAST 2.0 (Basic Local Alignment Search Tool (Altschul *et al.*, 1990) (Altschul *et al.*, 1997) est utilisé pour détecter les similarités des séquences nucléotidiques avec celles des banques de données GenBank/EMBL et GenBank/EMBL updates (BLASTn). Il permet également de détecter les similarités des produits de traduction des différentes ORF identifiées avec les séquences protéiques des banques de données GenPept et updates (BLASTx et BLASTp). Les alignements de séquences nucléotidiques et protéiques sont réalisés grâce au programme ClustalW (Thompson *et al.*, 1994). Le logiciel mfold 3.0 (Zuker, 2003) permet de déceler les structures tiges boucles potentielles et d'estimer leur valeur d'énergie libre (SantaLucia, 1998). La prédiction d'hélices transmembranaires des protéines potentielles est effectuée à l'aide du programme SMART (Schultz *et al.*, 1998) qui permet également de déterminer avec précision leur localisation. La recherche d'éventuel peptide signal sur une séquence protéique putative est effectuée avec le programme SignalP3.0 (Bendtsen *et al.*, 2004). La recherche de domaines par comparaison avec des séquences protéiques connues est réalisée avec le logiciel CDART (Geer *et al.*, 2002) et en utilisant la banque CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) (Marchler-Bauer *et al.*, 2005).

8 CLONAGE DE FRAGMENTS D'ADN

8.1 Ligation

Un produit de 782 pb, de 452 pb et de 451 pb correspondant respectivement à une séquence interne d'*epsA*, d'*orf5* et d'*orf14.9* sont générés par PCR sur l'ADN génomique de la souche IP6756 à l'aide des amorces précisées dans le tableau 5. La localisation de ces amorces est précisée figure 46.

Les deux premiers produits sont digérés par *Pst*I et ligués ensemble. Le fragment *epsA-orf5* ainsi généré ainsi que le troisième produit sont digérés par *Eco*RI. Après digestion par *Eco*RI du plasmide pGHost9, le fragment *epsA-orf5* ou *orf14.9* est ajouté dans un rapport 1/5 afin de favoriser la ligation d'un insert dans le vecteur par rapport à la ligation intramoléculaire. La réaction est effectuée par la T4 DNA ligase (Roche Molecular Biochemicals) (1 UE/ µg d'ADN) pendant 16 heures à 4°C dans les conditions préconisées par le fournisseur. Les plasmides pGHost9*epsAorf5* et pGHost9*orf14.9* sont ainsi respectivement générés.

8.2 Electroporation

Les cellules compétentes d'*E. coli* TG1 ont été préparées selon le protocole décrit par Dower *et al.* (Dower *et al.*, 1988). L'électroporation a été réalisée sur des échantillons de 40 µl de cellules compétentes à l'aide de l'appareil Gene Pulser (Biorad), selon les paramètres conseillés par le fournisseur. Les transformants sont sélectionnés sur milieu LB contenant 150 µg/ml d'érythromycine.

L'électroporation des souches de *S. thermophilus* a été réalisée selon un protocole dérivé de celui de O'Sullivan et Fitzgerald (O'Sullivan & Fitzgerald, 1999) (Pavlovic, 2004). Les cultures utilisées pour la transformation des souches A054, ATCC19258, NST2280 et IP6756 sont réalisées en milieu Belliker Thréonine (20 mM). Lorsque que $DO_{600}=0,3$, le milieu est

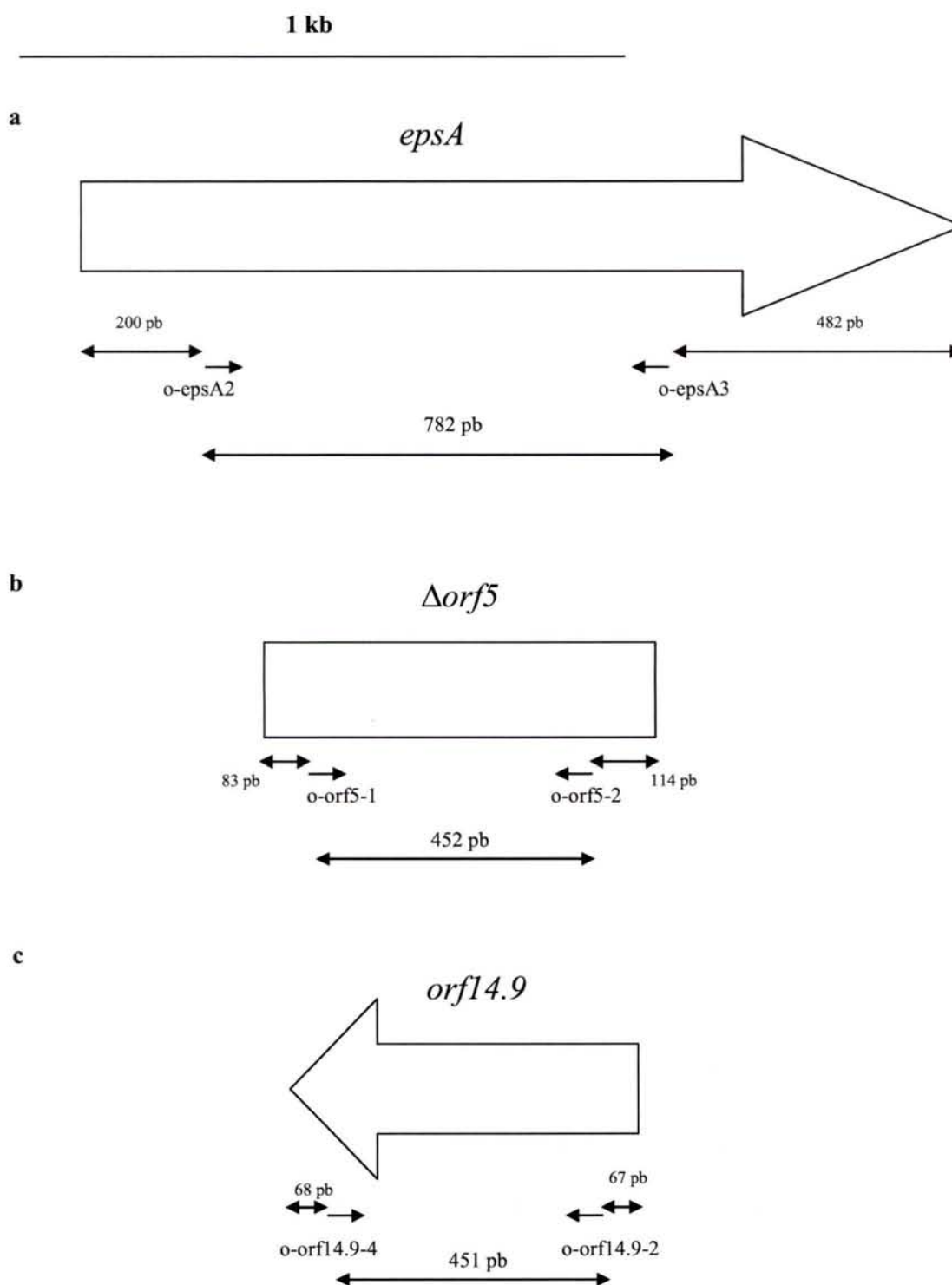


Figure 46 : Localisation des amorces utilisées pour amplifier les fragments internes d'*epsA* (a), *orf5* (b) et *orf14.9* (c) de *S. thermophilus* IP6756, qui seront clonés dans le plasmide pGHost9. La séquence de $\Delta orf5$ présente 98% d'identité avec la séquence d'*orf5* de *S. thermophilus* NST2280. La séquence de chacune des amorces est précisée dans le tableau 2.

remplacé par un milieu Belliker Thréonine (20 mM) supplémenté en glycine (10% p/v) et la culture est prolongée d'une heure. Les étapes suivantes de préparation des cellules compétentes sont effectuées selon le protocole de O'Sullivan et Fitzgerald (1999). 80 µl de cellules compétentes sont électroporées en présence de plasmide (250 ng d'ADN) en utilisant l'appareil Gene Pulser (Biorad) avec les paramètres suivants : 2,5 kV, 200 Ω et 25 µF.

La transformation de la souche IP6756 par l'ADN plasmidique d'*E. coli* étant difficile du fait notamment de la présence de l'élément mobilisable en cis CIME302 qui porte un système Restriction Modification (Pavlovic *et al.*, 2004), la transformation a été réalisée dans un premier temps avec la souche ATCC19258. L'ADN plasmidique du transformant obtenu est ensuite utilisé pour transformer IP6756.

9 DELETION DU CLUSTER *EPS* CHEZ *S. THERMOPHILUS* IP6756 ET NST2280

Le plasmide pGHost9*epsAorf5* est intégré dans le chromosome des souches IP6756 et NST2280 par simple crossing-over en présence d'érythromycine et à température restrictive (42°C) selon le protocole décrit par Burrus *et al.* (Burrus *et al.*, 2001). L'ADN total de 20 et 22 co-intégrants potentiels obtenus respectivement pour IP6756 et NST2280 est extrait et soumis à une analyse par Southern Blot pour vérifier la localisation de l'intégration du plasmide et son nombre de copie. Les clones recombinants IP6756-A17 et NST2280-A8, présentant une intégration multiple de pGHost9*epsAorf5* dans *epsA* des souches IP6756 et NST2280 respectivement, sont retenus pour réaliser la délétion du cluster *eps* de leurs souches respectives. La levée de la pression de sélection et le passage à température permissive (30°C) permet de sélectionner les événements d'excision du plasmide par recombinaison homologue. Cette excision s'accompagne soit de la restauration du cluster *eps* initial soit de sa délétion, c'est-à-dire de la perte des gènes chromosomiques situés entre *epsA* et *orf5*.

10 INTERRUPTION D'ORF14.9 CHEZ *S. THERMOPHILUS* NST2280 ET A054

Le plasmide pGHost9orf14.9 est intégré dans le chromosome des souches NST2280 et A054 par simple crossing-over selon le protocole décrit par Burrus *et al.* (Burrus *et al.*, 2001). L'ADN total de 12 et 15 co-intégrants potentiels obtenus respectivement pour NST2280 et A054 est extrait et analysé comme précédemment. Les clones recombinants NST2280-O7 et A054-O12, présentant une intégration multiple de pGHost9orf14.9 dans orf14.9 des souches NST2280 et A054 respectivement, sont retenus pour les analyses phénotypiques ultérieures.

11 DETERMINATION DU CARACTERE PRODUCTEUR D'EPS

11.1 Extraction des EPS

Les EPS des souches de *S. thermophilus* sont extraits selon le protocole décrit par Robson *et al.* (Robson *et al.*, 1974). Une culture en milieu lait reconstitué 10% est incubée pendant 12 heures à 42°C. Les protéines sont précipitées par addition d'un volume égal d'acide trichloroacétique (40%). Après centrifugation, un volume égal d'acétone est ajouté au surnageant pour précipiter les EPS. Ces derniers sont ensuite dissous dans 1 ml d'eau ultra-pure.

11.2 Dosage des EPS extraits

Les exopolysaccharides extraits sont ensuite dosés par la méthode de Dubois *et al.* (Dubois *et al.*, 1956). Cette méthode consiste à hydrolyser les exopolysaccharides en présence d'acide sulfurique et de phénol puis de mesurer l'absorbance à 490 nm des produits. La concentration est alors déterminée par comparaison avec une gamme étalon de glucose.

11.3 Composition des EPS

La nature des oses entrant dans la composition des molécules d'EPS produites par la souche IP6756 a été déterminée par analyse en chromatographie sur couche mince selon la méthode décrite par Seo *et al.* (Seo *et al.*, 2004). Les EPS extraits sont hydrolysés en présence d'acide trifluoroacétique 4M pendant 2 heures à 100°C puis concentrés dans un évaporateur rotatif. Les échantillons sont ensuite déposés sur la plaque de silice et élués dans une chambre de chromatographie avec le solvant n-propanol : eau : triéthylamine : 30% NH₃ 80 :20 :0,2 :4 (v/v/v/v) à température ambiante. Puis la plaque est séchée, aspergée d'une solution d'acide sulfurique à 30% dans du méthanol et chauffée à 170°C jusqu'à révélation complète.

12 DETERMINATION DES PARAMETRES DE CROISSANCE BACTERIENNE EN MILIEU LIQUIDE

Les souches NST2280 et A054 ainsi que leurs mutants insertionnels respectifs NST2280-O7 et A054-O12 sont cultivés dans 50 ml de milieu M17 (A054 et A054-O12) ou TPPY (NST2280 et NST2280-O7) à partir d'une préculture en milieu lait reconstitué de 12 heures. Quand DO₆₀₀=0,5, dix tubes de 10 ml de milieu M17 ou TPPY sontensemencés avec 0,2 ml de la culture. Chaque tube est utilisé pour mesurer la densité optique à 600 nm à un temps précis. Les courbes de croissance établies pour chacune des souches étudiées permettent de déterminer le taux de croissance spécifique (μ) et le temps de génération (g) de chaque souche.

REFERENCES BIBLIOGRAPHIQUES

- Albert, M. J., Bhuiyan, N. A., Rahman, A., Ghosh, A. N., Hultenby, K., Weintraub, A., Nahar, S., Kibriya, A. K., Ansaruzzaman, M. & Shimada, T. (1996).** Phage specific for *Vibrio cholerae* O139 Bengal. *J Clin Microbiol* **34**, 1843-1845.
- Almiron-Roig, E., Mulholland, F., Gasson, M. J. & Griffin, A. M. (2000).** The complete *cps* gene cluster from *Streptococcus thermophilus* NCFB 2393 involved in the biosynthesis of a new exopolysaccharide. *Microbiology* **146** (Pt 11), 2793-2802.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990).** Basic local alignment search tool. *J Mol Biol* **215**, 403-410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.
- Arrecubieta, C., Garcia, E. & Lopez, R. (1995).** Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. *Gene* **167**, 1-7.
- Arthur, M., Depardieu, F., Molinas, C., Reynolds, P. & Courvalin, P. (1995).** The *vanZ* gene of Tn1546 from *Enterococcus faecium* BM4147 confers resistance to teicoplanin. *Gene* **154**, 87-92.
- Aspiras, M. B., Ellen, R. P. & Cvitkovitch, D. G. (2004).** ComX activity of *Streptococcus mutans* growing in biofilms. *FEMS Microbiol Lett* **238**, 167-174.
- Battisti, L., Lara, J. C. & Leigh, J. A. (1992).** Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. *Proc Natl Acad Sci U S A* **89**, 5625-5629.
- Becker, A., Niehaus, K. & Puhler, A. (1995).** Low-molecular-weight succinoglycan is predominantly produced by *Rhizobium meliloti* strains carrying a mutated ExoP protein characterized by a periplasmic N-terminal domain and a missing C-terminal domain. *Mol Microbiol* **16**, 191-203.
- Becker, A., Katzen, F., Puhler, A. & Ielpi, L. (1998).** Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Appl Microbiol Biotechnol* **50**, 145-152.
- Bender, M. H. & Yother, J. (2001).** CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae*. *J Biol Chem* **276**, 47966-47974.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004).** Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-795.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarne, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**, 731-753.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G. D., Fonstein, M., Overbeek, R., Kyprides, N., Purnelle, B., Prozzi, D., Ngui, K., Masuy, D., Hancy, F., Burteau, S., Boutry, M., Delcour, J., Goffeau, A. & Hols, P. (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**, 1554-1558.

