



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>

Thèse présentée pour l'obtention du titre de
DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ, Nancy 1

en génétique moléculaire

par
Frédéric Borges

**Création de gènes par réassortiment de modules : caractères
chimérique et variable de *cse*, un gène impliqué dans la
ségrégation cellulaire chez la bactérie
*Streptococcus thermophilus***

Soutenue le 25 avril 2005 devant la commission d'examen :

Membres du jury :

<i>Président:</i>	D. EHRLICH	Directeur de recherche INRA, Jouy-en-Josas
<i>Rapporteurs :</i>	S. AYMERICH A. FOUET	Directeur de recherche INRA, Thiverval-Grignon Directeur de recherche CNRS, Institut Pasteur, Paris
<i>Directeurs de thèse :</i>	N. LEBLOND-BOURGET B. DECARIS	Maître de Conférence, Université Henri Poincaré, Nancy 1 Professeur, Université Henri Poincaré, Nancy 1

Remerciements

Je remercie Monsieur le Professeur Bernard Decaris d'avoir co-encadré ce travail de thèse. Je tiens à le remercier particulièrement pour ses conseils, son écoute et son goût prononcé pour les débats scientifiques.

Je remercie également Madame le Docteur Nathalie Leblond-Bourget d'avoir co-encadré ce travail. Je souhaite la remercier en particulier pour son ouverture d'esprit, ses critiques scientifiques et pour m'avoir fait bénéficier de son sens de la pédagogie. Je voudrais souligner combien il m'a été agréable d'effectuer ce travail de thèse au sein de l'équipe qu'elle dirige.

Je suis particulièrement reconnaissant à Madame le Docteur Agnès Fouet (Institut Pasteur, Paris) et Monsieur le Docteur Stéphane Aymerich (INRA, Thiverval-Grignon) d'avoir accepté de rapporter ce travail, et à Monsieur le Docteur Dusko Ehrlich (INRA, Jouy-en-Josas) d'avoir accepté de l'examiner.

Je remercie Monsieur le Docteur Saulius Kulakauskas (INRA, Jouy-en-Josas) de m'avoir accueilli pendant quelques semaines au sein de son équipe.

Merci à Monsieur le Docteur Pascal Hols (UCL, Belgique) de nous avoir permis l'accès aux séquences des loci *cse* et *pcsB* avant leur parution, et pour sa lecture attentive et ses remarques constructives lors de l'écriture de l'article portant sur le gène *cse*.

Je remercie également Monsieur le Professeur Sylvain Moineau (Université Laval, Canada) d'avoir accepté le travail collaboratif sur les phages avec tant d'enthousiasme.

Merci beaucoup à Monsieur le Professeur Pierre Leblond de toujours avoir laissé la porte de son bureau ouverte, afin de me faire bénéficier de sa clairvoyance et de son sens critique.

Mes remerciements s'adressent également à :

Bertrand Aigle, Gérard Guédon, Annabelle Thibessard, Annabelle Fernandez et Eric Chabrière pour leurs conseils et l'intérêt qu'ils ont porté à ce travail.

Séverine Layec pour son travail consciencieux et son implication qui ont contribué à rendre notre courte interaction productive.

Brigitte Gintz pour sa contribution technique et sa bonne humeur quotidienne.

Frédéric Choulet qui a accepté avec enthousiasme de débiter la recherche de gènes chimériques dans les banques de données de séquences.

à tous les autres membres du laboratoire ayant contribué d'une quelconque façon au bon déroulement de ce travail.

Enfin, je remercie infiniment mes proches de m'avoir supporté et soutenu durant ces quelques années pendant lesquelles je n'ai pas été tous les jours de bonne compagnie.

TABLE DES MATIERES

ABREVIATIONS.....	V
PREAMBULE.....	1
INTRODUCTION.....	2
1 LES CHAINES DE COQUES.....	2
1.1 LA PAROI CELLULAIRE DES BACTERIES A GRAM POSITIF	2
1.1.1 <i>Structure de la paroi</i>	2
1.1.1.1 Introduction	2
1.1.1.2 Le peptidoglycane	2
1.1.1.3 Les polymères secondaires	2
1.1.1.4 La couche S	3
1.1.2 <i>Métabolisme du peptidoglycane</i>	3
1.1.2.1 Biosynthèse du peptidoglycane	3
1.1.2.2 Les hydrolases du peptidoglycane	4
1.1.2.3 Dynamique de la biosynthèse du peptidoglycane	4
1.2 DIVERSITE DE L'ARRANGEMENT DES CELLULES CHEZ LES COQUES	5
1.3 LES FACTEURS GENETIQUES DETERMINANT LE CARACTERE CHAINES DE CELLULES	5
1.3.1 <i>La séparation des cellules résulte de l'hydrolyse de liaisons du peptidoglycane</i>	5
1.3.2 <i>La croissance en chaînes résulterait d'une déficience en hydrolase du peptidoglycane</i>	5
1.3.3 <i>Les gènes impliqués dans la formation des chaînes</i>	6
1.4 LIEN ENTRE LA DIVISION CELLULAIRE ET LA CROISSANCE EN CHAINES	6
1.5 AVANTAGE ADAPTATIF CONFERE PAR LA CROISSANCE EN CHAINES	7
1.5.1 <i>La longueur des chaînes est régulée en fonction de l'environnement</i>	7
1.5.2 <i>Paramètres de l'environnement influençant la longueur des chaînes</i>	7
1.5.3 <i>Acteurs transduisant le signal</i>	7
1.5.4 <i>Gènes régulés</i>	7
1.5.5 <i>La capacité à adhérer à des surfaces abiotiques est modulée en fonction de la longueur des chaînes de cellules</i>	8
2 VARIABILITE GENETIQUE DE REGIONS INTRAGENIQUES RICHES EN MOTIFS REPETES.....	9
2.1 ABONDANCE DES REPETITIONS DANS LES GENOMES BACTERIENS.....	9
2.2 TAILLE ET ORGANISATION DES REPETITIONS, NOTION DE COMPLEXITE.....	9
2.3 IMPORTANTE VARIABILITE DES REGIONS RICHES EN MOTIFS REPETES.....	10
2.4 L'INSTABILITE DES MOTIFS REPETES DE FAÇON DISPERSEE PEUT ENGENDRER LA CREATION DE NOUVEAUX MOTIFS REPETES	11
2.5 MECANISMES DE RECOMBINAISON A L'ORIGINE DE L'EVOLUTION DES REGIONS RICHES EN MOTIFS REPETES DE FAÇON DIRECTE	11
2.5.1 <i>La recombinaison : définition</i>	11
2.5.2 <i>La recombinaison illégitime</i>	12
2.5.2.1 Définition	12
2.5.2.2 Facteurs influençant la fréquence de recombinaison illégitime	12
2.5.2.3 Modèle moléculaire de recombinaison illégitime entre séquences répétées	12
2.5.3 <i>La recombinaison homologue</i>	13
2.5.3.1 Longueur minimale requise des régions homologues.....	13
2.5.3.2 Fréquence	13
2.5.3.3 Crossing over et conversion génique.....	13
2.5.4 <i>Impact de la recombinaison sur l'évolution des régions riches en motifs répétés</i>	13
2.6 CLASSIFICATION DES REPETITIONS D'ADN INTRAGENIQUES EN DEUX CLASSES	14
2.7 L'AVANTAGE ADAPTATIF EST-IL CONFERE PAR LE CARACTERE REITERATIF OU PAR LE CARACTERE INSTABLE DES REGIONS RICHES EN REPETITIONS ?	14
2.7.1 <i>Conséquences fonctionnelles du caractère réitératif des régions riches en répétitions</i>	15
2.7.1.1 Fixation coopérative des protéines à leur ligand : interaction avec les matrices extracellulaires des mammifères	15
2.7.1.2 Fixation à une structure répétée.....	15
2.7.1.2.1 Domaine d'attachement à la paroi.....	15
2.7.1.2.2 Interaction avec l'ADN : les protéines histone-like.....	16

2.7.1.3	Catalyses enzymatiques réitérées permettant la biosynthèse de molécules antibiotiques	16
2.7.2	<i>Conséquences fonctionnelles de la variabilité des régions riches en motifs répétés</i>	17
2.7.2.1	Variabilité des épitopes antigéniques	17
2.7.2.2	Variabilité de l'exposition d'un domaine à la surface des cellules	17
2.7.2.3	Modulation de l'activité du régulateur transcriptionnel CtsR d'un état « ON » à un état « OFF »	18
2.7.2.4	Modulation de la spécificité d'une enzyme de restriction de type I	18
2.7.2.5	Génération de nouvelles spécificités d'hôte chez un phytopathogène	19
3	CREATION DE NOUVELLES FONCTIONS PAR REASSORTIMENT DE DOMAINES	
	PROTEIQUES CHEZ LES BACTERIES	20
3.1	ARCHITECTURE DES PROTEINES	20
3.1.1	<i>Hiérarchie structurale des protéines</i>	20
3.1.2	<i>Organisation en domaines distincts de la structure tertiaire des protéines</i>	20
3.2	CREATION EXPERIMENTALE DE NOUVELLES FONCTIONS PAR CONSTRUCTION DE NOUVELLES COMBINAISONS DE DOMAINES	21
3.2.1	<i>Introduction</i>	21
3.2.2	<i>L'ATCase et l'OTCase</i>	21
3.2.3	<i>CPL7 et LytA</i>	21
3.3	DES REASSORTIMENTS DE DOMAINES ONT ETE SELECTIONNES AU COURS DE L'EVOLUTION	22
3.4	CARACTERE VERSATILE DES DOMAINES	22
3.4.1	<i>Construction de banques de données permettant d'établir un répertoire de combinaisons de domaines</i>	22
3.4.2	<i>Caractéristiques du répertoire de combinaisons de familles de domaines</i>	23
3.4.3	<i>Origine du caractère disparate de la versatilité des domaines</i>	24
3.4.3.1	Contraintes sur la recombinaison des séquences d'ADN	24
3.4.3.2	Certaines fonctions élémentaires sont requises en grand nombre, d'autres pas	24
3.4.3.3	Nécessité d'une co-adaptation des domaines	24
4	OBJECTIFS	26
	MATERIEL ET METHODES	28
1	CONSTRUCTION DE MUTANTS PAR REMPLACEMENT ALLELIQUE	28
2	ANALYSES PHENOTYPIQUES	28
2.1	STRESS OXYDANT	28
2.1.1	<i>Courbe de survie des cellules en présence de ménadione</i>	28
2.1.2	<i>Test de diffusion sur boîte</i>	29
2.2	ADHESION A UNE SURFACE EN VERRE	29
2.3	DOSAGE D'EXOPOLYSACCHARIDES	29
	RESULTATS	31
1	IDENTIFICATION DU GENE CSE PAR CARACTERISATION DE GENES IMPLIQUES DANS LA DEFENSE CONTRE LE STRESS OXYDANT	31
1.1	ISOLEMENT DE ONZE MUTANTS SENSIBLES A LA MENADIONE	31
1.2	IDENTIFICATION DE DIX MUTANTS SENSIBLES AU PARAQUAT	31
1.3	NIVEAU DE SENSIBILITE DU MUTANT 16D10 AU STRESS OXYDANT	32
1.4	LE MUTANT 16D10 PRESENTE DE LONGUES CHAINES	32
2	LE GENE CSE EST CHIMERIQUE, VARIABLE ET MOSAÏQUE ET CODE UNE PROTEINE EXTRACELLULAIRE IMPLIQUEE DANS LA SEGREGATION CELLULAIRE CHEZ STREPTOCOCCUS THERMOPHILUS	34
2.1	LE GENE CSE EST IMPLIQUE DANS LA SEGREGATION CELLULAIRE	34
2.1.1	<i>Le locus interrompu dans le génome du mutant 16D10 est l'ORF cse codant potentiellement une protéine de surface</i>	34
2.1.2	<i>Construction d'un mutant présentant une délétion de l'ORF cse</i>	35
2.1.3	<i>Longueur des chaînes du mutant CNRZ368-Δcse</i>	35
2.1.4	<i>Complémentation par ajout d'une copie sauvage de cse</i>	36
2.1.5	<i>Complémentation du phénotype de chaînes longues de la souche CNRZ368-Δcse</i>	37
2.2	INFLUENCE DE CSE SUR LA CROISSANCE EN MILIEU LIQUIDE ET EN MILIEU GELOSE	37
2.3	INFLUENCE DE CSE SUR L'ADHESION DES CELLULES A UNE SURFACE NON BIOTIQUE	37
2.4	CSE EST UNE PROTEINE EXTRACELLULAIRE	38
2.5	LA PROTEINE CSE EST-ELLE UNE HYDROLASE DU PEPTIDOLYCAN ?	38