Borges, F., Layec, S., Thibessard, A., Fernandez, A., Gintz, B., Decaris, B. & Leblond-Bourget, N. (2005). *cse*, a chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J Bacteriol*.

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B. & Guedon, G. (1999). Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* **233**, 151-161.

Bourgoin, F., Guedon, G., Pebay, M., Roussel, Y., Panis, C. & Decaris, B. (1996). Characterization of a mosaic *ISSI* element and evidence for the recent horizontal transfer of two different types of *ISSI* between *Streptococcus thermophilus* and *Lactococcus lactis*. *Gene* **178**, 15-23.

Bracquart, P. (1981). An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yoghurt. *J Appl Bacteriol* **51**, 303-305.

Bridge, P. D. & Sneath, P. H. (1983). Numerical taxonomy of *Streptococcus*. *J Gen Microbiol* **129**, 565-597.

Broadbent, J. R., McMahon, D. J., Welker, D. L., Oberg, C. J. & Moineau, S. (2003). Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J Dairy Sci* **86**, 407-423.

Browngardt, C. M., Wen, Z. T. & Burne, R. A. (2004). RegM is required for optimal fructosyltransferase and glucosyltransferase gene expression in *Streptococcus mutans*. *FEMS Microbiol Lett* **240**, 75-79.

Brussow, H., Bruttin, A., Desiere, F., Lucchini, S. & Foley, S. (1998). Molecular ecology and evolution of *Streptococcus thermophilus* bacteriophages--a review. *Virus Genes* **16**, 95-109.

Burne, R. A., Wen, Z. T., Chen, Y. Y. & Penders, J. E. (1999). Regulation of expression of the fructan hydrolase gene of *Streptococcus mutans* GS-5 by induction and carbon catabolite repression. *J Bacteriol* **181**, 2863-2871.

Burrus, V., Roussel, Y., Decaris, B. & Guedon, G. (2000). Characterization of a novel integrative element, *ICES1*, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* **66**, 1749-1753.

- Burrus, V., Bontemps, C., Decaris, B. & Guedon, G. (2001).** Characterization of a novel type II restriction-modification system, *Sth368I*, encoded by the integrative element *ICES1* of *Streptococcus thermophilus* CNRZ368. *Appl Environ Microbiol* **67**, 1522-1528.
- Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002a).** The *ICES1* element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**, 77-97.
- Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002b).** Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**, 601-610.
- Caimano, M. J., Hardy, G. G. & Yother, J. (1998).** Capsule genetics in *Streptococcus pneumoniae* and a possible role for transposition in the generation of the type 3 locus. *Microb Drug Resist* **4**, 11-23.
- Caroff, M. & Karibian, D. (2003).** Structure of bacterial lipopolysaccharides. *Carbohydr Res* **338**, 2431-2447.
- Cartee, R. T., Forsee, W. T., Schutzbach, J. S. & Yother, J. (2000).** Mechanism of type 3 capsular polysaccharide synthesis in *Streptococcus pneumoniae*. *J Biol Chem* **275**, 3907-3914.
- Cerning, J. (1995).** Production of exopolysaccharides by lactic acid bacteria and dairy propionibacteria. *Lait* **75**, 463-472.
- Charron-Bourgoin, F., Pluvinet, A., Morel, C., Guedon, G. & Decaris, B. (2001).** Polymorphism of *eps* loci involved in exopolysaccharide synthesis of *Streptococcus thermophilus*. *Lait* **81**, 281-288.
- Cho, E. H., Gumport, R. I. & Gardner, J. F. (2002).** Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *J Bacteriol* **184**, 5200-5203.
- Chomczynski, P. & Sacchi, N. (1987).** Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**, 156-159.
- Cieslewicz, M. J., Kasper, D. L., Wang, Y. & Wessels, M. R. (2001).** Functional analysis in type Ia group B *Streptococcus* of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci. *J Biol Chem* **276**, 139-146.
- Clarke, B. R., Pearce, R. & Roberts, I. S. (1999).** Genetic organization of the *Escherichia coli* K10 capsule gene cluster: identification and characterization of two conserved regions in group III capsule gene clusters encoding polysaccharide transport functions. *J Bacteriol* **181**, 2279-2285.
- Clarke, B. R., Esumeh, F. & Roberts, I. S. (2000).** Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5A: evidence for two distinct K5 lyase enzymes. *J Bacteriol* **182**, 3761-3766.

- Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona, R., Hryniewicz, W., Paton, J. C. & Spratt, B. G. (1998).** Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol Microbiol* **27**, 73-83.
- Crater, D. L. & van de Rijn, I. (1995).** Hyaluronic acid synthesis operon (*has*) expression in group A streptococci. *J Biol Chem* **270**, 18452-18458.
- Cvitkovitch, D. G., Li, Y. H. & Ellen, R. P. (2003).** Quorum sensing and biofilm formation in Streptococcal infections. *J Clin Invest* **112**, 1626-1632.
- Daveran-Mingot, M. L., Campo, N., Ritzenthaler, P. & Le Bourgeois, P. (1998).** A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. *J Bacteriol* **180**, 4834-4842.
- de Ruyter, P. G., Kuipers, O. P., Meijer, W. C. & de Vos, W. M. (1997).** Food-grade controlled lysis of *Lactococcus lactis* for accelerated cheese ripening. *Nat Biotechnol* **15**, 976-979.
- De Vuyst, L. & Degeest, B. (1999).** Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiol Rev* **23**, 153-177.
- De Vuyst, L., Vanderveken, F., Van de Ven, S. & Degeest, B. (1998).** Production by and isolation of exopolysaccharides from *Streptococcus thermophilus* grown in a milk medium and evidence for their growth-associated biosynthesis. *J Appl Microbiol* **84**, 1059-1068.
- De Vuyst, L., De Vin, F., Vaningelgem, F. & Degeest, B. (2001).** Recent developments in the biosynthesis and applications of heteropolysaccharides from lactic acid bacteria. *International Dairy Journal* **11**, 687-707.
- Defives, C., Werquin, M., Mary, P. & Hornez, J. P. (1996).** Roles of Exopolysaccharides and Lipopolysaccharides in the Adsorption of the Siphovirus Phage NM8 to *Rhizobium meliloti* M11S Cells. *Curr Microbiol* **33**, 371-376.
- Degeest, B., Vaningelgem, F. & De Vuyst, L. (2001).** Microbial physiology, fermentation kinetics, and process engineering of heteropolysaccharide production by lactic acid bacteria. *International Dairy Journal* **11**, 747-757.
- Delcour, J., Ferain, T., Deghorain, M., Palumbo, E. & Hols, P. (1999).** The biosynthesis and functionality of the cell-wall of lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**, 159-184.
- Deveau, H., Van Calsteren, M. R. & Moineau, S. (2002).** Effect of exopolysaccharides on phage-host interactions in *Lactococcus lactis*. *Appl Environ Microbiol* **68**, 4364-4369.
- Dillard, J. P. & Yother, J. (1994).** Genetic and molecular characterization of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 3. *Mol Microbiol* **12**, 959-972.

- Dillard, J. P., Vandersea, M. W. & Yother, J. (1995).** Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J Exp Med* **181**, 973-983.
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988).** High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-6145.
- Drummelsmith, J. & Whitfield, C. (1999).** Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol Microbiol* **31**, 1321-1332.
- Duboc, P. & Mollet, B. (2001).** Applications of exopolysaccharides in the dairy industry. *International Dairy Journal* **11**, 759-768.
- Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956).** Colorimetric method for determination of sugars and related substances. *Anal. Chem.* **28**, 350-356.
- Ercolini, D., Fusco, V., Blaiotta, G. & Coppola, S. (2005).** Sequence heterogeneity in the *lacSZ* operon of *Streptococcus thermophilus* and its use in PCR systems for strain differentiation. *Res Microbiol* **156**, 161-172.
- Farrow, J. A. & Collins, M. D. (1984).** DNA base composition, DNA-DNA homology and long-chain fatty acid studies on *streptococcus thermophilus* and *Streptococcus salivarius*. *J Gen Microbiol* **130**, 357-362.
- Fernandez, A., Thibessard, A., Borges, F., Gintz, B., Decaris, B. & Leblond-Bourget, N. (2004).** Characterization of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368. *Arch Microbiol* **182**, 364-372.
- Forde, A. & Fitzgerald, G. F. (2003).** Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid, pCI658. *Plasmid* **49**, 130-142.
- Forde, A., Daly, C. & Fitzgerald, G. F. (1999).** Identification of four phage resistance plasmids from *Lactococcus lactis* subsp. *cremoris* HO2. *Appl Environ Microbiol* **65**, 1540-1547.
- Frere, J. (1994).** Simple method for extracting plasmid DNA from lactic acid bacteria. *Lett Appl Microbiol* **18**, 227-229.
- Galas, D. J. & Chandler, M. (1989).** Bacterial insertion sequences. In *Mobile DNA*, pp. 109-162. Edited by D. E. Berg & M. M. Howe. Washington: American Society for Microbiology.
- Garcia, E., Llull, D., Munoz, R., Mollerach, M. & Lopez, R. (2000).** Current trends in capsular polysaccharide biosynthesis of *Streptococcus pneumoniae*. *Res Microbiol* **151**, 429-435.
- Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002).** CDART: protein homology by domain architecture. *Genome Res* **12**, 1619-1623.

- Giammarinaro, P. & Paton, J. C. (2002).** Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. *Infect Immun* **70**, 5454-5461.
- Gonzalez, J. E., Reuhs, B. L. & Walker, G. C. (1996a).** Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proc Natl Acad Sci U S A* **93**, 8636-8641.
- Gonzalez, J. E., York, G. M. & Walker, G. C. (1996b).** *Rhizobium meliloti* exopolysaccharides: synthesis and symbiotic function. *Gene* **179**, 141-146.
- Graves, M. C. & Rabinowitz, J. C. (1986).** In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J Biol Chem* **261**, 11409-11415.
- Grobben, G. J., Sikkema, J., Smith, M. R. & de Bont, J. A. M. (1995).** Production of extracellular polysaccharides by *Lactobacillus delbrueckii* ssp. *bulgaricus* NCFB 2772 grown in a chemically defined medium. *J Appl Bacteriol* **79**, 103-107.
- Grobben, G. J., Smith, M. R., Sikkema, J. & J.A.M., d. B. (1996).** Influence of fructose and glucose on the production of exopolysaccharides and the activities of enzymes involved in the sugar metabolism and the synthesis of sugar nucleotides in *Lactobacillus delbrueckii* subsp. *bulgaricus* NCFB 2772. *Appl Microbiol Biotechnol* **46**, 279-284.
- Guedon, G., Bourgoïn, F., Burrus, V., Pluvinet, A. & Decaris, B. (2000).** Implication of horizontal transfers in genetic polymorphism of lactic acid bacteria. *Sc. Alim* **20**, 85-95.
- Guedon, G., Bourgoïn, F., Pebay, M., Roussel, Y., Colmin, C., Simonet, J. M. & Decaris, B. (1995).** Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergeneric transfer of insertion sequences occur in lactic acid bacteria co-cultures? *Mol Microbiol* **16**, 69-78.
- Hall, T. A. (1999).** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95-98.
- Hardy, G. G., Caimano, M. J. & Yother, J. (2000).** Capsule biosynthesis and basic metabolism in *Streptococcus pneumoniae* are linked through the cellular phosphoglucomutase. *J Bacteriol* **182**, 1854-1863.
- Hardy, G. G., Magee, A. D., Ventura, C. L., Caimano, M. J. & Yother, J. (2001).** Essential role for cellular phosphoglucomutase in virulence of type 3 *Streptococcus pneumoniae*. *Infect Immun* **69**, 2309-2317.
- Hassan, A. N., Frank, J. F. & Shalabi, S. I. (2001).** Factors affecting capsule size and production by lactic acid bacteria used as dairy starter cultures. *Int J Food Microbiol* **64**, 199-203.

- Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Ehrlich, D., Guédon, E., Monnet, V., Renault, P. & Kleerebezem, M. (2005).** New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by a comparative genomics. *FEMS Microbiol Rev* Sous presse.
- Jolly, L. & Stingle, F. (2001).** Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. *International Dairy Journal* **11**, 733-745.
- Keith, L. M. & Bender, C. L. (1999).** AlgT (sigma22) controls alginate production and tolerance to environmental stress in *Pseudomonas syringae*. *J Bacteriol* **181**, 7176-7184.
- Kim, J. O. & Weiser, J. N. (1998).** Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. *J Infect Dis* **177**, 368-377.
- Lamothe, G. T., Jolly, L., Mollet, B. & Stingle, F. (2002).** Genetic and biochemical characterization of exopolysaccharide biosynthesis by *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Arch Microbiol* **178**, 218-228.
- Laws, A. P. & Marshall, V. M. (2001).** The relevance of exopolysaccharides to the rheological properties in milk fermented with ropy strains of lactic acid bacteria. *International Dairy Journal* **11**, 709-721.
- Lazarevic, V., Margot, P., Soldo, B. & Karamata, D. (1992).** Sequencing and analysis of the *Bacillus subtilis* *lytRABC* divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier. *J Gen Microbiol* **138**, 1949-1961.
- Leblond, P., Fischer, G., Francou, F. X., Berger, F., Guerineau, M. & Decaris, B. (1996).** The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* **19**, 261-271.
- Lemoine, J., Chirat, F., Wieruszeski, J. M., Strecker, G., Favre, N. & Neeser, J. R. (1997).** Structural characterization of the exocellular polysaccharides produced by *Streptococcus thermophilus* SFi39 and SFi12. *Appl Environ Microbiol* **63**, 3512-3518.
- Lin, W. S., Cunneen, T. & Lee, C. Y. (1994).** Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J Bacteriol* **176**, 7005-7016.
- Lindmark-Mansson, H., Fonden, R. & Pettersson, H. E. (2003).** Composition of swedish dairy milk. *Int Dairy J* **13**, 409-425.
- Lloret, J., Wulff, B. B., Rubio, J. M., Downie, J. A., Bonilla, I. & Rivilla, R. (1998).** Exopolysaccharide II production is regulated by salt in the halotolerant strain *Rhizobium meliloti* EFB1. *Appl Environ Microbiol* **64**, 1024-1028.