2.5.1	<i>Hypothèse</i>	38
2.5.2	<i>Activité lytique d'extraits protéiques bruts de S. thermophilus</i>	39
2.5.3	<i>Activité lytique de fragments de Cse surproduits chez E. coli</i>	39
2.5.4	<i>Complémentation avec cse d'un phénotype de longues chaînes résultant d'une déficience en une hydrolase du peptidoglycane</i>	40
2.6	LA REGION VAR-CSE DU GENE CSE EST VARIABLE ET RICHE EN REPETITIONS.....	40
2.6.1	<i>Mise en évidence de la variabilité de la région var-cse</i>	40
2.6.2	<i>La région var-cse est riche en motifs répétés</i>	41
2.6.3	<i>Structure et organisation des motifs répétés de la région var-cse</i>	41
2.7	CSE EST UN GENE MOSAÏQUE.....	42
2.8	LE GENE CSE EST CHIMERIQUE.....	44
2.9	LES ORF PCSB ET ORF1 PRESENTENT DES REGIONS RICHES EN REPETITIONS.....	45
2.10	L'ORF PCSB EST VARIABLE AU NIVEAU INTRASPECIFIQUE.....	45
2.11	IMPACT DE LA VARIABILITE DE VAR-CSE SUR L'ACTIVITE DE SEGREGATION CELLULAIRE.....	46
2.11.1	<i>Variabilité intraspécifique de la longueur des chaînes de cellules</i>	46
2.11.2	<i>La variabilité intraspécifique de la longueur des chaînes de cellules est corrélée à un facteur de variabilité du locus cse</i>	46
2.11.2.1	Choix des souches.....	46
2.11.2.2	Influence de l'allèle cse _{CNRZ368} sur la longueur des chaînes de la souche LMG18311.....	46
2.11.2.3	Influence d'une copie surnuméraire de cse _{CNRZ368} sur la longueur des chaînes de la souche CNRZ368.....	47
2.11.3	<i>Remplacement de la région var-cse de la souche LMG18311 par celle de la souche CNRZ368 : absence d'effet sur la ségrégation cellulaire</i>	47
3	SELECTION DE VARIANTS [EPS-, ADHESION+] PAR CULTURE EN MILIEU SEMI-LIQUIDE	49
3.1	CORRELATION ENTRE VITESSE DE SEDIMENTATION EN MILIEU SEMI-LIQUIDE ET LONGUEUR DES CHAINES DE CELLULES.....	49
3.2	SELECTION DE VARIANTS DE CROISSANCE EN MILIEU SEMI-LIQUIDE CHEZ S. THERMOPHILUS.....	49
3.3	LES VARIANTS DE SEDIMENTATION SONT AFFECTES DANS LA PRODUCTION D'EPS.....	50
3.4	LE VARIANT DE SEDIMENTATION CNRZ368-6.1 A UN POUVOIR D'ADHESION A UNE SURFACE EN VERRE AUGMENTE.....	50
	DISCUSSION	51
1	LE GENE CSE CODE UNE PROTEINE EXTRACELLULAIRE IMPLIQUEE DANS LA SEGREGATION CELLULAIRE	51
1.1	ARCHITECTURE MODULAIRE DES PROTEINES EXTRACELLULAIRES IMPLIQUEES DANS LA SEGREGATION CELLULAIRE.....	51
1.2	EST-CE QUE LA PROTEINE EXTRACELLULAIRE CSE EST UNE HYDROLASE DU PEPTIDOGLYCANE ?.....	51
1.3	FONCTION DE LA REGION VAR-CSE.....	52
1.3.1	<i>Une charnière interdomaine</i>	52
1.3.2	<i>Un espaceur</i>	53
1.3.3	<i>Un domaine de liaison à la paroi</i>	53
1.3.4	<i>Avantage adaptatif conféré par la variabilité de la région Var-Cse</i>	54
2	ORIGINE ET EVOLUTION DU GENE CSE	55
2.1	ORIGINE ET CONSEQUENCES FONCTIONNELLES DU CARACTERE CHIMERIQUE DE CSE.....	55
2.1.1	<i>Le gène cse aurait été généré par réassortiment de modules</i>	55
2.1.2	<i>La protéine Cse résulterait d'un réassortiment de domaines</i>	55
2.1.3	<i>Conséquence fonctionnelle du réassortiment de domaines à l'origine de Cse</i>	55
2.1.3.1	Introduction.....	55
2.1.3.2	Fonction de PcsB.....	56
2.1.3.3	Divergence fonctionnelle entre Cse et PcsB.....	56
2.1.3.4	Recouvrement fonctionnel partiel entre PcsB et Cse.....	56
2.2	MECANISMES D'EVOLUTION DE LA REGION CENTRALE DU GENE CSE.....	57
2.2.1	<i>Variabilité de la région var-cse</i>	57
2.2.2	<i>La variabilité : une conséquence de l'instabilité génétique résultant de la richesse en répétitions</i>	57
2.2.3	<i>Implication du transfert horizontal</i>	59
2.3	HISTOIRE EVOLUTIVE DE CSE.....	60
2.4	IMPORTANCE DES REGIONS DE FAIBLE COMPLEXITE ET TOLERANT LA VARIABILITE DANS LES REASSORTIMENTS DE DOMAINES.....	60

2.4.1	<i>Exemples de protéines de surface ayant été potentiellement générées par réassortiment de domaines</i>	60
2.4.1.1	La protéine Cse.....	60
2.4.1.2	La protéine PcsB ou SagA.....	61
2.4.1.3	La protéine APF1	62
2.4.2	<i>Les régions variables et de faible complexité pourraient favoriser la génération de nouvelles protéines par réassortiments de domaines</i>	62
3	FACTEURS INFLUENÇANT L'ADHESION DE <i>S. THERMOPHILUS</i> A UNE SURFACE NON BIOTIQUE	64
3.1	INTRODUCTION	64
3.2	INFLUENCE DU GENE <i>CSE</i> SUR LE POUVOIR ADHESIF DE <i>S. THERMOPHILUS</i>	65
3.2.1	<i>L'augmentation de la longueur des chaînes résultant de la délétion de cse induirait une diminution du pouvoir adhésif de S. thermophilus</i>	65
3.2.2	<i>Variabilité de la longueur des chaînes et variabilité du pouvoir adhésif</i>	65
3.3	INFLUENCE DES EXOPOLYSACCHARIDES SUR LE POUVOIR ADHESIF DE <i>S. THERMOPHILUS</i>	65
	PERSPECTIVES	67
1	CONSEQUENCES FONCTIONNELLES DU REASSORTIMENT DE DOMAINES A L'ORIGINE DE <i>CSE</i>	67
2	FONCTION DE LA REGION VAR-CSE DANS LA SEGREGATION CELLULAIRE	68
3	IMPORTANCE DES REGIONS VARIABLES ET DE FAIBLE COMPLEXITE DANS LA GENERATION DE NOUVELLES FONCTIONS	68
4	ADHESION DE <i>S. THERMOPHILUS</i> A LA SURFACE DE LA PAROI INTESTINALE	69
	RÉFÉRENCES	70
	ANNEXES	80

Abréviations

aa	: Acide aminé
ADN	: Acide DésoxyriboNucléique
ARN	: Acide RiboNucléique
CDD	: Conserved Domain Database
CHAP	: Cysteine, Histidine-dependent Amidohydrolase/Peptidase
CNRZ	: Centre National de Recherche Zootechnique
Cse	: Cell segregation
EPS	: ExoPolySaccharide
HPLC	: High-Pressure Liquid Chromatography
HRC1 et 2	: Homologous Region, Copy 1 et 2
IPTG	: Isopropyl- β -D-Thio-Galacto-Pyranoside
Mpb	: Méga paires de bases
ORF	: Open Reading Frame
pb	: paire de bases
PCR	: Polymerase Chain Reaction
PcsB	: Protein for cell separation B
RBS	: Ribosome Binding Site
Rppk	: ribose-phosphate pyrophosphokinase
SDS-PAGE	: Sodium Dodecyl Sulfate – PolyAcrylamide Gel Electrophoresis
SIP	: Surface Immunogenic Protein

Préambule

L'adaptation d'un organisme vivant à un environnement peut passer par la génération de protéines assurant de nouvelles fonctions qui seront sélectionnées si elles s'avèrent utiles à la colonisation de cet environnement. Plusieurs mécanismes de génération de nouvelles fonctions existent. L'un d'eux consiste à accumuler des mutations ponctuelles dans un gène. Toutefois, nécessitant de nombreux événements mutationnels, celui-ci est un mécanisme d'évolution lente. Un autre mécanisme tire parti de l'architecture multi-domaine des protéines dans laquelle chaque domaine assure une fonction élémentaire. Ce mécanisme consiste à créer de nouvelles protéines par réassortiment des domaines protéiques. Dans ce cas, un événement mutationnel unique peut entraîner la génération d'une nouvelle fonction, c'est un mécanisme d'évolution rapide.

Ce dernier mécanisme de génération de nouvelles fonctions est bien connu chez les organismes dont les gènes sont morcelés, c'est-à-dire constitués d'introns et d'exons. D'autres exemples montrent que ce mécanisme pourrait également contribuer à la génération de nouvelles fonctions chez les organismes dont les gènes ne sont pas morcelés, tels que les bactéries. Toutefois, chez ces derniers, les facteurs génétiques qui en sont responsables sont mal connus.

Le travail de thèse présenté ici concerne principalement l'étude du gène *cse* de la bactérie *Streptococcus thermophilus*. Les résultats obtenus montrent notamment que (i) *cse* est impliqué dans la ségrégation cellulaire, (ii) qu'il présente une région centrale riche en répétitions et de séquence variable, et suggèrent (iii) qu'il a été formé par réassortiment de fragments de gène (ou module). Pour cette raison, l'introduction de ce manuscrit s'articulera en 3 parties, dans lesquelles la synthèse des connaissances sur ces 3 thèmes sera présentée.

Plus précisément, le premier thème concerne la ségrégation cellulaire chez les bactéries se développant sous la forme de chaînes de coques. Compte tenu de l'importance de la paroi cellulaire dans la formation de chaînes, la structure et la biosynthèse de la paroi des bactéries à gram positif seront brièvement présentées. Ensuite les facteurs génétiques connus pour être impliqués dans la ségrégation cellulaire et les éléments en faveur de l'hypothèse d'un avantage adaptatif conféré par la croissance en chaînes seront présentés en détails.