- Llull, D., Garcia, E. & Lopez, R. (2001).** Tts, a processive beta-glucosyltransferase of *Streptococcus pneumoniae*, directs the synthesis of the branched type 37 capsular polysaccharide in Pneumococcus and other gram-positive species. *J Biol Chem* **276**, 21053-21061.
- Llull, D., Lopez, R., Garcia, E. & Munoz, R. (1998).** Molecular structure of the gene cluster responsible for the synthesis of the polysaccharide capsule of *Streptococcus pneumoniae* type 33F. *Biochim Biophys Acta* **1443**, 217-224.
- Llull, D., Munoz, R., Lopez, R. & Garcia, E. (1999).** A single gene (*tts*) located outside the *cap* locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. *J Exp Med* **190**, 241-251.
- Looijesteijn, P. J. & Hugenholtz, J. (1999).** Uncoupling of growth and exopolysaccharide production by *Lactococcus lactis* subsp. *cremoris* NIZO B40 and optimization of its synthesis. *Journal of Bioscience and Bioengineering* **88**, 159-163.
- Looijesteijn, P. J., Trapet, L., de Vries, E., Abee, T. & Hugenholtz, J. (2001).** Physiological function of exopolysaccharides produced by *Lactococcus lactis*. *Int J Food Microbiol* **64**, 71-80.
- Lopez, R. & Garcia, E. (2004).** Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiol Rev* **28**, 553-580.
- Mack, D. (1999).** Molecular mechanisms of *Staphylococcus epidermidis* biofilm formation. *J Hosp Infect* **43 Suppl**, S113-125.
- Mack, D., Rohde, H., Dobinsky, S., Riedewald, J., Nedelmann, M., Knobloch, J. K., Elsner, H. A. & Feucht, H. H. (2000).** Identification of three essential regulatory gene loci governing expression of *Staphylococcus epidermidis* polysaccharide intercellular adhesin and biofilm formation. *Infect Immun* **68**, 3799-3807.
- Maguin, E., Prevost, H., Ehrlich, S. D. & Gruss, A. (1996).** Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J Bacteriol* **178**, 931-935.
- Maguin, E., Duwat, P., Hege, T., Ehrlich, D. & Gruss, A. (1992).** New thermosensitive plasmid for gram-positive bacteria. *J Bacteriol* **174**, 5633-5638.
- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. & Bryant, S. H. (2005).** CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33 Database Issue**, D192-196.
- Marck, C. (1988).** 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res* **16**, 1829-1836.
- Mercenier, A. (1990).** Molecular genetics of *Streptococcus thermophilus*. *FEMS Microbiol Rev* **7**, 61-77.

- Mollerach, M., Lopez, R. & Garcia, E. (1998).** Characterization of the *galU* gene of *Streptococcus pneumoniae* encoding a uridine diphosphoglucose pyrophosphorylase: a gene essential for capsular polysaccharide biosynthesis. *J Exp Med* **188**, 2047-2056.
- Monsan, P., Bozonnet, S., Albenne, C., Joucla, G., Willemot, R. M. & Remaud-Simeon, M. (2001).** Homopolysaccharides from lactic acid bacteria. *International Dairy Journal* **11**, 675-685.
- Morona, J. K., Morona, R. & Paton, J. C. (1999).** Analysis of the 5' portion of the type 19A capsule locus identifies two classes of *cpsC*, *cpsD*, and *cpsE* genes in *Streptococcus pneumoniae*. *J Bacteriol* **181**, 3599-3605.
- Morona, J. K., Paton, J. C., Miller, D. C. & Morona, R. (2000).** Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *streptococcus pneumoniae*. *Mol Microbiol* **35**, 1431-1442.
- Morona, J. K., Morona, R., Miller, D. C. & Paton, J. C. (2002).** *Streptococcus pneumoniae* capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. *J Bacteriol* **184**, 577-583.
- Munoz, R., Mollerach, M., Lopez, R. & Garcia, E. (1997).** Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. *Mol Microbiol* **25**, 79-92.
- O'Sullivan, T. F. & Fitzgerald, G. F. (1999).** Electrotransformation of industrial strains of *Streptococcus thermophilus*. *J Appl Microbiol* **86**, 275-283.
- Pavlovic, G. (2004).** Evolution d'une famille d'éléments intégratifs potentiellement conjuguatifs et/ ou mobilisables de *Streptococcus thermophilus*: Université Henri Poincaré, Nancy1.
- Pavlovic, G., Burrus, V., Gintz, B., Decaris, B. & Guedon, G. (2004).** Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICES_{St1}-related elements from *Streptococcus thermophilus*. *Microbiology* **150**, 759-774.
- Petry, S., Furlan, S., Crepeau, M. J., Cerning, J. & Desmazeaud, M. (2000).** Factors affecting exocellular polysaccharide production by *Lactobacillus delbrueckii* subsp. *bulgaricus* grown in a chemically defined medium. *Appl Environ Microbiol* **66**, 3427-3431.
- Pluvinet, A. (2003).** Polymorphisme génétique du locus *eps* (Exopolysaccharide) chez la bactérie lactique *Streptococcus thermophilus*: Université Henri Poincaré, Nancy1.
- Pluvinet, A., Charron-Bourgoin, F., Morel, C. & Decaris, B. (2004).** Polymorphism of *eps* loci in *Streptococcus thermophilus* : sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. *International Dairy Journal* **14**, 627-634.
- Poyart, C., Quesne, G., Coulon, S., Berche, P. & Trieu-Cuot, P. (1998).** Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**, 41-47.

- Rahn, A., Drummelsmith, J. & Whitfield, C. (1999).** Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J Bacteriol* **181**, 2307-2313.
- Roberts, I. S. (1996).** The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol* **50**, 285-315.
- Robson, R. L., Robson, R. M. & Morris, J. G. (1974).** The biosynthesis of granulose by *Clostridium pasteurianum*. *Biochem J* **144**, 503-511.
- Rocha, E. P., Danchin, A. & Viari, A. (1999).** Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res* **27**, 3567-3576.
- Roussel, Y., Pebay, M., Guedon, G., Simonet, J. M. & Decaris, B. (1994).** Physical and genetic map of *Streptococcus thermophilus* A054. *J Bacteriol* **176**, 7413-7422.
- Roussel, Y., Bourgoïn, F., Guedon, G., Pebay, M. & Decaris, B. (1997).** Analysis of the genetic polymorphism between three *Streptococcus thermophilus* strains by comparing their physical and genetic organization. *Microbiology* **143** (Pt 4), 1335-1343.
- Rozen, S. & Skaletsky, H. (2000).** Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386.
- Ruas-Madiedo, P., Hugenholtz, J. & Zoon, P. (2002).** An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *International Dairy Journal* **12**.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988).** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular cloning : a laboratory manual*. Cold Spring Harbor, New-York: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467.
- SantaLucia, J., Jr. (1998).** A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-1465.
- Schaffer, C. & Messner, P. (2005).** The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. *Microbiology* **151**, 643-651.
- Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998).** SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857-5864.

- Segura, M., Gottschalk, M. & Olivier, M. (2004).** Encapsulated *Streptococcus suis* inhibits activation of signaling pathways involved in phagocytosis. *Infect Immun* **72**, 5322-5330.
- Seo, E. J., Yoo, S. H., Oh, K. W., Cha, J., Lee, H. G. & Park, C. S. (2004).** Isolation of an exopolysaccharide-producing bacterium, *Sphingomonas* sp. *CS101*, which forms an unusual type of sphingan. *Biosci Biotechnol Biochem* **68**, 1146-1148.
- Stingle, F. & Mollet, B. (1996b).** Disruption of the gene encoding penicillin-binding protein 2b (*pbp2b*) causes altered cell morphology and cease in exopolysaccharide production in *Streptococcus thermophilus* Sfi6. *Mol Microbiol* **22**, 357-366.
- Stingle, F., Neeser, J. R. & Mollet, B. (1996a).** Identification and characterization of the *eps* (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J Bacteriol* **178**, 1680-1690.
- Stingle, F., Newell, J. W. & Neeser, J. R. (1999a).** Unraveling the function of glycosyltransferases in *Streptococcus thermophilus* Sfi6. *J Bacteriol* **181**, 6354-6360.
- Stingle, F., Vincent, S. J., Faber, E. J., Newell, J. W., Kamerling, J. P. & Neeser, J. R. (1999b).** Introduction of the exopolysaccharide gene cluster from *Streptococcus thermophilus* Sfi6 into *Lactococcus lactis* MG1363: production and characterization of an altered polysaccharide. *Mol Microbiol* **32**, 1287-1295.
- Terzaghi, B. & Sandine, W. (1975).** Improved medium for lactic streptococci and their bacteriophages. *Appl Environ Microbiol* **29**, 807-813.
- Thibessard, A., Fernandez, A., Gintz, B., Decaris, B. & Leblond-Bourget, N. (2002).** Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can J Microbiol* **48**, 473-478.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.
- Tsakalidou, E., Zoidou, E., Pot, B., Wassill, L., Ludwig, W., Devriese, L. A., Kalantzopoulos, G., Schleifer, K. H. & Kersters, K. (1998).** Identification of streptococci from Greek Kasseri cheese and description of *Streptococcus macedonicus* sp. nov. *Int J Syst Bacteriol* **48 Pt 2**, 519-527.
- van den Bogaard, P. T., Kleerebezem, M., Kuipers, O. P. & de Vos, W. M. (2000).** Control of lactose transport, beta-galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus*: evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *J Bacteriol* **182**, 5982-5989.
- van Kranenburg, R., Marugg, J. D., van, S., II, Willem, N. J. & de Vos, W. M. (1997).** Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Mol Microbiol* **24**, 387-397.

- van Kranenburg, R., Vos, H. R., van, S., II, Kleerebezem, M. & de Vos, W. M. (1999a).** Functional analysis of glycosyltransferase genes from *Lactococcus lactis* and other gram-positive cocci: complementation, expression, and diversity. *J Bacteriol* **181**, 6347-6353.
- van Kranenburg, R., van, S., II, Marugg, J. D., Kleerebezem, M. & de Vos, W. M. (1999b).** Exopolysaccharide biosynthesis in *Lactococcus lactis* NIZO B40: functional analysis of the glycosyltransferase genes involved in synthesis of the polysaccharide backbone. *J Bacteriol* **181**, 338-340.
- Vuong, C., Kocianova, S., Voyich, J. M., Yao, Y., Fischer, E. R., DeLeo, F. R. & Otto, M. (2004).** A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J Biol Chem* **279**, 54881-54886.
- Walsh, F. M. & Amyes, S. G. (2004).** Microbiology and drug resistance mechanisms of fully resistant pathogens. *Curr Opin Microbiol* **7**, 439-444.
- Wang, L., Liu, D. & Reeves, P. R. (1996).** C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598-2604.
- Wang, L. Y., Li, S. T. & Li, Y. (2003).** Identification and characterization of a new exopolysaccharide biosynthesis gene cluster from *Streptomyces*. *FEMS Microbiol Lett* **220**, 21-27.
- Weiner, R., Langille, S. & Quintero, E. (1995).** Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* **15**, 339-346.
- Wexler, D. L., Hudson, M. C. & Burne, R. A. (1993).** *Streptococcus mutans* fructosyltransferase (*fif*) and glucosyltransferase (*gtfBC*) operon fusion strains in continuous culture. *Infect Immun* **61**, 1259-1267.
- Whitfield, C. & Roberts, I. S. (1999).** Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* **31**, 1307-1319.
- Whitfield, C. & Paiment, A. (2003).** Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. *Carbohydr Res* **338**, 2491-2502.
- Zuker, M. (2003).** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415.

ANNEXE 1

Article n°1 : sous presse (*International Dairy Journal*)

**The *eps* locus of *Streptococcus thermophilus* IP6756 is not involved
in exopolysaccharide production**

Guillaume TYVAERT^a, Catherine MOREL^a, Jean-Pierre JOLY^b, Bernard DECARIS^a
and Florence CHARRON-BOURGOIN^{a*}

^a Laboratoire de Génétique et Microbiologie, UMR INRA1128 – IFR110, Université Henri Poincaré Nancy 1, Faculté des Sciences et Techniques, BP239, Vandoeuvre-lès-Nancy 54506, France

^b Groupe SUCRES, UMR CNRS7565, Université Henri Poincaré Nancy 1, Faculté des Sciences et Techniques, BP239, Vandoeuvre-lès-Nancy 54506, France

*Corresponding author. Tel.: 33-3-83-68-42-06; Fax: 33-3-83-68-44-99, *E-mail address*:

Florence.charron-bourgoin@scbiol.uhp-nancy.fr

Abstract

The sequence analysis of the *eps* locus of *Streptococcus thermophilus* IP6756 shows that it is very different from all other *eps* loci studied in *S. thermophilus*: it is smaller and displays only few open reading frames (ORF) compared to other *eps* loci. Moreover, none of the identified ORF encodes glycosyltransferases, which are required for EPS production, whereas IP6756 is an EPS-producing strain. The functional study of this locus shows that it is organized in operon and that its polar interruption does not lead to any modification in EPS yield. All these results indicate that the *eps* locus of *S. thermophilus* IP6756 is not involved in EPS production and that, for the first time in *S. thermophilus*, EPS production is under the control of gene(s) encoded elsewhere in the genome.

Keywords: exopolysaccharide; sequence replacement; functional study; lactic acid bacteria

1. Introduction

Many strains of *Streptococcus thermophilus* produce extracellular polysaccharides. These molecules are completely secreted into the medium as slime (exopolysaccharides or EPS) or tightly linked to the cell wall as capsular polysaccharide (CPS). The term of EPS has been used to design either type of polysaccharide (Ruas-Madiedo, Hugenholtz & Zoon, 2002). Extracellular polysaccharides were widespread among bacteria species. They could protect the cells from different stresses or phagocytosis, and could be involved in symbiosis or in biofilm formation.

S. thermophilus has a GRAS (Generally Recognized As Safe) status in the dairy industry and produces EPS with interesting rheological properties. They could reduce syneresis and enhance texture and viscosity of yoghurt. Biochemical and genetic studies showed that the structural diversity of these molecules is correlated with the genetic diversity

of the *eps* locus, which contains the genes involved in EPS biosynthesis. The distribution of regulatory and structural genes within each of these clusters shows a modular organization which is conserved (Jolly & Stingle, 2001). The first module is strongly maintained among studied strains and contains genes involved in regulation of EPS production and in determination of polymerization and chain length of the EPS. The second module is highly polymorphic and contains genes coding glycosyltransferases involved in repeating unit formation and genes involved in EPS polymerization and export. Some authors consider that the *eps* locus is located between the ORF *deoD* and *orf14.9* (Jolly & Stingle, 2001). However, it has been demonstrated that some genes located immediately downstream of *orf14.9* are also implicated in EPS production. The ORF *orf5* and *pgm* can be considered as the right border of *eps* loci (Pluvinet, Charron-Bourgoin, Morel & Decaris, 2004).