Le second thème porte sur l'évolution rapide des régions riches en répétitions. Les mécanismes à l'origine de l'instabilité génétique des régions riches en motifs répétés seront présentés. L'avantage adaptatif conféré par les caractères réitératifs et variables des régions riches en répétitions sera détaillé.

Le troisième thème concerne la génération de nouvelles fonctions par réassortiments de domaines. Après une présentation brève de certains aspects architecturaux des protéines, suivront les données expérimentales démontrant que le réassortiment de domaines est un mécanisme de génération de nouvelles fonctions chez les bactéries. Enfin, la question de l'ampleur de ce mécanisme d'évolution, tant du point de vue quantitatif que qualitatif, sera posée.

INTRODUCTION

1 Les chaînes de coques

1.1 La paroi cellulaire des bactéries à gram positif

1.1.1 Structure de la paroi

1.1.1.1 *Introduction*

L'enveloppe cellulaire bactérienne est constituée de la membrane plasmique et de la paroi cellulaire (Figure 1). La paroi cellulaire est constituée de peptidoglycane, d'acides téichoïques, d'acides lipotéichoïques et de polysaccharides. Certaines bactéries présentent également une couche de protéines appelée couche S ou une capsule de polysaccharides (pour revue Delcour, 1999). Lors de la division cellulaire, un système protéique complexe est mis en oeuvre dans le but de néosynthétiser une nouvelle paroi cellulaire. Dans cette partie seront brièvement présentés la structure et le métabolisme de la paroi des bactéries à gram positif.

1.1.1.2 *Le peptidoglycane*

Le peptidoglycane est une macromolécule constituée de chaînes d'oses reliées les unes aux autres par des ponts peptidiques (Figure 2)(pour revue Delcour, 1999). Les chaînes d'oses sont des polymères du disaccharide N-acétyl-glucosamine- β (1-4)-N-acétyl-muramique (GlcNAc- β (1-4)-MurNAc). Dans la majorité des cas, les résidus d'acide N-acétyl-muramique sont liés, via une liaison amide, à un petit peptide de cinq acides aminés, le pentapeptide. La séquence de ce peptide varie d'une espèce à l'autre (Figure 3). Les chaînes d'oses sont reliées entre elles par la formation d'un pontage entre des pentapeptides portés par des chaînes différentes. La formation de ce pontage implique la réduction du pentapeptide en un peptide de quatre acides aminés. Le pontage peut être constitué d'une simple liaison amide reliant deux pentapeptides, comme c'est le cas chez *L. monocytogenes*. Il peut également être constitué d'un petit peptide dont la séquence varie d'une espèce à l'autre (Figure 3).

1.1.1.3 *Les polymères secondaires*

En plus du peptidoglycane, la paroi cellulaire est composée d'autres molécules appelées polymères secondaires. Les mieux connus sont les acides téichoïques et lipotéichoïques.

Les acides téichoïques sont des polymères anioniques constitués de poly-glycérol phosphate, poly-ribitol phosphate, ou poly-glucosyl phosphate (Figure 4). Ces polymères peuvent être ornés de groupements de nature glucidique ou protéique. Ils sont liés de façon covalente au peptidoglycane *via* une unité de liaison à constitution variable d'une espèce à l'autre (pour revue Navarre, 1999).

Les acides lipotéichoïques sont également des polymères constitués de poly-glycérol phosphate, poly-ribitol phosphate, ou poly-glucosyl phosphate. A la différence des acides téichoïques, ils sont ancrés de façon covalente à la membrane plasmique (Navarre et Schneewind, 1999).

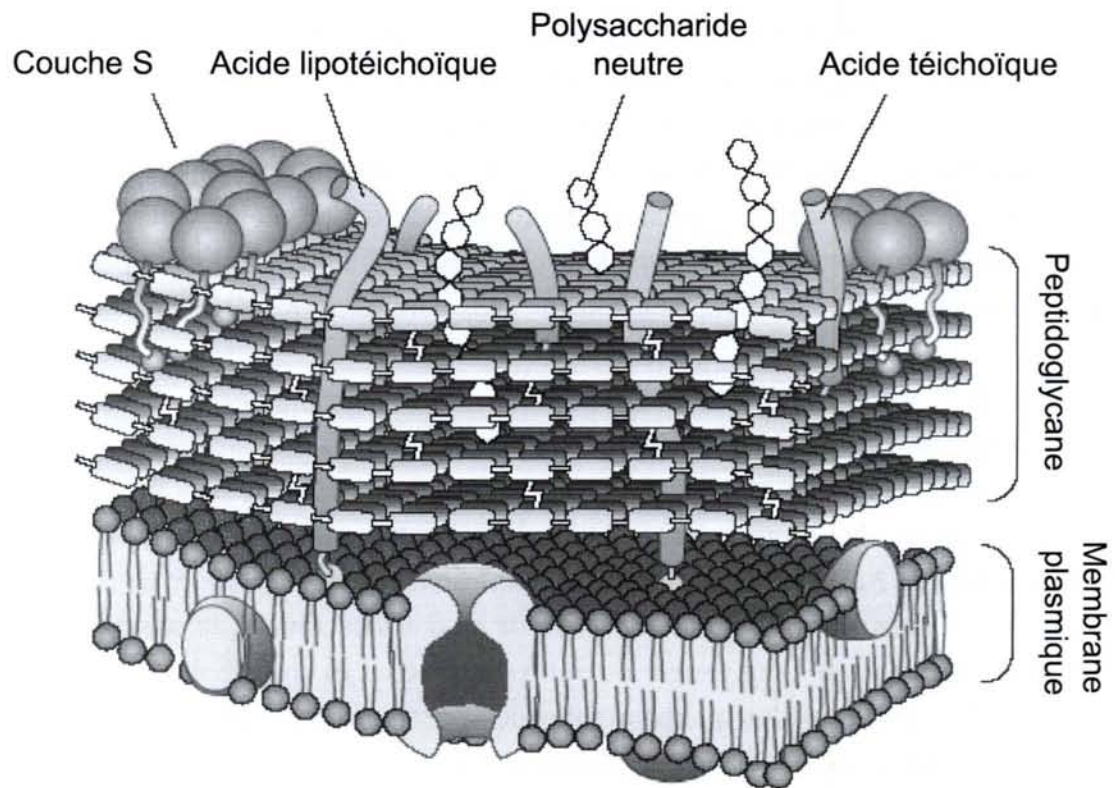


Figure 1 : Représentation schématique de l'enveloppe cellulaire des bactéries à Gram positif (d'après Delcour et al., 1999)

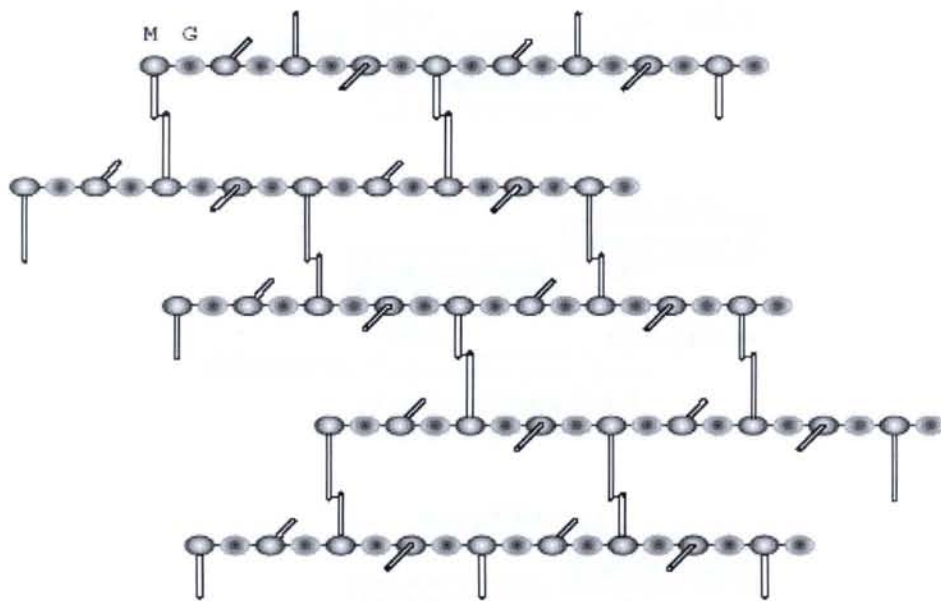


Figure 2 : Représentation schématique de la structure en réseau du peptidoglycane (d'après Delcour et al., 1999)

Les lettres M et G désignent respectivement les résidus d'acide N-acetyl-muramique et N-acetyl-glucosamine. Les pentapeptides et les peptides de pontage sont respectivement représentés par une barre et un trait fin reliant les barres entre elles.

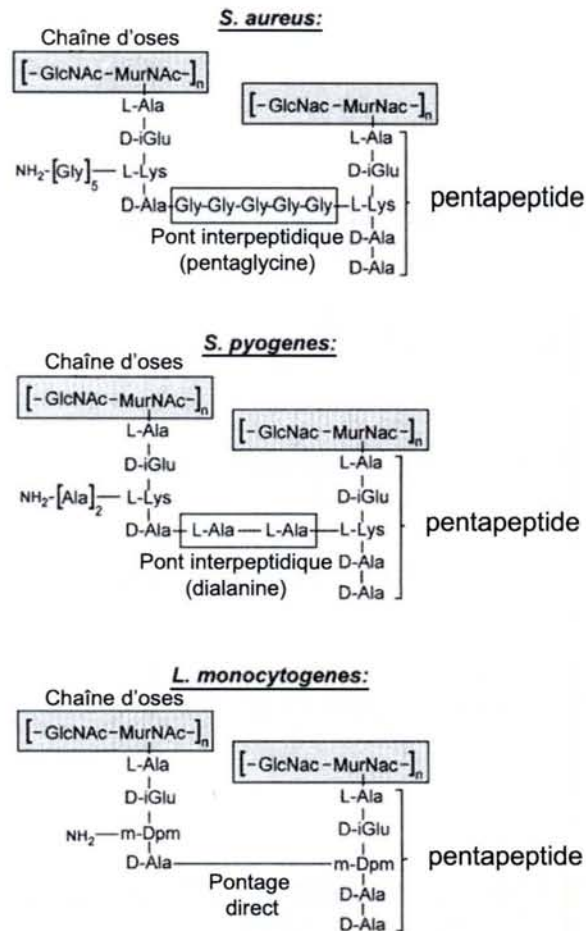


Figure 3 : Diagramme représentant la structure du peptidoglycane de *S. aureus*, *S. pyogenes* et *L. monocytogenes* (d'après Navarre et Schneewind, 1999)

D-iGlu et m-Dpm signifient respectivement D-isoglutamate et acide méso-diaminopimélique.

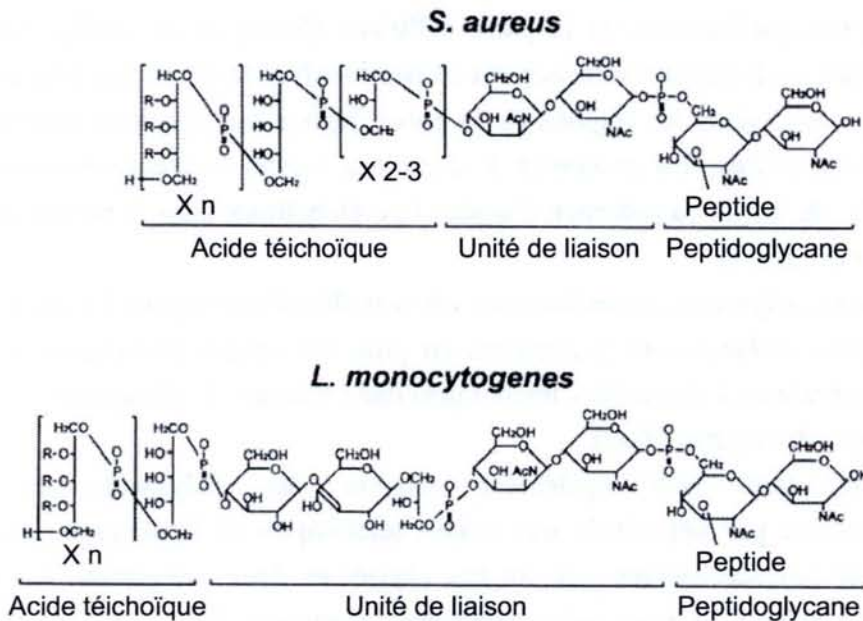


Figure 4 : Diagramme représentant la structure des acides téichoïques de *S.aureus* et *L. monocytogenes* (d'après Navarre et Schneewind, 1999)

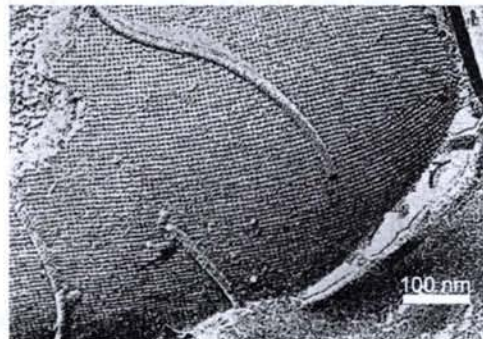


Figure 5 : Surface d'une cellule d'*Aneurinibacillus thermophilus* (d'après Kadurugamuwa et al., 1998)

La surface de la cellule a été obtenue par cryofracture, puis photographiée en microscopie électronique à balayage. Cette surface est constituée de glycoprotéines assemblées en une structure cristalline. Un flagelle, désigné par une flèche noire, est également observable.

Le rôle des acides téichoïques et lipotéichoïques est mal connu. Toutefois, il a été montré que certaines hydrolases de la paroi, telles que LytA de *Streptococcus pneumoniae*, sont liées à la paroi par interaction avec la choline des acides téichoïques (Holtje et Tomasz, 1975), suggérant un rôle dans l'attachement des protéines à la paroi. Concernant les acides lipotéichoïques, ils sont supposés être impliqués dans la localisation de protéines à domaine LysM en des régions particulières de la paroi cellulaire (Steen *et al.*, 2003). En effet, le domaine LysM serait un domaine d'attachement au peptidoglycane qui ne pourrait assurer son rôle d'ancre uniquement dans les régions de la paroi dépourvues d'acides lipotéichoïques. Ainsi, chez *S. thermophilus*, les protéines à domaines LysM seraient restreintes à une localisation septale, du fait de la présence d'acides lipotéichoïques dans le restant de la paroi cellulaire (Steen *et al.*, 2003).

La composition en polymères secondaires est très variable d'une espèce à l'autre. En effet, *B. subtilis* et *Bacillus licheniformis* présentent, en plus des acides téichoïques, des acides téichuroniques, équivalents à des acides téichoïques dans lesquels le phosphate est remplacé par l'acide uronique (Beveridge, 2000).

L'enveloppe cellulaire peut également contenir des polysaccharides, appelés polysaccharides neutres par opposition aux acides téichoïques et lipotéichoïques qui sont polyanioniques. Les polysaccharides peuvent être classés en deux catégories : les EPS pour exopolysaccharides et les CPS pour polysaccharides capsulaires. Les CPS sont attachés de façon covalente au peptidoglycane, contrairement aux EPS (Broadbent *et al.*, 2003).

1.1.1.4 La couche S

La couche S n'est présente que chez certains genres bactériens tels que *Bacillus*, *Clostridium*, *Lactobacillus* ou encore *Sporosarcina* (Beveridge, 2000). Elle est localisée à la surface du peptidoglycane, et est constituée de protéines ou glycoprotéines assemblées en une structure cristalline (Figure 5)(Kadurugamuwa *et al.*, 1998). Le rôle de la couche S n'est pas bien connu. Cependant il a été montré qu'elle est un facteur de virulence de *B. anthracis* (Mesnage *et al.*, 1997).

1.1.2 Métabolisme du peptidoglycane

1.1.2.1 Biosynthèse du peptidoglycane

Chez *S. aureus*, comme chez les autres bactéries probablement, la biosynthèse du peptidoglycane est un processus complexe se déroulant dans les trois compartiments cellulaires : le cytoplasme, la membrane plasmique et la paroi cellulaire (pour revue Navarre et Scheewind, 1999)(Figure 6). Les précurseurs, constitués d'un disaccharide relié au pentapeptide, lui-même lié au peptide de pontage, sont synthétisés dans le cytoplasme puis transportés au travers de la membrane plasmique grâce aux molécules d'undecaprenyl-pyrophosphate (C55-PP) auquel ils sont liés covalamment. Les étapes d'assemblage du peptidoglycane sont catalysées par les PBP (pour Penicillin Binding Proteins). Les PBP présentent trois types d'activité enzymatique : les activités transglycosylase, transpeptidase et carboxypeptidase.

Les PBP de classe A sont des enzymes bi-fonctionnelles permettant d'une part la polymérisation des chaînes glycaniques grâce à leur activité transglycosylase, et d'autre part

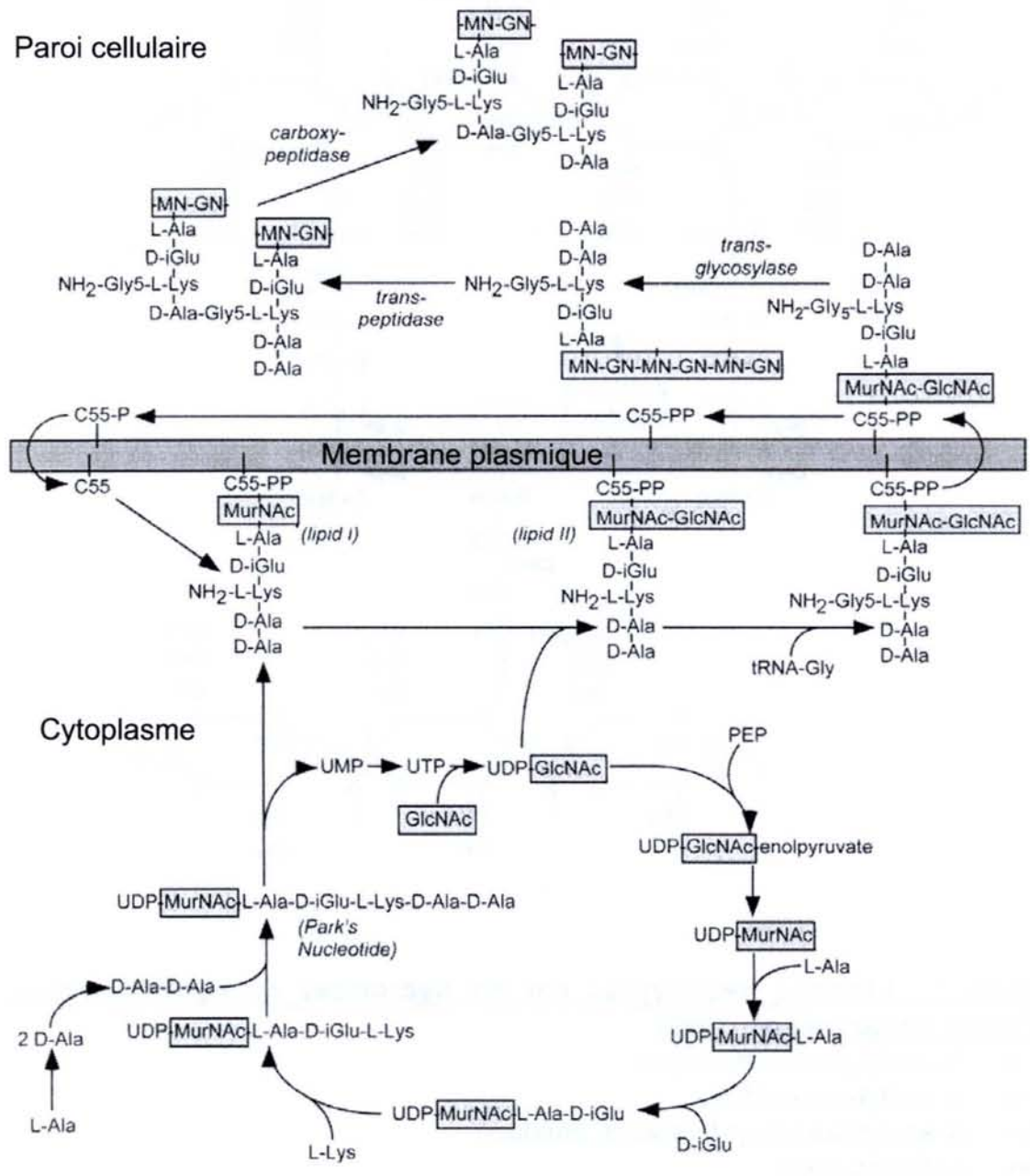


Figure 6 : Diagramme représentant les différentes étapes de biosynthèse du peptidoglycane (d'après Navarre et Schneewind, 1999)

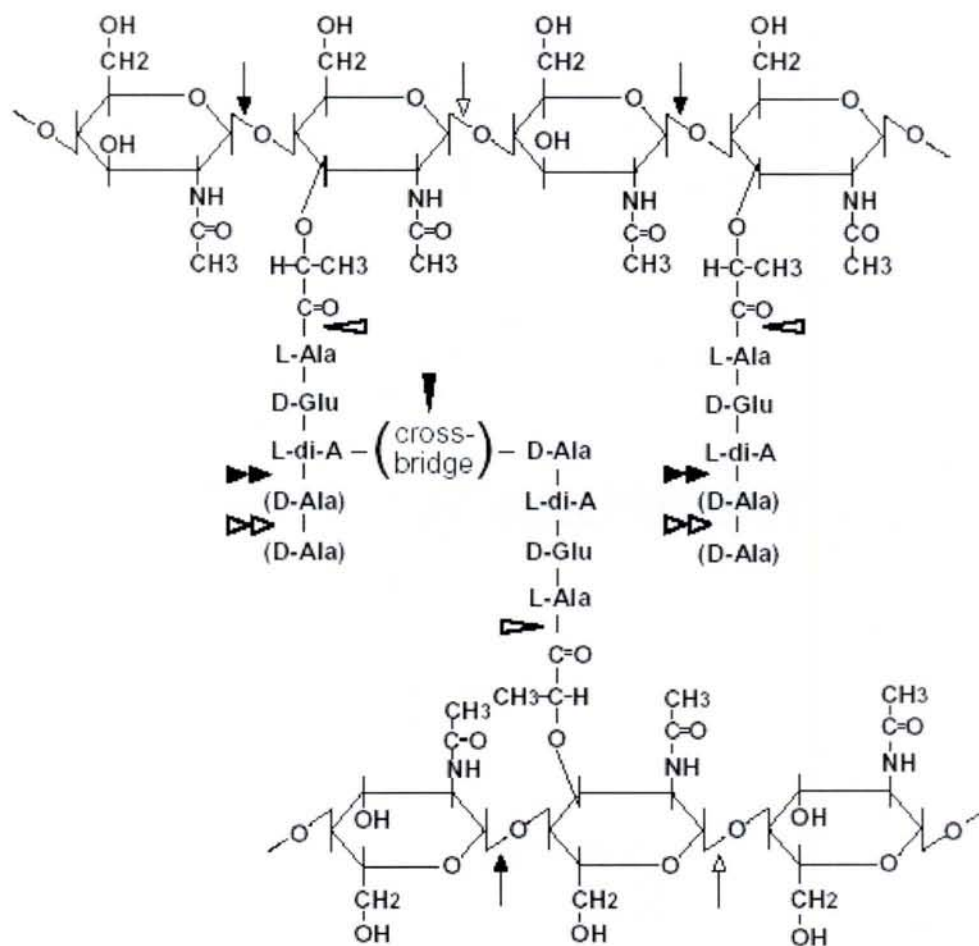


Figure 7 : Liaisons hydrolysées par les hydrolases du peptidoglycane (d'après Delcour *et al.*, 1999)