An illustration of the *eps* locus polymorphism in *S. thermophilus* was obtained by hybridization of probes isolated from the *S. thermophilus* CNRZ368 *eps* locus with DNA of 17 *S. thermophilus* strains (Charron-Bourgoin, Pluvinet, Morel & Decaris, 2001). These strains could be arranged in six groups according to the presence or the absence of hybridization signal with the different probes tested. One of these groups only contained the IP6756 strain, which has a specific hybridization profile: the upstream region displayed signals only with the *epsA* and *epsB* specific probes and not with the *epsC* and *epsD* specific probes, whereas all other strains displayed signals with each of these four probes. Our study was focused on the IP6756 *eps* locus to determine its organization and its involvement in EPS production.

2. Materials and methods

2.1 Bacterial strains, plasmid and culture conditions

S. thermophilus IP6756, purchased from the Institut Pasteur (IP), was routinely grown in 10% reconstituted skim milk, M17 or Belliker media, like other *S. thermophilus* strains used : from the IP collection IP6757, from the Centre National de la Recherche Zootechnique (CNRZ) collection CNRZ-7, -308, -368, -385, -388, -391, -404, -445, -464, -702, -1066, -1068 and -1402, from the BCCM/LMG collection LMG18311, from the American Type Culture Collection (ATCC) ATCC-19258, -19987, and from the laboratory collection NST-1, -3, -5, -10, -11, -12 and -2280. For *S. thermophilus* transformants and recombinant clones, 5 $\mu\text{g mL}^{-1}$ of erythromycin was added.

Cultures used for transformation experiments of *S. thermophilus* ATCC19258, NST2280 and IP6756 were performed in Belliker Threonine (20 mM) medium. At O.D.= 0.3, the medium was replaced by Belliker Threonine (20 mM) medium supplemented with 10% (w/v) glycine. Then following steps were carried out as previously described (O'Sullivan & Fitzgerald, 1999).

The recombinant plasmid pGHost9epsA was used to transform *Escherichia coli* TG1 (Stratagene, La Jolla, USA) by electroporation according to Dower, Miller and Ragsdale (1988). Resulting transformants were selected on Luria-Bertani medium containing 150 $\mu\text{g mL}^{-1}$ of erythromycin.

2.2 DNA manipulation

DNA manipulations such as preparation of plasmid DNA from *E. coli*, restriction nuclease digestions, ligation, agarose gel electrophoresis and Southern blot hybridizations of DNA were performed according to Sambrook, Fritsch and Maniatis (1989). *S. thermophilus* genomic DNA was extracted as previously described (Bourgoin, Pluvinet, Gintz, Decaris &

Guedon, 1999). *S. thermophilus* plasmidic DNA was extracted as previously described by Frere (1994).

Conventional PCR were performed as previously reported (Saiki et al., 1988) and were used to produce specific probes of *epsJ* (primers 5'-ACAGACCTGCCTCCTTTC-3' and 5'-GCAAGACGTCCACAGAAG-3') and *epsZ* (primers 5'-GGGTGCCTTGTCATCGAC-3' and 5'-GAGGCCCAACAAAGCCA-3') of *S. thermophilus* IP6756 and a specific probe of *tts* (primers 5'-TGAGGCCTTTGGTATTGG-3' and 5'-GCCTTAAAGCCATTTCGAT-3') of *Streptococcus pneumoniae* type 37. The whole *eps* locus was amplified by Long-Range PCR according to supplier conditions (Takara, Gennevilliers, France) (primers 5'-GGAAGAAGAAGACTTATAATA-3' and 5'-CCTGAATTCTGACTGAAAGGCAC-3').

Sequencing was carried out with the DNA sequencing kit (Applied Biosystems, Villebon sur Yvette, France) on a ABI Prism 377 genetic analyser (Applied Biosystems). Sequence data were analysed with BLAST, mfold3.0, ClustalW, Autoassembler 2.1.1, SignalP 3.0 and HMMTOP softwares. The 8121 bp DNA sequence obtained in this work was registered in GenBank under accession number AY916787.

2.3 RNA manipulation

Extractions of total RNA were performed as previously described (Fernandez, Thibessard, Borges, Gintz, Decaris & Leblond-Bourget, 2004) with modifications: cells were cultured in M17 medium and harvested at the beginning of stationary phase. DNA-free RNA was obtained by DnaseI (Amersham Pharmacia Biotech, Orsay, France) treatment (0.5 units μg^{-1} RNA) for 15 min at 37°C. The absence of contaminating genomic DNA was checked by PCR.

cDNA was generated from 2.5 µg of RNA with 100 units of MMLV-reverse transcriptase at 37°C for 60 min in the presence of hexamer oligonucleotide random primers at 100 ng µL⁻¹ (Amersham Pharmacia Biotech, Orsay, France).

Primers for RT-PCR (Table 1) were designed with primer3 software and purchased from MWG Biotech (Villebon sur Yvette, France) and Invitrogen (Cergy-Pontoise, France). Amplification of specific products was performed on the iCycler (BioRad, Marnes la Coquette, France) with the following program : 1 cycle at 94°C for 5 min; 35 cycles at 94°C for 30 s, 49°C for 30 s and 72°C for 30 s; 1 cycle at 72°C for 7 min.

2.4 Integration of pGHost9epsA in IP6756 *epsA* by homologous recombination

A 730 bp PCR product corresponding to the internal part of *epsA* of IP6756 (primers 5'-CCTTCAGAATTCTAAGAAATTACC-3' and 5'-GTCAATCAACTGCAGGAAAGAT-3') was inserted into the pGHost9 plasmid (Maguin, Prevost, Ehrlich & Gruss, 1996) to generate pGHost9epsA. It was introduced into *S. thermophilus* NST2280 and ATCC19258 by electroporation. Plasmid DNA of ATCC19258 transformant was extracted and used to transform *S. thermophilus* IP6756. Its integration into the chromosome by single-crossover was performed as described by Burrus, Bontemps, Decaris and Guedon (2001). Total DNA of putative pGHost9epsA co-integrants of IP6756 and NST2280 was extracted and subjected to Southern Blot analysis to verify integrated plasmid location and copy number. Recombinant clones IP6756-A17 and NST2280-A8 were obtained by multiple integration of pGHost9epsA plasmid in IP6756 and NST2280 *epsA* gene respectively.

2.5 EPS isolation, quantification and composition

EPS of *S. thermophilus* strains were extracted as described by Robson, Robson and Morris (1974), on a 12 hours culture in 10% reconstituted skim milk at 42°C. Proteins were removed from the culture with an equal volume of trichloroacetic acid (40%). After

centrifugation, an equal volume of acetone was added to precipitate the EPS. Then the EPS were dissolved in water. The total neutral osidic content was determined by the phenol-sulphuric acid method (Dubois, Gilles, Hamilton, Rebers and Smith, 1956).

The sugar components of EPS produced by *S. thermophilus* IP6756 were examined by thin layer chromatography (TLC) as previously described by Seo, Yoo, Oh, Cha, Lee and Park (2004). The extracted EPS were hydrolyzed with 4M trifluoroacetic acid (TFA) at 100°C for 2 h and concentrated in a rotative evaporator (Büchi, Switzerland). The TLC plate was activated in an oven at 100°C for 30 min. Prepared samples were spotted on the plate and then eluted in a TLC chamber with the solvent n-propanol:water:triethylamine:30% NH₃ = 80:20:0.2:4 (v/v/v/v) at room temperature. Then the plate was dried, sprayed with 30% H₂SO₄ in methanol and heated at 170°C until complete development.

3. Results and discussion

3.1 EPS production of *S. thermophilus* IP6756

The EPS yield of *S. thermophilus* IP6756 was 120 ± 32 mg L⁻¹. To determine if this strain could be qualified as EPS-producing strain, the EPS yield of milk medium without bacteria inoculation was measured (11 mg L⁻¹). Otherwise, the EPS-producing strain NST2280 (170 ± 36 mg L⁻¹) produced only 25 mg L⁻¹ when *epsN* was interrupted (Pluvinet et al., 2004). Thus, *S. thermophilus* IP6756 is an EPS-producing strain.

The TLC analysis showed that the IP6756 EPS contains mainly glucose, some galactose and traces of an unidentified monosaccharide (Figure 1).

3.2 The *eps* locus of *S. thermophilus* IP6756

The *eps* locus of *S. thermophilus* IP6756 was amplified using primers corresponding to conserved regions which border *S. thermophilus* *eps* loci (Stingele, Neeser & Mollet, 1996;

Pluvinet et al., 2004). The 8121 bp amplified DNA includes the sequence between the *deoD-epsA* intergenic region and *pgm*. The ORF *deoD* would be the left border of the identified *eps* locus, as in the other *eps* loci studied previously. Four ORF strongly related to *eps* genes of *S. thermophilus* (*epsA*, *epsB*, *epsZ* and *epsJ*), two truncated IS elements, *orf14.9* and a partial sequence corresponding to *orf5* were identified between these extremities (Figure 2A). Upstream of each *eps* putative genes, putative ribosome binding sites (RBS) were detected. *orf14.9*, which is identified in the locus, is present in all the studied *eps* loci of *S. thermophilus* (hybridization results and available sequences) and its sequence is highly conserved. The 3'-region of the obtained sequence contains $\Delta ISS1$, $\Delta orf5$ and the 5'-end of *pgm* and displayed 98.2% and 97% identity with the similar region of the *eps* locus of respectively *S. thermophilus* IP6757 (Pluvinet, Charron-Bourgoin, Morel & Decaris, accession number : AJ289861) and MTC310 (Broadbent, McMahon, Welker, Oberg & Moineau, accession number: AF448502). The ORF *orf5* and *pgm* were present in the common ancestor of the *S. thermophilus* strains and were considered as the right border of *eps* loci (Pluvinet et al., 2004).

The first 2091 bp region includes the *deoD-epsA* intergenic region, *epsA* and the 465 bp 5'-end of *epsB* and shares 98% identity with the corresponding sequences of the *eps* loci of *S. thermophilus* Sfi39 (Germond & Delley, accession number: AF373595) and MTC310 (Figure 2B) and more than 96% with the corresponding sequences of other sequenced *eps* loci of *S. thermophilus*. In IP6756, *epsB* is 489 bp long compared to 732 bp long in the other *eps* loci. The second region includes the 24 bp 3'-end of *epsB*, a 15 bp intergenic region and the 86 bp 5'-end of *epsZ*. This 125 bp sequence shares no homology with any known sequence. The third region includes the 1330 bp 3'-end of *epsZ* and *epsJ* and shares 98% identity with the corresponding sequences of Sfi39 and MTC310 (Figure 2B). In Sfi39, this region contains a 207 bp intergenic sequence, the pseudo-ORF *epsI**, a 148 bp intergenic sequence and *epsJ*.

In MTC310, it contains the 1195 bp 3'-end of the pseudogene *rut*, a 146 bp intergenic region and *epsJ*. The 1330 bp 3'-end of *epsZ* also displayed 98 and 97 % identity with *eps10L* and *eps6L* respectively, encoding putative repeating unit transporter in *eps* loci type X and VI respectively (Rallu, Ehrlich and Renault, accession number : AF454500, and AF454497 respectively). Hybridization of *epsZ* and *epsJ* specific probes with genomic DNA of 23 *S. thermophilus* strains (IP-6756, -6757, CNRZ-7, -308, -368, -385, -388, -391, -404, -445, -464, -702, -1066, -1068 and -1402, LMG18311, ATCC19987, NST-1, -3, -5, -10, -11 and -12) showed a single signal for IP6756, CNRZ388 and CNRZ445 with the *epsZ* specific probe and for IP6756 and CNRZ388 with the *epsJ* specific probe (data not shown). The presence of both ORF was only detected in *S. thermophilus* IP6756 and CNRZ388.

This 8121 bp sequence revealed 98% identity with the similar sequence of the *eps* loci of *S. thermophilus* Sfi39 and MTC310, except the 125 bp including the 3'-end of *epsB* and the 5'-end of *epsZ* (Figure 2). This difference observed in the central part of the above-mentioned *eps* loci suggests the involvement of sequence replacement, and horizontal transfer acquisition. Such a phenomenon was previously suggested for the formation of the *eps* locus of *S. thermophilus* IP6757 (Pluvinet et al., 2004). In *S. pneumoniae*, a similar mechanism has been described for 19F and 23F serotypes (Coffey et al., 1998). In IP6756, *epsZ* displays a putative ribosome binding site and contains two sequences with different origins: the 86 bp 5'-end shares no homology with any known sequence and the 1330 bp 3'-end shares 98% identity with a 1330 bp sequence containing *epsI** in Sfi39 and *rut* in MTC310, which are both pseudo-ORF. The putative ORF *epsZ* has a mosaic structure and could be functional.

The *eps* locus of *S. thermophilus* IP6756 only contains four ORF, for which functions could be proposed by sequence homology: *epsA* and *epsB* would be involved in regulation of gene expression as previously described (Jolly & Stinglele, 2001), *epsZ* in repeating units transport and *epsJ* in conversion of pyranose molecule to the furanose form, since the putative

EpsJ protein shares 99.4% identity with the putative EpsJ mutase of Sfi39. Although *S. thermophilus* IP6756 produces EPS, none of the ORF in the identified locus would encode a glycosyltransferase. This strongly suggests that EPS production of this strain involves gene(s) located elsewhere in the genome.

3.3 Functional study of the *eps* locus of *S. thermophilus* IP6756

Sequence analysis indicated that the *deoD-epsA* region contains a putative promoter and that a putative transcription terminator is located downstream of *epsJ* ($\Delta G = -33 \text{ kcal mol}^{-1}$ at 37°C according to mfold3.0 software). The *eps* gene expression could be directed by a single transcript containing the whole cluster as described in *S. thermophilus* Sfi6 (Stingele, Vincent, Faber, Newell, Kamerling & Neeser, 1999) and *Lactococcus lactis* NIZO B40 (van Kranenburg, Marugg, van Swam, Willem & de Vos, 1997).

A specific transcript for *epsA*, *epsB*, *epsZ* and *epsJ* (344 bp, 370 bp, 420 bp and 510 bp respectively, figure 3A) and transcripts overlapping *epsA-epsB*, *epsB-epsZ* and *epsZ-epsJ* (410 bp, 468 bp and 611 bp respectively, figure 3B) were detected. These data suggest that the IP6756 *eps* locus displays an operonic transcript, like *eps* loci studied in the previously mentioned strains.