- : N-acetylglucosaminidase
- ▷ : N-acetylmuramidase
- ▷ : N-acetylmuramoyl-L-alanine amidase
- ▴ : endopeptidase
- ▷▷ : D,D-carboxypeptidase
- ▷▷ : L,D-carboxypeptidase

la formation du pontage entre les pentapeptides grâce à leur activité transpeptidase (Figure 6). Les pentapeptides non pris en charge par les transpeptidases peuvent perdre leur résidu D-Ala terminal sous l'action d'une carboxypeptidase, activité caractérisant les PBP de faible poids moléculaire (LMW PBP pour *low molecular weight* PBP). Des mutants dépourvus de la protéine PBP5 et d'au moins deux autres LMW PBP sont affectés dans la morphologie cellulaire (pour revue Popham et Young, 2003). Il est supposé que cet effet sur la morphologie cellulaire résulterait d'un équilibre entre transpeptidation et carboxypeptidation. L'état de cet équilibre aurait une influence sur la structuration tridimensionnelle du peptidoglycane (Popham et Young, 2003).

1.1.2.2 Les hydrolases du peptidoglycane

Les hydrolases du peptidoglycane interviennent à différentes étapes du métabolisme du peptidoglycane. Elles sont essentielles à l'insertion de molécules nouvellement synthétisées, puisqu'il faut hydrolyser des liaisons existantes pour pouvoir les insérer. Elles sont également responsables du démantèlement des couches externes du peptidoglycane.

Les enzymes capables d'hydrolyser le peptidoglycane peuvent être classées en quatre catégories selon le type de liaison qu'elles ciblent (Figure 7) :

- les N-acétylglucosaminidases, catalysant l'hydrolyse des liaisons osidiques reliant les résidus N-acétyl-glucosamine aux résidus d'acides N-acétyl-muramique,
- les N-acétylmuramidases, catalysant l'hydrolyse des liaisons osidiques reliant les résidus d'acides N-acétyl-muramique aux résidus N-acétyl-glucosamine,
- les N-acétylmuramoyl-L-alanines amidases, catalysant l'hydrolyse des liaisons reliant les résidus d'acides N-acétyl-muramique aux pentapeptides,
- les peptidases, catégorie regroupant les endopeptidases et les carboxypeptidases.

1.1.2.3 Dynamique de la biosynthèse du peptidoglycane

Le peptidoglycane assure de multiples fonctions dont celle de protéger la cellule de la lyse. En effet, en l'absence de peptidoglycane les cellules lyseraient sous l'effet de la pression osmotique consécutive à la différence de concentration saline entre les milieux intracellulaire et extracellulaire. Or, lors de la division cellulaire une nouvelle paroi doit être synthétisée. Et cette synthèse nécessite d'hydrolyser, et donc de fragiliser, le peptidoglycane pour pouvoir insérer de nouveaux fragments. La biosynthèse de la paroi repose sur deux principes permettant de pallier à ces contraintes : « synthétiser avant de casser » (« make before break ») et « de l'intérieur vers l'extérieur » (« inside-to-outside ») (Delcour *et al.*, 1999; Beveridge, 2000).

Le peptidoglycane peut être divisé en trois couches : la couche interne d'assemblage du nouveau peptidoglycane, la couche intermédiaire faisant barrière à la pression osmotique interne, et la couche externe, qui est constitué du peptidoglycane le plus ancien et qui est en cours d'hydrolyse par les hydrolases du peptidoglycane. Du nouveau matériel est constamment assemblé au niveau de la couche interne pendant que le peptidoglycane le plus âgé est hydrolysé dans la couche externe. Lorsque les cellules sont en division, la quantité de matériel ajouté dans la face interne est supérieure à celle retirée sur la couche externe. Ce

déséquilibre, en combinaison avec la pression osmotique interne, serait à l'origine de l'expansion du peptidoglycane (Beveridge, 2000).

1.2 Diversité de l'arrangement des cellules chez les coques

Les coques sont des bactéries à gram positif dont les cellules ont une forme sphérique ou ovoïde. Ces cellules peuvent être associées entre elles en différents types d'arrangements. Elles peuvent en effet s'organiser en grappes, en tétrades ou en chaînes (Schleifer, 1986). Les bactéries se développant sous la forme de chaînes appartiennent notamment aux genres *Streptococcus*, *Lactococcus*, *Peptostreptococcus*, *Ruminococcus* et *Coprococcus* (Schleifer, 1986). Le nombre de cellules par chaîne varie considérablement d'une espèce à l'autre : certaines espèces se développent majoritairement sous la forme de chaînes de deux cellules, on parle alors de diplocoques, comme c'est le cas de *S. pneumoniae*, alors que d'autres, comme *Streptococcus mitior*, se caractérisent par des chaînes constituées d'environ 100 cellules (Schleifer, 1986).

1.3 Les facteurs génétiques déterminant le caractère chaînes de cellules

1.3.1 La séparation des cellules résulte de l'hydrolyse de liaisons du peptidoglycane

L'observation au microscope électronique de chaînes de cellules de streptocoques de groupe A révèle l'existence d'un pont intercellulaire reliant les cellules entre elles (couverture de Fischetti *et al.*, 2000). Différents éléments laissent supposer que le lien intercellulaire des chaînes est constitué de peptidoglycane. AcmA, la muramidase-2-like, LytA et LytB sont des protéines dont l'activité de ségrégation cellulaire résulte de leur capacité à hydrolyser le peptidoglycane. En effet, AcmA de *L. lactis* ainsi que la muramidase-2-like de *E. faecalis* auraient une activité N-acétylmuramidase, et les protéines LytA et LytB de *S. pneumoniae* auraient respectivement une activité N-acétylmuramyl-L-alanine amidase et acétylglucosaminidase (Holtje et Tomasz, 1976; Buist *et al.*, 1995; Qin *et al.*, 1998; De Las Rivas *et al.*, 2002)(Figure 7). De plus, il a été montré que la protéine LytB préalablement purifiée est capable d'hydrolyser des extraits de paroi de *S. pneumoniae* (De Las Rivas *et al.*, 2002). Les protéines AcmA et LytB présentent, en plus de leur domaine catalytique, des domaines de liaison à la paroi responsables de leur localisation aux pôles cellulaires (De Las Rivas *et al.*, 2002; Steen *et al.*, 2003). Cette localisation aux pôles permettrait de rapprocher ces enzymes des liens intercellulaires dont elles doivent catalyser l'hydrolyse.

1.3.2 La croissance en chaînes résulterait d'une déficience en hydrolase du peptidoglycane

L'existence de chaînes de coques implique que les hydrolases du peptidoglycane impliquées dans le raccourcissement des chaînes ne rompent pas tous les liens intercellulaires formés. Chez *S. thermophilus*, le nombre de liens non rompus est élevé puisque les chaînes de la souche CNRZ7 peuvent contenir plus de 100 cellules. Il a été montré que les chaînes de cette souche peuvent être raccourcies jusqu'à une moyenne de deux cellules par chaîne, lorsque du lysozyme d'œuf de poule est ajouté dans le milieu de culture à une concentration

non limitante pour la croissance (Mercier *et al.*, 2000). Le même résultat est obtenu lorsque le lysozyme est remplacé par AcmA de *L. lactis* (Mercier *et al.*, 2000). Or les chaînes de *L. lactis* MG1363 sont plus courtes (moins de dix cellules par chaîne) que celles de la souche *S. thermophilus* CNRZ7. Ainsi, ce résultat suggère que les chaînes de *S. thermophilus* sont plus longues que celles de *L. lactis* car *S. thermophilus* est davantage déficient en hydrolase du peptidoglycane.

Il est possible d'obtenir des mutants se développant sous la forme de chaînes chez *Escherichia coli*, *Listeria monocytogenes*, *Bacillus subtilis* et *Enterococcus faecalis* dont la souche sauvage se développe sous la forme de cellules isolées (Ishikawa *et al.*, 1998; Qin *et al.*, 1998; Margot *et al.*, 1999; Ohnishi *et al.*, 1999; Heidrich *et al.*, 2001; Carroll *et al.*, 2003). Dans chacun des cas, la croissance en chaînes de cellules peut résulter de l'inactivation d'un gène codant une hydrolase du peptidoglycane. Ces résultats tendent à confirmer que la croissance en chaînes chez les coques résulterait d'une déficience en hydrolases du peptidoglycane impliquées dans la ségrégation cellulaire.

1.3.3 Les gènes impliqués dans la formation des chaînes

Deux gènes intervenant dans la formation des chaînes ont été identifiés. L'inactivation du gène *ponA* de *L. lactis* dans un contexte génétique $\Delta acmA$, mutant présentant de longues chaînes, entraîne un raccourcissement des chaînes. Le gène *ponA* code la protéine PBP1a, enzyme bi-fonctionnelle impliquée dans la synthèse du peptidoglycane chez *E. coli* (pour revue Goffin, 1998). Plus précisément, elle présente une activité transglycosylase ainsi qu'une activité transpeptidase (Figure 6). Cela suggère que le lien reliant les cellules entre elles est constitué d'au moins une des deux liaisons covalentes dont la formation est catalysée par PBP1a.

L'inactivation de *gls24* chez *E. faecalis* confère une diminution de la longueur des chaînes indiquant que *gls24* est impliqué dans la formation des chaînes. Ce phénotype est accompagné d'une diminution de la vitesse de croissance et d'une diminution de la résistance aux sels biliaires (Giard *et al.*, 2000). De plus, il été montré que l'inactivation de *gls24* est associée à l'augmentation de la quantité de trois enzymes du métabolisme du pyruvate : la L-lactate déshydrogénase, la lipoamide déshydrogénase et la pyruvate décarboxylase (Giard *et al.*, 2000). Cela suggère qu'il existe un lien entre le métabolisme du pyruvate et la formation des chaînes de cellules.

1.4 Lien entre la division cellulaire et la croissance en chaînes

Les mutants *gls24*, *acmA* et *lytB* sont affectés dans la longueur de leurs chaînes mais ne présentent pas de modification de la morphologie des cellules (Buist *et al.*, 1995; De Las Rivas *et al.*, 2002). En revanche les mutants *rodA*, *pbp2b* et *mreD* de *S. thermophilus* (Thibessard *et al.*, 2002b; Thibessard *et al.*, 2004), et *pcsB* de *S. pneumoniae* (Ng *et al.*, 2004) sont affectés dans la morphologie cellulaire, en plus de présenter de longues chaînes.

Les cellules des mutants *rodA*, *pbp2b* et *mreD* de *S. thermophilus* ont une forme sphérique, alors que la souche sauvage présente des cellules de forme ovoïde (Figure 8). Les protéines RodA et PBP2b seraient impliquées dans l'élongation du peptidoglycane (Ishino *et al.*, 1986).

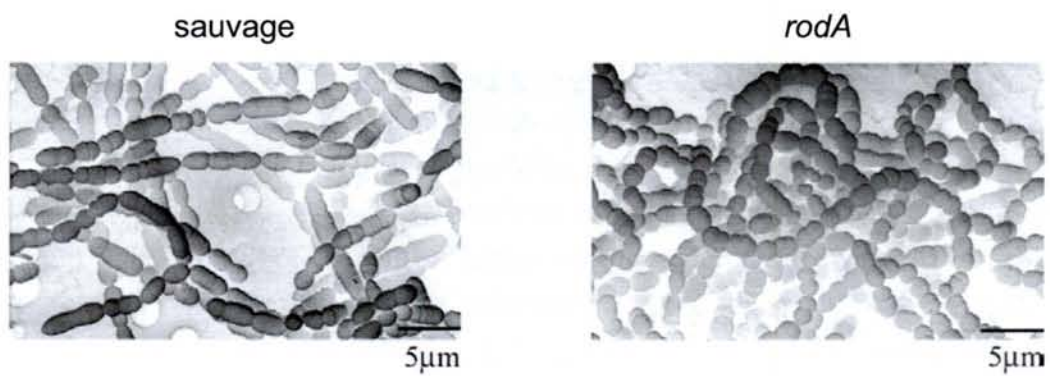


Figure 8 : Implication de RodA dans la morphologie cellulaire chez *S. thermophilus* (d'après Thibessard et al., 2002)

Les cellules ont été photographiées en microscopie électronique à balayage.

Le rôle de ces protéines dans la division cellulaire suggère que la formation de chaînes de cellules est dépendante du processus de division cellulaire. Cette idée serait renforcée par les travaux effectués sur la protéine PcsB de *S. pneumoniae* (Ng *et al.*, 2004). Le gène *pcsB* de *S. pneumoniae*, étant essentiel à sa croissance, sa fonction a été étudiée via la construction d'un mutant conditionnel pour lequel le niveau d'expression du gène est fonction de la concentration du milieu en fucose. La diminution de l'expression de ce gène entraîne la croissance de *S. pneumoniae* sous la forme de chaînes, la souche sauvage étant un diplocoque. Une diminution drastique de l'expression de ce même gène entraîne, en plus de la croissance sous la forme de longues chaînes, l'apparition de cellules à morphologie aberrante ainsi que la perte du parallélisme des plans équatoriaux et septaux. Par ailleurs, il a été montré que la PcsB est impliquée dans la biosynthèse de la paroi (Ng *et al.*, 2004).

L'ensemble de ces résultats tend à montrer que la formation des chaînes de cellules est intimement liée au processus de division cellulaire.

1.5 Avantage adaptatif conféré par la croissance en chaînes

1.5.1 La longueur des chaînes est régulée en fonction de l'environnement

Un des mécanismes d'adaptation des bactéries à des modifications des paramètres de leur environnement implique trois étapes : la perception de ces modifications, la transduction du signal perçu et la réponse pouvant se manifester par la modification de l'expression de gènes. Certains acteurs assurant certaines de ces étapes sont connus dans le cas où la réponse de la bactérie inclut une modification de la longueur des chaînes.

1.5.2 Paramètres de l'environnement influençant la longueur des chaînes

Deux facteurs de l'environnement influençant la longueur des chaînes ont été identifiés. Chez *S. sobrinus* la longueur des chaînes varie en fonction du pH (Thibodeau et Ford, 1991). A pH alcalin les chaînes sont courtes tandis qu'à pH acide les chaînes sont longues. Chez *S. gordonii*, la longueur des chaînes augmente lorsque les cellules sont en interaction avec des fibrilles de collagène de type I (Heddle *et al.*, 2003).

1.5.3 Acteurs transduisant le signal

Les protéines Stk1 et Stp1 de *S. agalactiae* sont respectivement une serine/threonine kinase et une phosphatase potentiellement impliquées dans la transduction de signaux (Rajagopal *et al.*, 2003). L'inactivation des gènes codant ces protéines entraîne une augmentation de la longueur des chaînes de cellules de cette bactérie. La nature du signal transduit et les gènes régulés par Stk1 et Stp1 ne sont pas connus (Rajagopal *et al.*, 2003).

1.5.4 Gènes régulés

Les protéases HtrA et GelE respectivement de *L. lactis* et *E. faecalis* ont toutes deux une influence sur la longueur des chaînes. En effet, l'inactivation des gènes codant ces protéases entraîne une augmentation de la longueur des chaînes de cellules (Foucaud-Scheunemann et Poquet, 2003; Waters *et al.*, 2003). L'influence de HtrA sur la longueur des chaînes résulterait de l'activité protéolytique qu'elle exercerait sur la protéine AcmA, protéine de ségrégation

cellulaire (Poquet *et al.*, 2000). Concernant GelE, les auteurs font l'hypothèse qu'elle activerait, par protéolyse, une autolysine responsable de la ségrégation cellulaire (Waters *et al.*, 2003). Or l'expression du gène codant HtrA augmente en condition de stress et celle du gène codant GelE est induite par le système de quorum-sensing codé par *fsr* (Qin *et al.*, 2000; 2001). Il est donc vraisemblable qu'en condition de stress ou en fonction de la densité cellulaire, la longueur des chaînes respectivement de *L. lactis* ou *E. faecalis* diminue, et que cette diminution soit le résultat de l'augmentation de l'expression des gènes codant HtrA et GelE.

Ainsi, la longueur des chaînes de cellules est un caractère morphologique qui est modulé suite à des modifications des paramètres de l'environnement. Cela suggère que la croissance sous la forme de chaînes de cellules peut conférer un avantage adaptatif.

1.5.5 La capacité à adhérer à des surfaces abiotiques est modulée en fonction de la longueur des chaînes de cellules

La capacité à adhérer à des surfaces abiotiques telles que le verre, le polystyrène ou l'acier, a été estimée pour la souche sauvage de *L. lactis* et pour un mutant $\Delta acmA$ présentant de longues chaînes de cellules. Le mutant $\Delta acmA$ présente une capacité réduite à adhérer à ces surfaces par rapport à la souche sauvage (Mercier *et al.*, 2002). L'adhésion peut être restaurée de deux façons différentes : lorsque les chaînes de cellules sont préalablement traitées avec du lysozyme, traitement ayant pour conséquence de réduire la longueur des chaînes, ou alors par inactivation du gène *ponA* dans le contexte génétique $\Delta acmA$ (Mercier *et al.*, 2002). Ainsi, la capacité à adhérer à des surfaces abiotiques diminue lorsque la longueur des chaînes augmente.

L'hypothèse défendue par les auteurs pour expliquer ce résultat est une diminution de la probabilité de contact d'une cellule avec la surface abiotique lorsque le nombre de cellules par chaîne augmente. Cette probabilité diminuerait du fait d'une réduction du nombre moyen de degré de liberté d'une cellule lorsque le nombre de cellules par chaîne augmente (Mercier *et al.*, 2002).

La longueur des chaînes ayant une influence sur la capacité à adhérer à une surface, un avantage adaptatif de la régulation de la longueur des chaînes de cellules pourrait être de moduler le pouvoir adhésif en fonction des conditions environnementales.

2 Variabilité génétique de régions intragéniques riches en motifs répétés

2.1 Abondance des répétitions dans les génomes bactériens

La présence de séquences répétées est un caractère ubiquitaire des génomes bactériens. Certaines répétitions sont très abondantes comme par exemple le motif HRM (pour Highly Repetitive Motif) de 13 pb qui est représenté environ 1000 fois dans le génome de 2,4 Mpb de *Lactococcus lactis* (Mrazek *et al.*, 2002). Toutefois, le contenu en répétitions peut varier considérablement d'une espèce bactérienne à l'autre (Romero *et al.*, 1999). Romero *et al.* (1999) ont recherché dans chacun des génomes bactériens, tous les motifs parfaitement répétés et d'une taille supérieure à 300 pb.

Les résultats montrent que les Chlamydiales telles que *Chlamydia pneumoniae* et *Chlamydia trachomatis*, la protéobactérie- α *Rickettsia prowasekii*, et les Spirochetes tels que *Treponema pallidum* et *Borrelia burgdorferi* présentent peu de répétitions de cette taille dans leur génome (Figure 9). D'autres génomes au contraire présentent une forte proportion de motifs répétés tels que la cyanobactérie *Synechocystis sp.* et la bactérie Firmicutes *Bacillus subtilis*. Ces différences ne sont pas corrélées à la taille du génome. En effet, *Mycobacterium pneumoniae* et *Borrelia burgdorferi* ont un génome de taille semblable (respectivement 0,81 et 0,91 Mb) mais ne présentent pas la même quantité de répétitions (Figure 9).

En plus de varier quantitativement, les répétitions varient qualitativement. Tandis que le génome d'*E. coli* est particulièrement riche en unités répétées longues (de taille comprise entre 85 et 200 pb), le génome de *Haemophilus influenzae* est riche en motifs répétés courts (3 à 4 pb) (Yeramian et Buc, 1999).

2.2 Taille et organisation des répétitions, notion de complexité

Les motifs d'ADN répétés peuvent avoir une taille mesurant entre 1 nucléotide (pour revue Henderson *et al.*, 1999 ; Hallet *et al.*, 2001) et plusieurs méga paires de bases (Wenner *et al.*, 2003). Les motifs peuvent être répétés de façon directe ou inversée, et être positionnés en tandem ou de façon dispersée (Figure 10). Si des motifs répétés sont localisés à l'intérieur de gènes, et qu'ils sont positionnés dans la même phase de lecture les uns par rapport aux autres, alors la séquence des protéines codées par ces gènes contiendra des motifs protéiques répétés.

Dans le cas des répétitions en tandem, trois grandes classes peuvent être distinguées selon la taille du motif : les régions constituées d'unités répétées mesurant plus de 100 pb sont appelées satellites, celles dont les unités répétées ont une taille allant de 6 à 100 pb sont des minisatellites, et celles dont les unités répétées mesurent de 1 à 5 pb sont des microsatellites (Charlesworth *et al.*, 1994).