The IP6756-A17 mutant was obtained by interruption of *epsA*, which is the first ORF of this operonic locus. In this mutant, a specific transcript for *orf14.9*, an ORF located at the 3'-end of the *eps* locus of IP6756 and which does not belong to the *eps* operon (Figure 2A), was detected by RT-PCR and corresponds to a positive control. In return, no transcript for *epsB*, *epsZ* and *epsJ* could be detected by RT-PCR (data not shown), showing the polar effect of this interruption. The determination of EPS yield showed no difference between this IP6756-A17 mutant and *S. thermophilus* IP6756 (Figure 4). When such a polar interruption of the *eps* locus of *S. thermophilus* NST2280 was performed (NST2280-A8 mutant), the EPS

production is abolished (Figure 4). So, contrary to what is observed in *S. thermophilus* NST2280, the *eps* locus of IP6756 may be not involved in EPS production on quantitative term.

This result is in accordance with the sequence analysis of the *eps* locus of *S. thermophilus* IP6756, since the absence of ORF encoding putative glycosyltransferase indicates that other gene(s) would be involved in EPS production in the IP6756 strain. This is the first time that such a situation is described in *S. thermophilus*. A lot of glycosyltransferase genes could act in EPS biosynthesis. Genes encoding the priming glycosyltransferase found in various gram-positive and proteobacteria are fairly conserved among strains (Wang, Liu & Reeves, 1996), whereas sequence diversity of other glycosyltransferase genes is huge. Hybridization of glycosyltransferase-specific probes elaborated from the *eps* locus of *S. thermophilus* CNRZ368, NST2280 and IP6757 on genomic DNA of *S. thermophilus* IP6756 did not reveal any signal (Bourgoin et al., 1999; Pluvinet et al., 2004), indicating that the other genes involved in EPS production of IP6756 would be totally different from those of these *eps* loci.

Our observation is similar to that described in *Streptococcus pneumoniae* type 37 (Llull, Muñoz, Lopez & Garcia, 1999), where a classical *cps* locus, identical to the *cps* locus of type 33, contains inactivated ORF and is not involved in CPS production. This strain uses a single enzyme encoded by the *tts* gene to produce a homopolysaccharide (Llull et al., 1999). Hybridization of a *tts* specific probe with genomic DNA of IP6756 did not reveal any signal (data not shown) indicating that sequences strongly related to this *tts* gene are not present in the IP6756 genome. Since *S. thermophilus* IP6756 produces a heteropolysaccharide, the mechanism for EPS production in this strain is different from that of *S. pneumoniae* type 37 and should involved several glycosyltransferases.

4. Conclusions

The *eps* locus of *S. thermophilus* IP6756 displays a peculiar organization compared to all other *eps* loci studied in *S. thermophilus* and represents a good illustration of the huge variability of *eps* loci in this species. The sequence analysis of the IP6756 *eps* locus showed that its 125 bp central region from unknown origin could result from sequence replacement involving unknown horizontal transfer and integration mechanisms. Moreover, it showed that none of the identified ORF encode glycosyltransferases, which are required for EPS production, whereas IP6756 is an EPS-producing strain. The functional study indicated that the identified *eps* locus is not involved in EPS production. This study shows, for the first time in this species, that gene(s) located outside the *eps* locus, especially glycosyltransferase gene(s), are required for EPS production. Identification of the involved gene(s) would indicate if this situation is specific to the IP6756 strain and if gene(s) involved in other cellular functions could play a role in EPS production. More generally, this study leads to a better comprehension of biology of EPS production in *S. thermophilus*. In the dairy industry, such knowledge may allow for genetic engineering of novel EPS.

Acknowledgements

We thank Ernesto Garcia (Departamento de Microbiología Molecular, CSIC, Madrid, Spain) for providing genomic DNA of *S. pneumoniae* type37.

References

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B., & Guedon, G. (1999). Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene*, 233, 151-161.

- Burrus, V., Bontemps, C., Decaris, B., & Guedon, G. (2001). Characterisation of a novel type II restriction-modification system, *Sth368I*, encoded by the integrative element *ICES1* of *Streptococcus thermophilus* CNRZ368. *Applied and Environmental Microbiology*, 67, 1522-1528.
- Charron-Bourgoin, F., Pluvinet, A., Morel, C., & Decaris, B. (2001). Polymorphism of *eps* loci involved in exopolysaccharide synthesis of *Streptococcus thermophilus*. *Lait*, 81, 281-288.
- Coffey, T.J., Enright, M.C., Daniels, M., Morona, J.K., Morona, R., Hryniewicz, W., Paton, J.C., & Spratt, B.G. (1998). Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Molecular Microbiology*, 27, 73-83.
- Dower, W.J., Miller, J.F., & Ragsdale, C.W. (1988). High efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Research*, 16, 6127-6145.
- Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., & Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Analytical Chemistry*, 28, 350-356.
- Fernandez, A., Thibessard, A., Borges, F., Gintz, B., Decaris, B., & Leblond-Bourget, N. (2004). Characterisation of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368. *Archives of Microbiology*, 182(5), 364-372.
- Frere, J. (1994). Simple method for extracting plasmid DNA from lactic acid bacteria. *Letters Applied Microbiology*, 18, 227-229.
- Jolly, L., & Stinglele, F. (2001). Molecular organization und fonctionnalité of exopolysaccharide genes clusters in lactic acid bacteria. *International Dairy Journal*, 11, 733-745.

- Llull, D., Munoz, R., Lopez, R., & Garcia, E. (1999). A single gene (*tts*) located outside the *cap* locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide: type 37 pneumococci are natural, genetically binary strains. *Journal of Experimental Medicine*, 190, 241-251.
- Maguin, E., Prevost, H., Ehrlich, S.D., & Gruss, A. (1996). Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *Journal of Bacteriology*, 178, 931-935.
- O'Sullivan, T.F., & Fitzgerald, G.F. (1999). Electrotransformation of industrial strains of *Streptococcus thermophilus*. *Journal of Applied Microbiology*, 86, 275-283.
- Pluvinet, A., Charron-Bourgoin, F., Morel, C., & Decaris, B. (2004). Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. *International Dairy Journal*, 14, 627-634.
- Robson, R.L., Robson, R.M., & Morris, J.G. (1974). The biosynthesis of granulose by *Clostridium pasteurianum*. *Journal of Biochemistry*, 144, 503-511.
- Ruas-Madiedo, P., Hugenholtz, J., & Zoon, P. (2002). An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *International Dairy Journal*, 12, 163-171.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., & Ehrlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, 487-491.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Seo, E.J., Yoo, S.H., Oh, K.W., Cha, J., Lee, H.G., & Park, C.S. (2004). Isolation of an exopolysaccharide-producing bacterium, *Sphingomonas* sp. CS101, which forms an unusual type of sphingan. *Bioscience Biotechnology and Biochemistry*, 68, 1146-1148.

- Stingele, F., Neeser, J.R., & Mollet, B. (1996). Identification and characterization of the *eps* (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *Journal of Bacteriology*, 178, 1680-1690.
- Stingele, F., Vincent, S.J.F., Faber, E.J., Newell, J.W., Kamerling, J.P., & Neeser, J.R. (1999). Introduction of the exopolysaccharide gene cluster from *Streptococcus thermophilus* Sfi6 into *Lactococcus lactis* MG1363: production and characterization of an altered polysaccharide. *Molecular Microbiology*, 32, 1287-1295.
- van Kranenburg, R., Marrug, J.D., van Swam, I.I., Willem, N.J., & de Vos, W.M. (1997). Molecular characterization of the plasmid encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Molecular Microbiology*, 24, 387-397.
- Wang, L., Liu, D., & Reeves, P.R. (1996). C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *Journal of Bacteriology*, 178, 2598-2604.

Figure captions

Figure 1: TLC analysis of the carbohydrate composition of EPS produced by *S. thermophilus* IP6756.

A: standards, B: hydrolyzed EPS and standards, C: hydrolyzed EPS. The standards are galacturonic acid (1), glucuronic acid (2), galactose (3), glucose (4) and rhamnose (5).

Figure 2: Comparison of the *eps* loci of *S. thermophilus* IP6756 (A), Sfi39 (accession number: AF373595) and MTC310 (accession number: AF448502) (B).

Since nucleotide sequences of the *eps* loci of Sfi39 and MTC310 strains display 99% identity, only the Sfi39 *eps* locus is shown. Nucleotide sequences with more than 98% identity are represented by the same symbols. Localization of the specific region of each *eps* locus is indicated (white box in IP6756 and hatching box in Sfi39). The putative *eps* locus promoter is symbolized by an arrow. The putative transcription terminator is symbolized by . *epsI** is a pseudo-ORF. Δ corresponds to deleted ORF or IS element. The double crossed line does not respect the scale.

Figure 3: Transcription study of the *S. thermophilus* IP6756 *eps* locus.

A: Identification of specific transcript for *epsJ* (1), *epsZ* (2), *epsB* (3) and *epsA* (4) on 2% agarose gel electrophoresis.

B: Identification of transcripts overlapping *epsZ-epsJ* (1'), *epsB-epsZ* (2') and *epsA-epsB* (3') on 2% agarose gel electrophoresis.

c, r, g: amplification by PCR on obtained cDNA (c), on DNA-free RNA (r), on genomic DNA (g); n: negative control; λ: DNA of λ phage digested by *Pst*I.

Figure 4: Yield of EPS in IP6756, NST2280 and their respective *epsA* insertional mutants.

Grey and open bars correspond to EPS produced by the wild strains and by their *epsA* insertional mutants respectively. Standard error is figured.