Au sein d'une région riche en répétitions, les unités répétées présentes aux extrémités ont tendance à diverger davantage des motifs répétés internes, que ces derniers ne divergent les uns des autres (Kehoe, 1994). Il semble que cette caractéristique soit indépendante de la longueur des unités répétées. En effet, le gène de *Staphylococcus aureus* codant la protéine

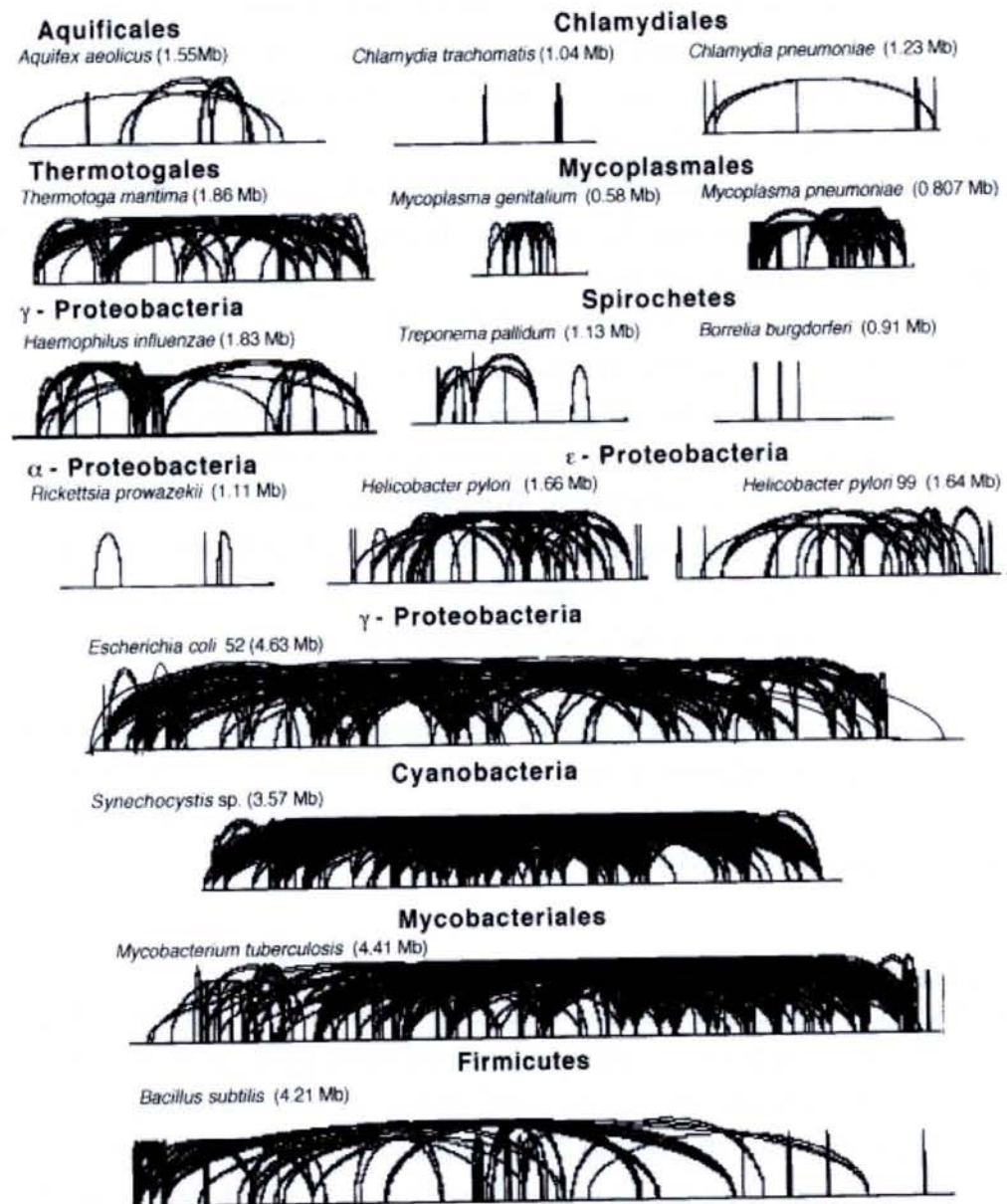


Figure 9 : Contenu en répétitions d'ADN dans quelques génomes bactériens (d'après Romero et al., 1999)

Les auteurs ont recherché les motifs répétés parfaitement et de taille supérieure à 300 pb au sein des génomes bactériens. Les chromosomes sont représentés par un trait fin horizontal. Les motifs répétés sont reliés les uns aux autres par un arc de cercle.

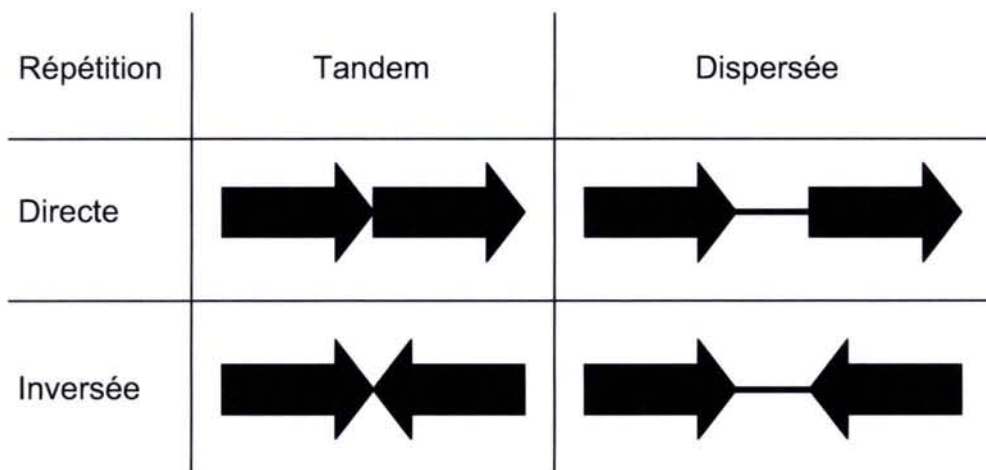


Figure 10 : Tableau présentant la variété de disposition des motifs répétés les uns par rapport aux autres

Les flèches représentent les unités répétées, les traits noirs épais les séquences d'ADN non répétées.

ClfB présente une régions riche en répétitions constituée d'un motif répété court de 18 pb (Ni Eidhin *et al.*, 1998). En revanche, le gène de streptocoques de groupe B codant la protéine alpha C présente lui aussi une telle région mais constituée quant à elle d'un motif répété de 246 pb (Michel *et al.*, 1992).

Lorsque les motifs sont de petite taille (de l'ordre de la dizaine de nucléotides ou d'acides aminés), et qu'ils sont proches les uns des autres sur la molécule d'ADN ou sur la structure primaire d'une protéine, ceux-ci peuvent induire un biais dans la composition en nucléotides ou acides aminés de la séquence. Ce biais rend la séquence plus « simple » ou moins complexe, c'est pourquoi ces régions sont qualifiées de régions de faible complexité (Wootton et Federhen, 1996). C'est le cas du gène de *S. aureus* codant la protéine ClfB citée précédemment qui présente une région de 768 pb quasi exclusivement constituée du motif GAY TCN GAY TCN GAY AGY répété en tandem direct (Ni Eidhin *et al.*, 1998). Ce motif étant traduit en la séquence protéique DSDSDS, la protéine ClfB présente une région quasi exclusivement constituée de la répétition en tandem du motif DS sur 256 résidus d'acides aminés. Lorsque la séquence de ClfB est analysée par le logiciel SEG (Wootton et Federhen, 1996), logiciel permettant de détecter une séquence de faible complexité, celui-ci reconnaît cette région riche en répétitions comme une séquence de faible complexité (Figure 11).

2.3 Importante variabilité des régions riches en motifs répétés

Les régions riches en motifs répétés sont caractérisées par une forte instabilité. Les génomes de *Yersinia pestis* et de *Bacillus anthracis* contiennent respectivement 76 et 30 régions minisatellites dont l'unité répétée mesure plus de 9 pb (Le Fleche *et al.*, 2001). L'analyse de 5 souches de *Y. pestis* révèle un polymorphisme de taille de 45 des 76 régions minisatellites. De façon analogue, l'analyse de 31 souches de *B. anthracis* montre un polymorphisme de 14 des 30 régions minisatellites. Ainsi, environ la moitié des régions riches en motifs répétés est polymorphe à l'intérieur d'une espèce.

Le polymorphisme de taille des régions minisatellites résulte de la modification du nombre de motifs répétés (pour revue Van Belkum *et al.*, 1998)(Yang et Gabriel, 1995; Wilton *et al.*, 1998; Lukomski *et al.*, 2000; Schupp *et al.*, 2000; Rasmussen et Bjorck, 2001; Areschoug *et al.*, 2002; Koroleva *et al.*, 2002; Karatzas *et al.*, 2003; de Benito *et al.*, 2004). Un exemple est le gène *fbsA* de *Streptococcus agalactiae* qui présente en son centre une région minisatellite constituée de la répétition en tandem d'un motif de 48 pb (Schubert *et al.*, 2002). Cette région arbore un polymorphisme de taille important, au sein de l'espèce, résultant de la modification du nombre de motifs répétés, nombre pouvant varier de 3 à 30, il en résulte que la taille de la protéine FbsA varie de 186 aa à 618 aa (Schubert *et al.*, 2002). La variabilité du nombre de motifs répétés au sein de ces régions leur vaut leur qualificatif de VNTR (pour variable number of tandem repeat).

Ce type de variabilité n'est pas restreint aux régions minisatellites, mais est également observé pour les régions microsatellites (pour revue Henderson *et al.*, 1999)(Snyder *et al.*, 2001), ainsi que pour les régions satellites (Mouw *et al.*, 1988; Madoff *et al.*, 1991; Smith *et al.*, 1993; Zhang et Wise, 1996; Brady *et al.*, 1998).

Ces régions peuvent évoluer rapidement. Par exemple, il a été montré qu'au sein des régions constituées de la répétition en tandem du motif CA, un événement d'insertion ou de délétion du motif CA a lieu toutes les 1000 réplifications chez *Escherichia coli* (Levinson et Gutman, 1987).

L'instabilité des régions riches en répétitions est d'ailleurs utilisée pour différencier des souches d'une même espèce très proches les unes des autres (pour revue Van Belkum, 1999).

2.4 L'instabilité des motifs répétés de façon dispersée peut engendrer la création de nouveaux motifs répétés

Au sein des gènes et de leur produit, les répétitions décrites sont majoritairement des répétitions directes positionnées en tandem ou de façon dispersée. La protéine alpha C de la souche A909, un streptocoque de groupe B, présente 9 unités d'un motif de 82 résidus d'acides aminés répété en tandem direct (Michel *et al.*, 1992)(Figure 12). A l'inverse, la protéine G de la souche GX7809, un streptocoque de groupe G, présente deux types de motifs répétés de façon dispersée. Les motifs A, de 37 résidus d'acides aminés et B, de 55 résidus d'acides aminés, sont chacun présents en deux exemplaires (Fahnestock *et al.*, 1986)(Figure 12).

La séquence de ces deux protéines est variable. La variabilité de la séquence de la protéine alpha C, dont les motifs sont répétés en tandem, se traduit par une modification du nombre de motifs répétés (Madoff *et al.*, 1991; Madoff *et al.*, 1996). Celle de la protéine G, dont les motifs sont répétés de façon dispersée, se traduit non seulement par une modification du nombre d'unités répétées d'un même type de motifs, mais également par une modification de la composition en motifs répétés (Figure 13). En effet, la protéine G de la souche GX7809 présente 2 types de motifs répétés différents, alors que celle de la souche GX7805 en présente 4 (Fahnestock *et al.*, 1986; Filpula *et al.*, 1987). D'un point de vue mécanistique, ces différences s'interprètent par des événements de recombinaison entre unités répétées d'un même type de motifs, résultant en des événements d'insertion ou de délétion. Par exemple, un événement de recombinaison entre les motifs A1 et A2 du gène codant la protéine G de la souche GX7809 peut résulter non seulement en la duplication du motif A, mais également en la duplication de la région située entre les motifs A1 et A2. Cet événement entraînerait la création d'un nouveau motif répété : le motif B de la séquence de la protéine G de la souche GX7805 (Fahnestock *et al.*, 1986; Filpula *et al.*, 1987; Kehoe, 1994).