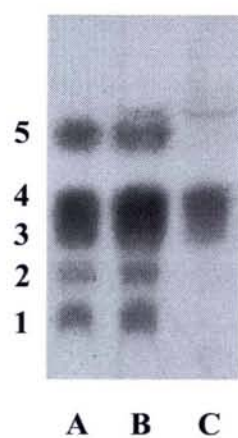


Figure 1

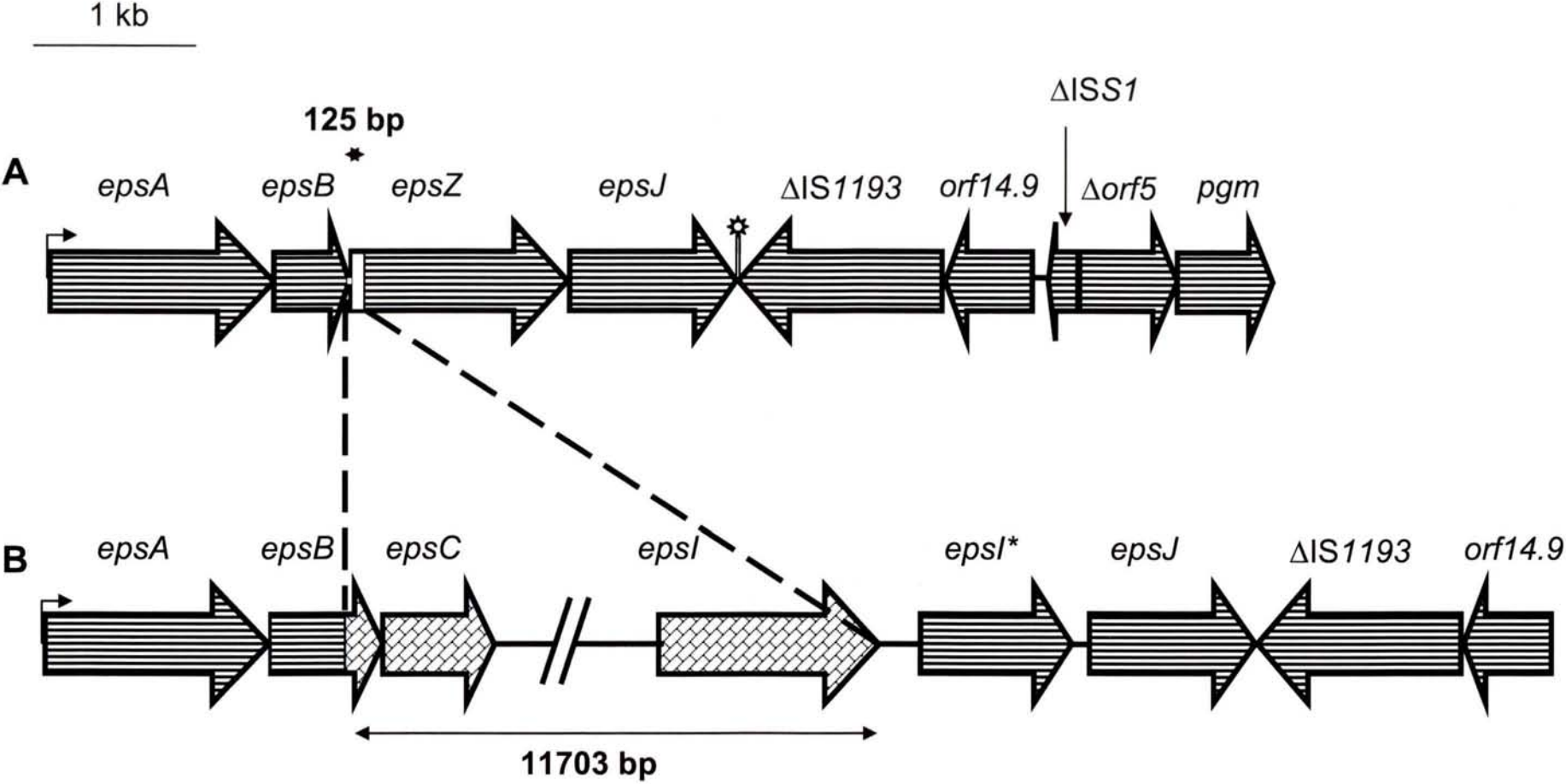


Figure 2

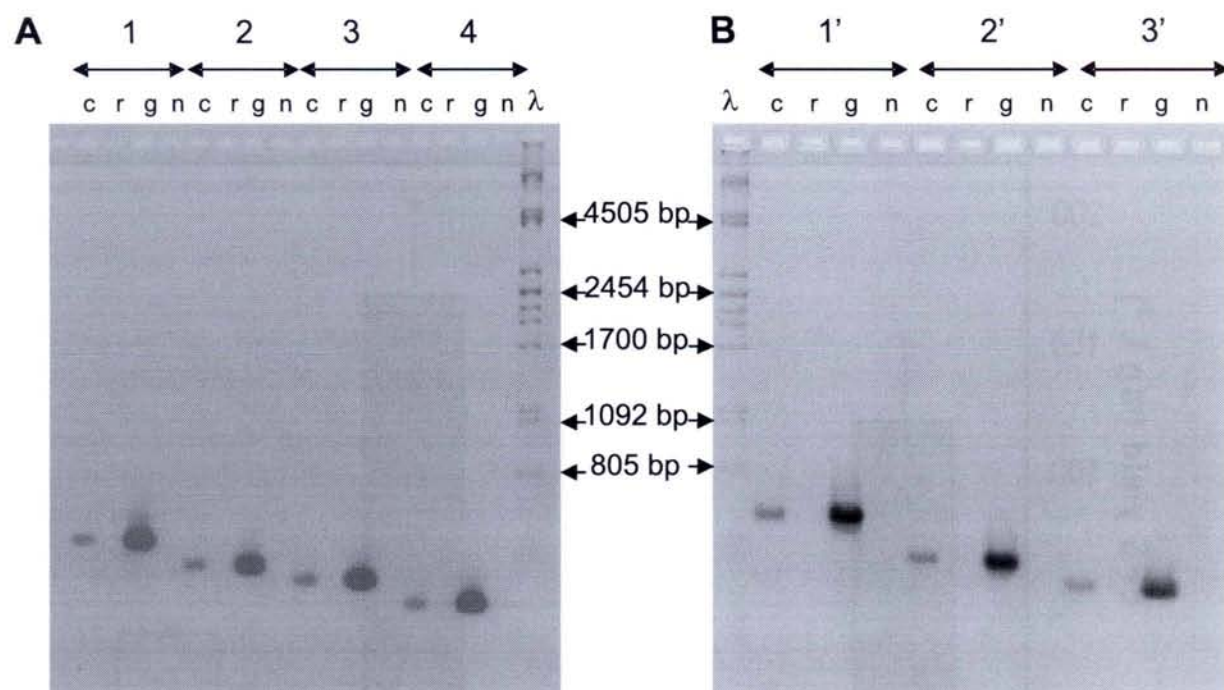


Figure 3

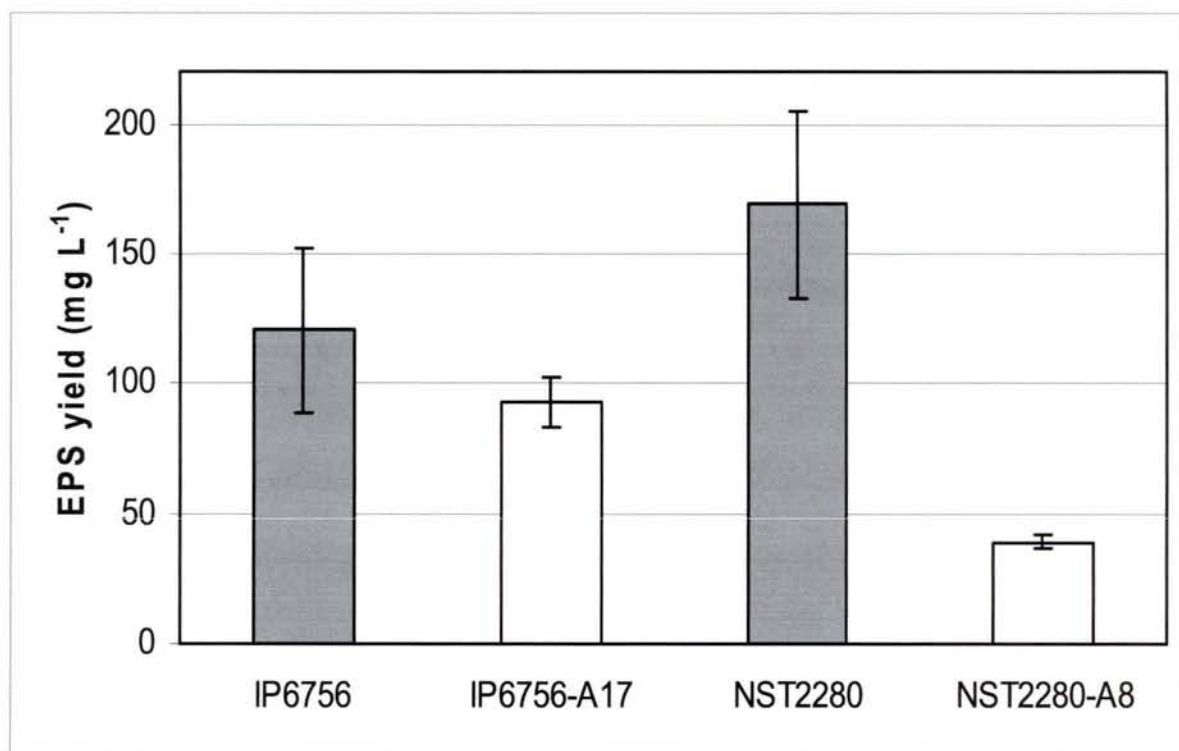
**Figure 4**

Table 1: Primers used in transcription studies.

Primers	Sequence (5'-3')	Primer function
o-epsA5	CAGGCACCTACTAAGGTTG	transcription of <i>epsA</i>
o-epsA6a	ACATCTGAACGTGACACTG	
o-epsB1	CATTAGTGAAAGTTACGCCC	transcription of <i>epsB</i>
o-epsA8	CTTGCAATTCCTAGTGTCG	
o-epsI*a	GGGTGCCTTGTCATCGAC	transcription of <i>epsZ</i>
o-epsI*b	GAGGCCCAACAAAGCCA	
o-epsJA'	ACAGACCTGCCTCCTTTC	transcription of <i>epsJ</i>
o-epsJ1	GCAAGACGTCCACAGAAG	
o-epsA7	GACCGTGGTAAAAACCAGG	transcript overlapping <i>epsA</i> -
o-epsBa	GGTCTAAACTTTCTTCTAAAG	<i>epsB</i>
o-epsB3'	TCATAGTGGGCTTAGTAATG	transcript overlapping <i>epsB</i> -
o-eps1	TGCATACCAGGAAAGAACAA	<i>epsZ</i>
o-eps2	AGCAGTAGATGGTCTTGTC	transcript overlapping <i>epsZ</i> -
o-epsJ5'	GAAAGGAGGCAGGTCTGT	<i>epsJ</i>

ANNEXE 2

Article n°2 : soumis (*Archives of Microbiology*)

**The constant gene *orf14.9*, which belongs to the variable *eps* (exopolysaccharide) cluster,
is involved in the cell growth of *Streptococcus thermophilus***

Guillaume TYVAERT, Catherine MOREL, Bernard DECARIS
and Florence CHARRON-BOURGOIN*

Laboratoire de Génétique et Microbiologie, UMR INRA/UHP 1128 - IFR110 GEEF,
Université Henri Poincaré (UHP) Nancy 1, Faculté des Sciences et Techniques, BP239,
Vandoeuvre-lès-Nancy 54506, France

* Corresponding author. Tel.: 33-(0)3-83-68-42-06; Fax: 33-(0)3-83-68-44-99

E-mail address: Florence.charron-bourgoin@scbiol.uhp-nancy.fr

Abstract: In *Streptococcus thermophilus*, the *eps* clusters involved in exopolysaccharides (EPS) biosynthesis are very polymorphic, nevertheless all contain a highly conserved sequence corresponding to *orf14.9*. This ORF is oriented oppositely to the operonic ORF of the 5'-end of the *eps* cluster and is transcribed. Sequence analysis showed that the putative Orf14.9 protein is a transmembrane protein, but its function is unknown. Insertional mutant of *orf14.9* was obtained in NST2280 and A054 *S. thermophilus* strains. EPS yields of the obtained mutants are similar to those of their respective wild strains suggesting that *orf14.9* does not modify the quantity of produced EPS. Another phenotypic analysis showed that *orf14.9* is involved in the cell growth of *S. thermophilus*.

Keywords: *Streptococcus thermophilus*, *eps* cluster, *orf14.9*, sequence analysis, EPS biosynthesis, cell growth

Introduction

Streptococcus thermophilus is a Firmicute bacterium: the genome of this Gram-positive bacteria displays a low guanine and cytosine content. *S. thermophilus* belongs to the lactic acid bacteria group and is used in the manufacture of dairy products, especially yoghurts. Most strains of this species are able to produce exopolysaccharides (EPS). These molecules are either completely secreted into the medium as slime or tightly linked to the cell wall as a capsule. The EPS of *S. thermophilus* display some interesting rheological properties, since they reduce syneresis and enhance texture and viscosity of yoghurt (Ruas-Madiedo et al. 2002).

Biochemical and genetic studies showed that the structural diversity of the molecules is correlated with the genetic diversity of the *eps* cluster, which contains the genes involved in EPS biosynthesis. The *eps* clusters of *S. thermophilus* can be divided into two parts. The 5'-

part displays a modular and operonic organization (Jolly and Stingele 2001; Stingele et al. 1999): one module is constant and is involved in EPS biosynthesis regulation whereas the other is highly variable and is involved in EPS biosynthesis. It contains genes encoding glycosyltransferases which are involved in repeating unit formation. The 3'-part of the *eps* clusters of *S. thermophilus* contains various ORF and insertion sequences (IS) among strains. Nevertheless, according to the available sequences of *eps* clusters, this part always contains *orf14.9*. Some authors considered this ORF as the right border of the *eps* cluster in *S. thermophilus* (Jolly and Stingele 2001). However, it has been demonstrated that some genes located immediately downstream of *orf14.9* are also involved in EPS production, indicating that the right border of these clusters would be *orf5* and *pgm* (Pluvinet et al. 2004).

To date, the function of *orf14.9* is unknown. Since this ORF is present in the 3'-part of all the *eps* clusters sequenced and is not affected by its variability, this ORF could confer a selective advantage to *S. thermophilus*. After an *in-silico* analysis of *orf14.9* sequence, the functional study of this ORF is presented.

Materials and methods

Bacterial strains, plasmid and culture conditions

The 13 *S. thermophilus* strains used (Table 1) and the IP6756 and NST2280 *S. thermophilus* strains from the Institut Pasteur (IP) and the laboratory collections respectively are routinely grown in 10% reconstituted skim milk, M17, TPPY or Belliker media. For *S. thermophilus* transformants and recombinant clones, 5 $\mu\text{g mL}^{-1}$ of erythromycin is added.

Cultures used for transformation experiments of *S. thermophilus* NST2280 and A054 are performed in Belliker Threonine (20 mM) medium. When Optical Density at 600 nm (O.D.₆₀₀) reaches 0.3, the medium is replaced by Belliker Threonine (20 mM) medium

supplemented with 10% (w/v) glycine. Then following steps are carried out as previously described (O'Sullivan and Fitzgerald 1999).

The recombinant plasmid pGHost9orf14.9 was used to transform *Escherichia coli* TG1 (Stratagene, La Jolla, USA) by electroporation according to Dower et al. (1988). Resulting transformants were selected on Luria-Bertani medium containing 150 $\mu\text{g mL}^{-1}$ of erythromycin.

DNA manipulation

DNA manipulations such as preparation of plasmid DNA from *E. coli*, restriction nuclease digestions, ligation, agarose gel electrophoresis and Southern blot hybridizations of DNA are performed according to Sambrook et al. (1989). *S. thermophilus* genomic and plasmidic DNA are extracted as previously described by Bourgoïn et al. (1999) and Frere (1994) respectively.

Conventional PCR was performed as previously reported (Saiki et al. 1988) by using the Biolabs *Taq* polymerase (Saint Quentin Yveline, France) at a melting temperature (T_m) of 49°C, in order to produce an internal fragment of *orf14.9* in 13 *S. thermophilus* strains and a specific probe of *orf14.9* of *S. thermophilus* NST2280 (primers 5'-CTCCCAACAACAATAATACC-3' and 5'-CCTCAGCAGTTTATTGGTC-3').

Sequencing is carried out with the DNA sequencing kit (Applied Biosystems, Villebon sur Yvette, France) on an ABI Prism 377 genetic analyser (Applied Biosystems). Sequence data are analysed with BlastN, BlastX and BlastP of the NCBI (Altschul et al. 1990 ;1997), mfold3.0 (Zuker 2003), ClustalW (Thompson et al. 1994), SignalP3.0 (Bendtsen et al. 2004), SMART (Schultz et al. 1998) and CDART (Geer et al. 2002) softwares.

RNA manipulation

Extractions of total RNA are performed as previously described (Chomczynski and Sacchi 1987; Fernandez et al. 2004) with some modifications: cells are cultured in M17 medium and harvested at the beginning of stationary phase. DNA-free RNA is obtained by DnaseI (Amersham Pharmacia Biotech, Orsay, France) treatment ($0.5 \text{ units } \mu\text{g}^{-1} \text{ RNA}$) for 15 min at 37°C . The absence of contaminating genomic DNA is checked by PCR.

Complementary DNA is generated from $2.5 \text{ } \mu\text{g}$ of RNA with 100 units of MMLV-reverse transcriptase at 37°C for 60 min in the presence of hexamer oligonucleotide random primers at $100 \text{ ng } \mu\text{L}^{-1}$ (Amersham Pharmacia Biotech, Orsay, France).

Primers for RT-PCR (primers $5'\text{-CTCCCAACAACAATAATACC-3'}$ and $5'\text{-CCTCAGCAGTTTATTGGTC-3'}$) are designed with primer3 software (Rozen and Skaletsky 2000) and purchased from MWG Biotech (Villebon sur Yvette, France). Amplification of specific products is performed on the iCycler (BioRad, Marnes la Coquette, France) with the following program: 1 cycle at 94°C for 5 min; 35 cycles at 94°C for 30 s, 49°C for 30 s and 72°C for 30 s; 1 cycle at 72°C for 7 min.

Integration of pGHost9orf14.9 in NST2280 and A054 *orf14.9* by homologous recombination

A 451 bp PCR product corresponding to the internal part of *orf14.9* of NST2280 (primers $5'\text{-GCGACGAATTCTTGGCTGATTATGCCCG-3'}$ and $5'\text{-TAAGAATTCTAATGGCAAATGTGCCC-3'}$) is inserted into the unique *EcoRI* site of the pGHost9 plasmid (Maguin et al. 1992) to generate pGHost9orf14.9. This PCR product is introduced into *S. thermophilus* NST2280 and A054 by electroporation. Its integration into the chromosome by homologous recombination is performed as described by Burrus et al. (2001). Total DNA of putative pGHost9orf14.9 co-integrants of NST2280 and A054 are

extracted and subjected to Southern Blot analysis to verify location and copy number of integrated plasmid.

EPS isolation and quantification

EPS of *S. thermophilus* strains are extracted on a 12 hours culture in 10% reconstituted skim milk at 42°C as described by Pluvinet et al. (2004). Proteins are removed from the culture with an equal volume of trichloroacetic acid (40%). After centrifugation, an equal volume of acetone is added to precipitate the EPS. Then the EPS are dissolved in water. The total neutral osidic content is determined by the phenol-sulphuric acid method (Dubois et al. 1956).

Determination of the growth parameters

A culture is performed in 50 mL of M17 (A054 and A054-O12) or TPPY (NST2280 and NST2280-O7) medium. At O.D.₆₀₀ = 0.5, ten 1/50 cultures are performed. Each culture is used to measure O.D.₆₀₀ at a precise time and to establish the growth curve. This experiment is performed three times for each strain.

Results and discussion

***orf14.9* is conserved in all the *S. thermophilus* strains studied**

Sequences closely related to *orf14.9* were detected in the 13 *S. thermophilus* strains studied by hybridization of an *orf14.9* specific probe isolated from the *S. thermophilus* NST2280 *eps* cluster (data not shown). A 260 bp internal fragment of *orf14.9* was sequenced and showed 96 to 100% identity with the NST2280 allele (Table 1). More, *orf14.9* is present in all the sequenced *S. thermophilus eps* clusters. To date, twenty three sequences of *orf14.9* alleles are available and their coding DNA sequence (CDS) showed the same variability as

above mentioned (Table 2). The comparison of all these sequences showed the high conservation of the *orf14.9* CDS. The sequence comparison of *lacS* or *lacZ* of 29 *S. thermophilus* strains which encodes lactose permease and β -galactosidase respectively showed such low variability (Ercolini et al., 2005) and suggests that *orf14.9* could confer a selective advantage to *S. thermophilus*.

Nevertheless, a heterogeneity of sequence length can be observed: most of *orf14.9* alleles are 576 bp long but some display a length of 570 bp (MR-1C, CNRZ368 and type IV *eps* cluster), 564 bp (type V and VII *eps* cluster) and 558 bp (type X *eps* cluster) (Table 2). According to the author choice for the translational start codon (TTG or ATG, Figure 1a), the *orf14.9* allele displays 576 or 564 bp length. The translational stop codon of the 570 and 558 bp alleles is located upstream of that of the 576 bp length alleles, as a consequence of a punctual mutation in the 3'-end of *orf14.9* (Figure 1b). The *orf14.9* sequence is highly conserved among *S. thermophilus* strains and the length heterogeneity is due to difference in author annotation or to minor sequence modification in the 3'-end.

Transcriptional activity of *orf14.9* in *S. thermophilus*

The figure 2 shows the localization of *orf14.9* in three types of *eps* clusters of *S. thermophilus* which correspond to the different situations according to the presence and the number of ORF involved in EPS biosynthesis located upstream of *orf14.9*. In each *eps* cluster, *orf14.9* is oriented oppositely to the operonic ORF of the 5'-end.

Upstream of each *orf14.9* available sequence, more than 100 bp are highly conserved among the strains. This sequence contains a putative Ribosome Binding Site (RBS) highly conserved among the strains (mainly AAAGGAGT) and very similar to the consensus sequence AAAGGAGG. It also contains putative -10 boxes (mainly TAAAT) and -35 boxes (for example ATGACA in IP6756 *eps* cluster) which are closely related to their

respective consensus sequence TATAAT and TTGACA. Downstream of *orf14.9* of NST2280, mfold3.0 software detected a potential transcriptional terminator, which displays $\Delta G = -39.6 \text{ kcal mol}^{-1}$ at 37°C according to this software.

To determine if *orf14.9* is transcribed, total RNA of three EPS-producing *S. thermophilus* strains (NST2280, A054 and IP6756) were extracted and subjected to RT-PCR experiments. A 535 bp *orf14.9* specific transcript was observed for NST2280 (Figure 3) and for A054 and IP6756 (data not shown). According to the high sequence conservation of the CDS and the promoter sequence of *orf14.9*, this ORF should be transcribed and probably functional in all the *eps* clusters of *S. thermophilus*.

Involvement of *orf14.9* in the cell growth

The BlastX and BlastP tools of the NCBI detected no homology between the potential 191 aa Orf14.9 protein of the NST2280 strain and other proteins of known function. The SMART software indicates that the putative Orf14.9 protein should contain six transmembrane segments (from amino acids 7 to 29, 44 to 63, 70 to 92, 112 to 134, 141 to 160 and 164 to 186) and suggests that it should be located in the membrane. However, the signalP3.0 server detected no peptide signal. A domain research with the CDART software suggested that the amino acids 99 to 191 of the putative Orf14.9 protein correspond to a VanZ domain (E-Value: $4. \cdot 10^{-12}$). The VanZ protein is involved in teicoplanin resistance in *Enterococcus faecium* (Arthur et al. 1995), but the mechanism of this low-level resistance is to date unknown (Walsh and Amyes 2004).

To study the function of *orf14.9*, the recombinant clones NST2280-O7 and A054-O12 were obtained by *orf14.9* interruption. They display multiple integration of pGHost9*orf14.9* plasmid in the *orf14.9* gene of NST2280 and A054 *S. thermophilus* strains respectively. Since *orf14.9* is always located in the *eps* cluster, its function would be related to EPS biosynthesis.

The EPS yields of these *orf14.9* insertional mutants are similar to those of their respective wild strains (Figure 4) suggesting that *orf14.9* does not modify the quantity of produced EPS in *S. thermophilus*. However, no qualitative data are available (for example: monosaccharide composition, molecule length or anchorage of the EPS on the cell).

The comparison of the growth curves of NST2280 and NST2280-O7 strains showed that the mutant enters later in stationary phase (Figure 5). This was also observed for the A054 and A054-O12 strains (data not shown). The determination of the growth parameters showed that each mutant displays a generation time significantly longer than the wild strain and a specific growth rate significantly reduced compared to the wild strain (Table 3). These results suggest that *orf14.9* is involved in the cell growth of *S. thermophilus*.

References

- Almiron-Roig E, Mulholland F, Gasson MJ, Griffin AM (2000) The complete *cps* gene cluster from *Streptococcus thermophilus* NCFB2393 involved in the biosynthesis of a new exopolysaccharide. *Microbiology* 146:2793-2802
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402
- Arthur M, Depardieu F, Molinas C, Reynolds P, Courvalin P (1995) The *vanZ* gene of Tn1546 from *Enterococcus faecium* BM4147 confers resistance to teicoplanin. *Gene* 154:87-92
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340(4):783-795

- Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, Fonstein M, Overbeek R, Kyprides N, Purnelle B, Prozzi D, Ngui K, Masuy D, Hancy F, Burtéau S, Boutry M, Delcour J, Goffeau A, Hols P (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nature Biotech* 22(12):1554-1558
- Bourgoin F, Pluvinet A, Gintz B, Decaris B, Guédon G (1999) Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* 233(1-2):151-161
- Broadbent JR, McMahon DJ, Welker DL, Oberg CJ, Moineau S (2003) Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J Dairy Sci* 86:407-423
- Burrus V, Bontemps C, Decaris B, and Guédon G (2001) Characterisation of a novel type II restriction-modification system, *Sth368I*, encoded by the integrative element *ICES1* of *Streptococcus thermophilus* CNRZ368. *Appl Env Microbiol* 67:1522-1528
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156-159
- Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16(13):6127-6145
- Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric method for determination of sugars and related substances. *Anal Chem* 28:350-356
- Ercolini D, Fusco V, Blaiotta G, Coppola S (2005) Sequence heterogeneity in the *lacSZ* operon of *Streptococcus thermophilus* and its use in PCR systems for strain differentiation. *Res Microbiol* 156:161-172

- Fernandez A, Thibessard A, Borges F, Gintz B, Decaris B, Leblond-Bourget N (2004) Characterization of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368. Arch Microbiol 182:364-372
- Frere J (1994) Simple method for extracting plasmid DNA from lactic acid bacteria. Lett Appl Microbiol 18(4):227-229
- Geer LY, Domrachev M, Lipman DJ, Bryant SH (2002) CDART: protein homology by domain architecture. Genome Res 12(10):1619-1623
- Germond JE, Delley M, D'Amico N, Vincent SJF (2001) Heterologous expression and characterization of the exopolysaccharide from *Streptococcus thermophilus* Sfi39. Eur J Biochem 268:5149-5156
- Jolly L, Stingle F (2001) Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. Int Dairy J 11:733-745
- Maguin E, Duwat P, Hege T, Ehrlich D, Gruss A (1992) New thermosensitive plasmid for gram-positive bacteria. J Bacteriol 174:5633-5638
- O'Sullivan TF, Fitzgerald GF (1999) Electrotransformation of industrial strains of *Streptococcus thermophilus*. J Appl Microbiol 86:275-283
- Pluvinet A, Charron-Bourgoin F, Morel C, Decaris B (2004) Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. Int Dairy J 14:627-634
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386
- Ruas-Madiedo P, Hugenholtz J, Zoon P (2002) An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. Int Dairy J 12:163-171
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839):487-491

- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory (Cold Spring Harbor Laboratory, New-York)
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 95(11):5857-5864
- Stingele F, Neeser JR, Mollet B (1996) Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. J Bacteriol 178(6):1680-1690
- Stingele F, Vincent SJF, Faber EJ, Newell JW, Kamerling JP, Neeser JR (1999) Introduction of the exopolysaccharide gene cluster from *Streptococcus thermophilus* Sfi6 into *Lactococcus lactis* MG1363: production and characterization of an altered polysaccharide. Mol Microbiol 32(6):1287-1295
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-4680
- Walsh FM, Amyes SGB (2004) Microbiology and drug resistance mechanisms of fully resistant pathogens. Curr Op Microbiol 7(5):439-444
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13):3406-3415

Figure captions

Figure 1: Alignment of the 5' (a) and 3' (b)-ends of the NST2280 (576 bp), CNRZ368 (570 bp), type VII (564 bp) and type X (558 bp) *orf14.9* alleles.

The translational start (a) and stop (b) codons are surrounded. The nucleotides shared by at least half the sequences are shaded. The dash corresponds to gap introduced in the alignment.

Figure 2: Localization of *orf14.9* in the *eps* clusters of *S. thermophilus* IP6757 and LMG18311 (a), CNRZ1066, MR-2C, MTC330, MTC360 and NST2280 (b), CNRZ368 and MR-1C (c).

The *eps* clusters of the grouped strains display more than 99% identity. For the other *eps* clusters, no ORF involved in EPS biosynthesis was found upstream of *orf14.9* (IP6756 strain) or was sequenced. The putative *orf14.9* transcription promoter and terminator are symbolized by an arrow and  respectively. The hatched ORF are involved in EPS production. The names of the operonic ORF which belong to the 5'-part of the *eps* cluster are underlined. Δ corresponds to deleted ORF or IS element. The double crossed line does not respect the scale. The dotted boxes correspond to pseudo-ORF.

Figure 3: Identification of *orf14.9* specific transcript by RT-PCR in NST2280 strain on a 2% agarose gel electrophoresis.

Amplification by PCR was performed on obtained cDNA (1), on DNA-free RNA (2), on genomic DNA (3) and on sample without DNA (4). λ : DNA of λ phage digested by *Pst*I.

Figure 4: EPS yield in *S. thermophilus* NST2280 and A054 and in their respective *orf14.9* insertional mutants NST2280-O7 and A054-O12.

Grey and open bars correspond to EPS produced by the wild strains and by their *orf14.9* insertional mutants respectively. Standard error is figured.

Figure 5: Growth curve of the NST2280 (triangle) and NST2280-O7 (square) in TPPY medium at 42°C.

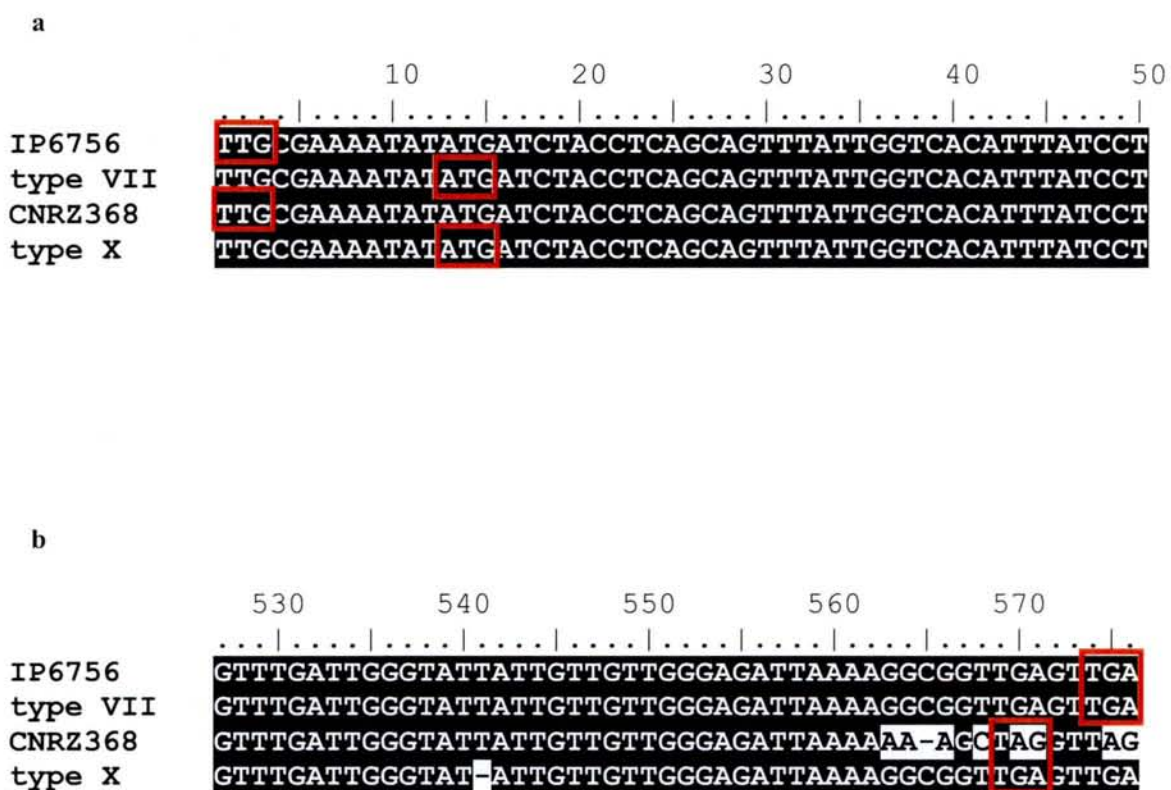


Figure 1

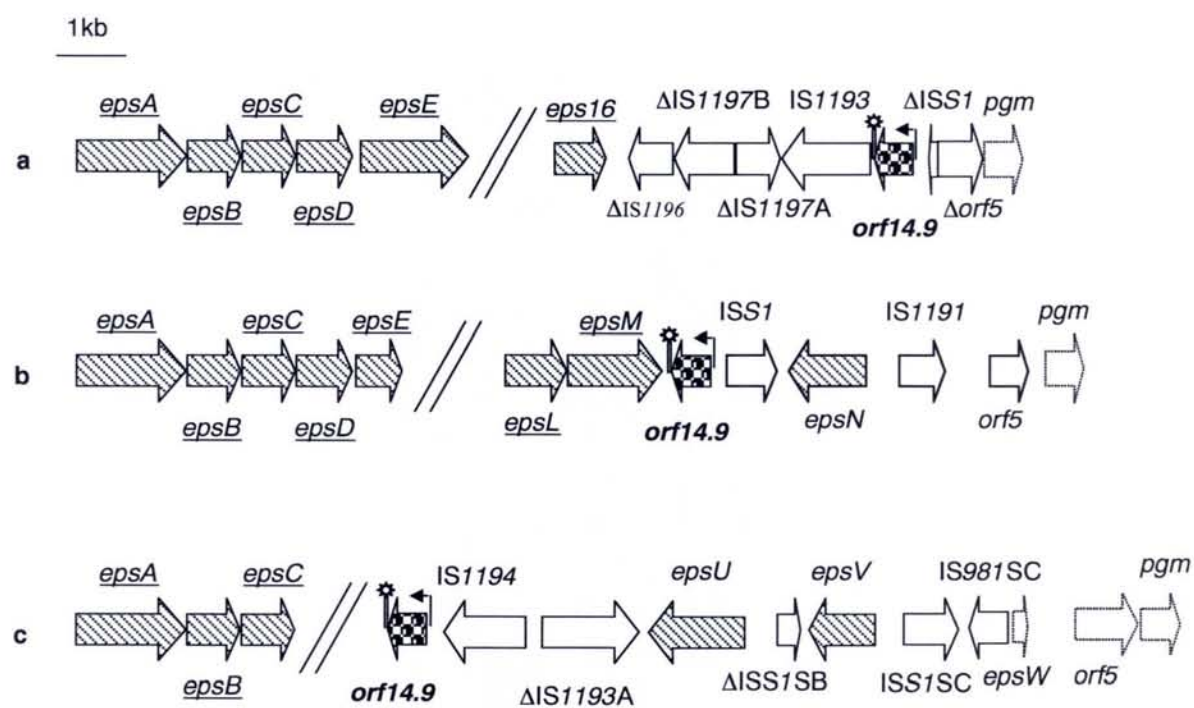
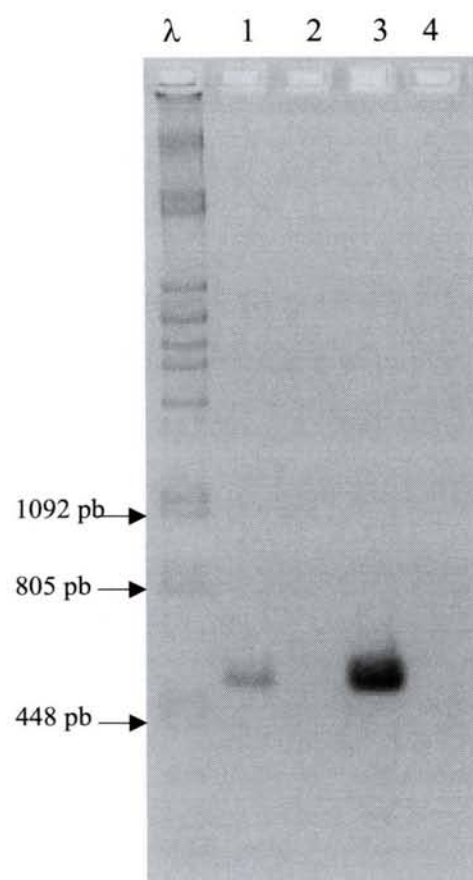
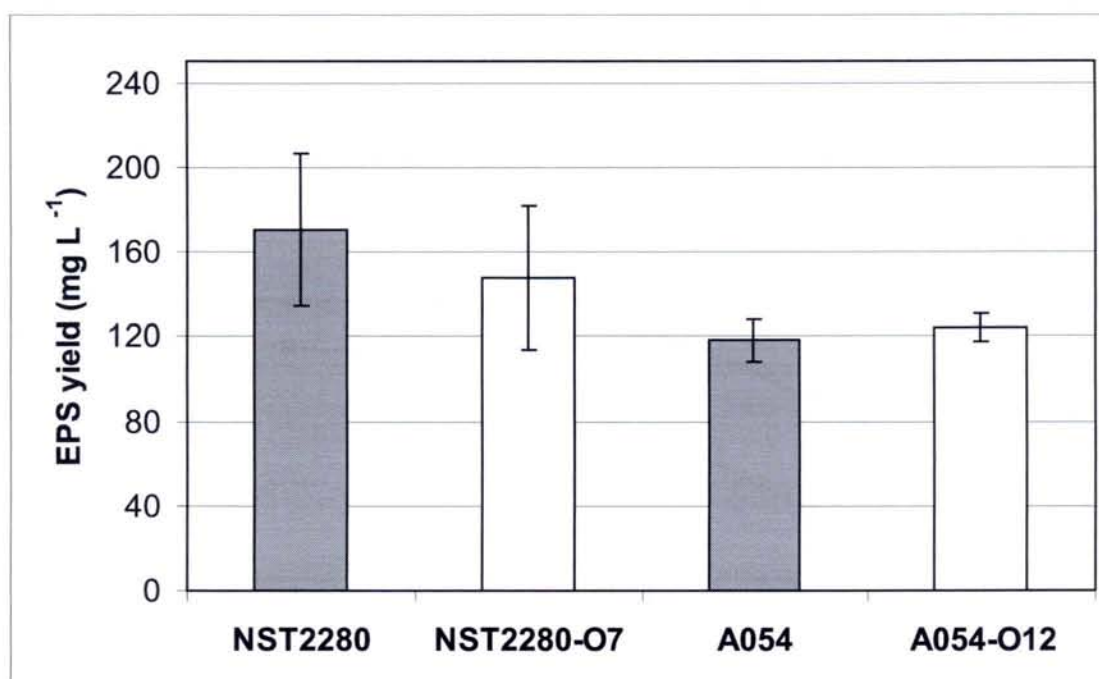


Figure 2

**Figure 3**

**Figure 4**

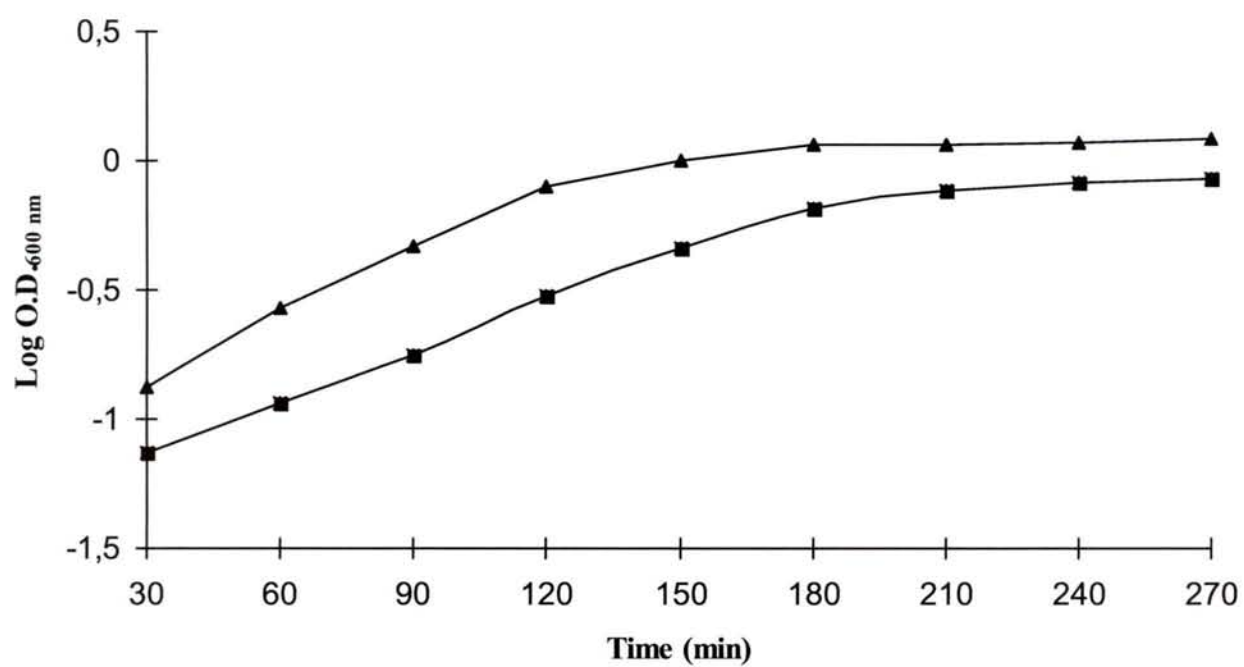


Figure 5

Table 1: Nucleotidic identity (%) of various *S. thermophilus* alleles with the *S. thermophilus* NST2280 allele

<i>S. thermophilus</i> strain ^a	Identity
A054	96,5
ATCC19258	95,8
ATCC19987	95,8
CNRZ302	96,5
CNRZ307	96,5
CNRZ308	97
CNRZ385	96,2
CNRZ388	96,5
CNRZ391	100
CNRZ455	96,2
CNRZ702	97,7
CNRZ1402	96,5
NST1	96,2

^a A054 and NST1 are industrial and laboratory strains respectively, the other are collection strains; CNRZ : Centre National de la Recherche Zootechnique; ATCC : American Type Culture Collection.

Table 2: Characteristics of the *orf14.9* alleles of *eps* clusters in *S. thermophilus*

Strain or <i>eps</i> cluster type	Accession number	Identity ^c	Size	Reference
NST2280	AF517575		576	Pluvinet et al. 2004
CNRZ1066	NC_006449	100	576	Bolotin et al. 2004
CNRZ368	Z98171	97	570	Bourgoin et al. 1999
IP6756	AY916787	97	576	Tyvaert et al. ^d
IP6757	AJ289861	97	576	Pluvinet et al. 2004
LMG18311	NC_006448	97	576	Bolotin et al. 2004
MR-1C	AF448249	97	570	Broadbent et al. 2003
MR-2C	AY061649	100	576	Broadbent et al. 2003
MTC310	AF448502	97	576	Broadbent et al. 2003
MTC330	AF430847	100	576	Welker et al. ^d
MTC360	AF434993	100	576	Broadbent et al. 2003
NCFB2393	Y17900	97	576	Almiron-Roig et al. 2000
Sfi39 ^a	AF373595	97	ND	Germond et al. 2001
Sfi6	U40830	99	576	Stingele et al. 1996
type I ^{a,b}	AF410175	97	ND	Rallu et al. ^d
type III ^b	AY057915	97	576	Rallu et al. ^d
type IV ^b	AF454495	97	570	Rallu et al. ^d
type V ^b	AF454496	97	564	Rallu et al. ^d
type VI ^{a,b}	AF454497	96	ND	Rallu et al. ^d
type VII ^b	AF454498	96	564	Rallu et al. ^d
type IX ^{a,b}	AF454499	100	ND	Rallu et al. ^d
type X ^b	AF454500	97	558	Rallu et al. ^d
type XI ^{a,b}	AF454501	ND	ND	Rallu et al. ^d

^aThe available sequence is partial; ^bThe strain(s) in which this type of *eps* clusters was isolated were not specified; ^cPercentage of nucleotidic identity with the NST2280 allele;

^dunpublished data; ND : not determined

Table 3: Growth parameters of NST2280 and A054 and their respective *orf14.9* insertional mutants NST2280-O7 and A054-O12.

	NST2280	NST2280-O7	A054	A054-O12
μ^a	$0.016 (\pm 2.10^{-3})$	$0.012 (\pm 1.10^{-3})$	$0.026 (\pm 1.10^{-3})$	$0.021 (\pm 1.10^{-3})$
g^b	$43 (\pm 7.5)$	$58 (\pm 5.2)$	$27 (\pm 1.0)$	$33 (\pm 0.8)$

^a μ : specific growth rate (min^{-1}); ^b g : generation time (min); Standard error is figured.

Monsieur TYVAERT Guillaume

DOCTORAT DE L'UNIVERSITE HENRI POINCARÉ, NANCY 1
en GENETIQUE MOLECULAIRE

VU, APPROUVÉ ET PERMIS D'IMPRIMER N° 112

Nancy, le 25 Novembre 2005

Le Président de l'Université



RESUME

Les exopolysaccharides (EPS) produits par *Streptococcus thermophilus* présentent une variabilité importante tant sur le plan structural que sur celui de leur déterminisme génétique. Le cluster *eps* de cette bactérie est ainsi soumis à de nombreux remaniements tels que des remplacements de séquences, par transferts horizontaux.

L'organisation du cluster *eps* de *S. thermophilus* IP6756, de même localisation que les clusters *eps* des autres souches de l'espèce, est très particulière. Ainsi, chez cette souche productrice d'EPS, le module de régulation n'est constitué que de deux (*epsA* et *epsB*) des quatre ORF présentes dans les autres clusters *eps* de l'espèce. De même, le module de biosynthèse et d'export n'est constitué que de deux ORF (*epsZ* et *epsJ*). Enfin, les produits potentiels de ces ORF ne présentent aucune homologie avec des glycosyltransférases, catalysant l'accrochage séquentiel d'unités osidiques lors de la formation des unités répétées constitutives des EPS. L'ensemble de ces observations suggère un fonctionnement différent du déterminisme génétique de la biosynthèse des EPS chez la souche IP6756. Des gènes extérieurs au cluster *eps* identifié seraient ainsi nécessaires à la biosynthèse des EPS.

Sur le plan fonctionnel, le cluster *eps* de cette souche présente une organisation opéronique. Son interruption s'accompagne du maintien de la production dans la souche IP6756. Ce résultat suggère l'intervention d'autres déterminants génétiques. Un ou plusieurs gènes extérieurs au cluster *eps* seraient donc nécessaires à la biosynthèse des EPS de la souche IP6756, notamment ceux codant les glycosyltransférases nécessaires. Chez une bactérie proche de *S. thermophilus*, *Streptococcus pneumoniae*, la biosynthèse des EPS de la souche de type 37 est assurée par un gène extérieur au cluster *eps*, le gène *tts*. Aucune séquence homologue à *tts* n'a été détectée dans le génome de la souche IP6756.

Comme les autres clusters *eps* de *S. thermophilus*, le cluster *eps* de la souche IP6756 contient *orf14.9* dont la fonction est inconnue. Lors de ce travail, il a été montré que cette ORF est transcrite. Les expériences d'interruption réalisées n'ont pas permis d'établir un lien entre la fonctionnalité de cette ORF et la production d'EPS, au moins sur un plan quantitatif. Par contre, il semble que l'intégrité de cette ORF soit nécessaire à une croissance optimale de *S. thermophilus*. La présence d'une copie fonctionnelle d'*orf14.9* dans le cluster *eps* pourrait conférer une valeur sélective supérieure à l'ensemble de ces gènes.

En conclusion, le déterminisme de la biosynthèse des EPS de la souche IP6756 présente une organisation originale, nécessitant l'intervention de gènes extérieurs au cluster *eps* identifié dans la souche. Sur un plan plus général, le cluster *eps* de *S. thermophilus* résulterait d'un regroupement de gènes tendant à augmenter la possibilité de transfert à d'autres bactéries présentes dans la même niche écologique.