2.5 Mécanismes de recombinaison à l'origine de l'évolution des régions riches en motifs répétés de façon directe

2.5.1 La recombinaison : définition

La recombinaison a été définie par Clark (1971) comme la création de nouvelles combinaisons de séquences nucléiques portées par une ou plusieurs molécules d'acides nucléiques. Les mécanismes de recombinaison impliqués dans l'instabilité des régions riches en motifs répétés sont la recombinaison illégitime et la recombinaison homologue.

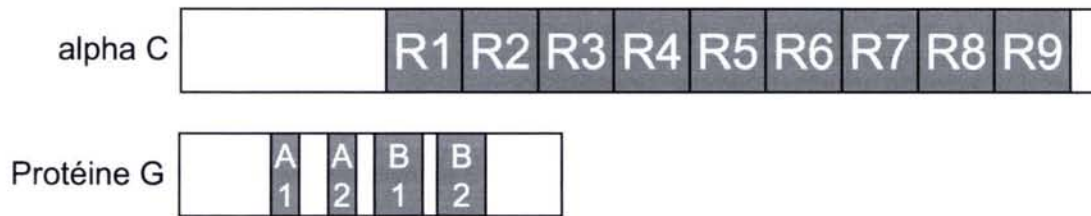


Figure 12 : Représentation schématique de la protéine alpha C et de la protéine G (d'après Fahnestock *et al.*, 1986 ; Michel *et al.*, 1992)

Les motifs répétés sont représentés sous la forme d'un rectangle gris contenant le nom du motif répété. Le type de motif est indiqué par une lettre et l'exemplaire du motif répété par un numéro.

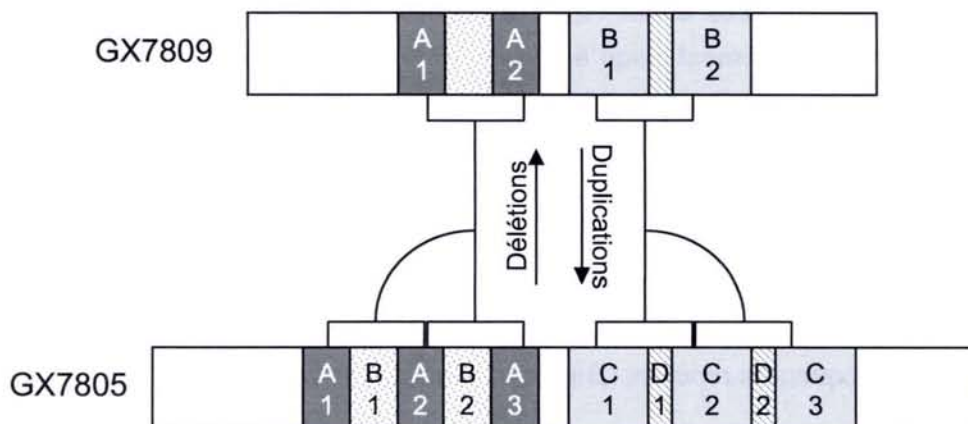


Figure 13 : Séquences répétées dans la protéine G (d'après Kehoe, 1984)

Représentation schématique de la structure primaire de la protéine G de deux souches de streptocoques de groupe G. Chaque motif répété est nommé d'une lettre, désignant le type de motif répété, suivie d'un chiffre, désignant la copie de l'unité répétée. La dénomination de chaque type de motif répété est effectuée par analyse de la séquence de l'extrémité N-terminale vers l'extrémité C-terminale. A chaque nouveau type de motif répété rencontré, une lettre lui est associée. Ainsi, le premier type de motif rencontré sera nommé A, le second B, et ainsi de suite.

2.5.2 La recombinaison illégitime

2.5.2.1 Définition

La recombinaison illégitime désigne la formation de nouvelles molécules d'ADN par jonction de séquences d'ADN présentant ou pas de courtes homologies (pour revue Michel, 1999). Elle s'oppose ainsi à la recombinaison homologue faisant intervenir RecA qui nécessite de plus grandes homologies entre les molécules recombinées. La recombinaison illégitime peut induire différents types de réarrangements tels que des délétions, des inversions, des amplifications et des translocations.

2.5.2.2 Facteurs influençant la fréquence de recombinaison illégitime

La fréquence d'événements de recombinaison illégitime dépend de la longueur de l'homologie entre les séquences lorsque celle-ci existe. Des études effectuées chez un mutant *recA* d'*E. coli* ont montré que des unités répétées en tandem direct dont la taille augmente de 14 pb à 100 pb recombinent à une fréquence augmentant de 10^{-5} à 10^{-4} . Toutefois, lorsque la taille des unités répétées augmente au-delà de 100 pb, la fréquence reste constante (Bi et Liu, 1994).

En plus de la longueur de l'homologie, la fréquence de recombinaison est fortement dépendante de la distance entre les séquences répétées. Chez *B. subtilis*, la fréquence de délétion par recombinaison entre deux unités de 18 pb répétées de façon directe, est relativement élevée lorsque les deux motifs sont espacés de 51 pb puisqu'elle est de $5 \cdot 10^{-6}$. Et, lorsque la distance entre les deux unités répétées augmente de 33 à 2313 pb, la fréquence de délétion diminue exponentiellement jusqu'à plus de 1000 fois (Chedin *et al.*, 1994).

D'autres facteurs favorisent également la recombinaison illégitime, notamment la présence de séquences palindromiques entre les unités répétées (Glickman et Ripley, 1984). Les séquences palindromiques ont l'aptitude de former des structures secondaires stables. La formation de ces structures a deux conséquences : la première est qu'elles permettent de rapprocher spatialement les unités répétées l'une de l'autre, et la seconde est qu'elles induisent une pause de la fourche de réplication. Ces pauses favorisent les événements de recombinaison entre séquences répétées (Bierne et Michel, 1994).

2.5.2.3 Modèle moléculaire de recombinaison illégitime entre séquences répétées

Deux modèles non exclusifs ont été proposés pour expliquer les événements de recombinaison entre courtes séquences homologues. Le premier modèle implique un glissement de fourche de réplication (« slipped-mispairing model », « replication slippage » ou encore « copy choice recombination »). Ce modèle prévoit une pause de la machinerie de réplication au niveau d'un exemplaire d'un motif répété, conduisant à la dissociation du brin d'ADN néo-synthétisé de la matrice, et sa réassociation avec un autre exemplaire du motif répété (Figure 14) (Michel, 1999). Ce modèle permet aussi bien d'expliquer les événements de délétion que les événements de duplication de motifs répétés.

Le second modèle se nomme « single-strand annealing » (SSA), et implique en premier lieu un événement de cassure double brin de l'ADN. La dégradation d'un brin d'ADN, à partir des extrémités double brin par les exonucléases, permet l'appariement des séquences

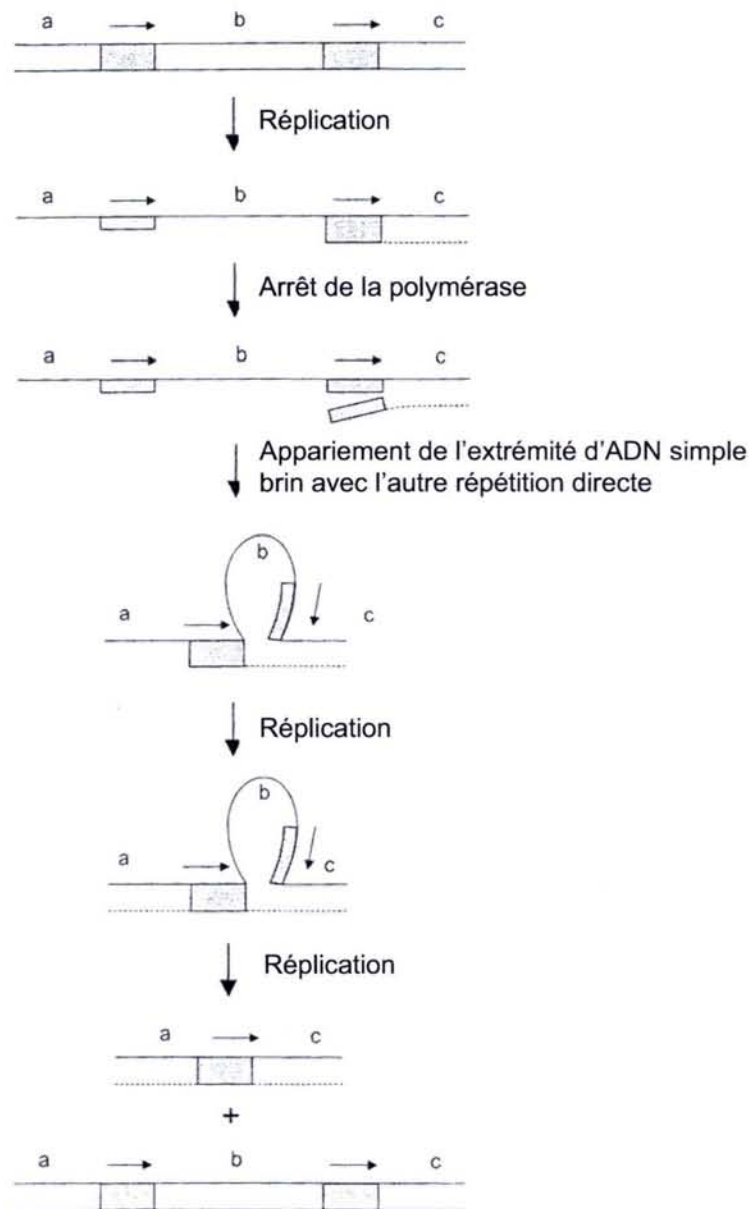


Figure 14 : Modèle de délétion par glissement de fourche de réplication entre des motifs répétés de façon directe (d'après Michel, 1999)

Les motifs répétés sont représentés par des rectangles et leur orientation est indiquée par une flèche. Les traits représentent les brins d'ADN, les pointillés les brins d'ADN nouvellement synthétisés. Les régions a et c flanquent la région contenant les motifs répétés, la région b est comprise entre les unités répétées. Une pause de la fourche de réplication au niveau de la première unité répétée rencontre induit une séparation des deux brins d'ADN. L'appariement de l'unité répétée néo-synthétisée avec l'autre unité répétée entraîne la délétion de la région b ainsi que de l'une de deux unités répétées.

complémentaires correspondant aux deux unités répétées (Figure 15) (Michel, 1999). Ce mécanisme conduit à la délétion de motifs répétés.

2.5.3 La recombinaison homologue

2.5.3.1 Longueur minimale requise des régions homologues

La recombinaison homologue requiert des molécules d'ADN présentant une homologie de séquence sur une longueur suffisante. Cette longueur doit être d'environ 20 et 30 pb respectivement chez *E. coli* et *B. subtilis* (Watt *et al.*, 1985; Majewski et Cohan, 1999). Chez *E. coli*, elle peut dépendre de la protéine RecA mais également d'autres systèmes Rec (pour revues Kowalczykowski *et al.*, 1994 ; Kuzminov, 1999).

2.5.3.2 Fréquence

Chez *E. coli*, la fréquence des événements de recombinaison homologue augmente de façon exponentielle, de 10^{-7} à 10^{-4} , lorsque la longueur des séquences homologues passe de 20 à 74 pb, puis augmente de façon linéaire au-delà de 74 pb (Watt *et al.*, 1985). Contrairement à la recombinaison illégitime, la recombinaison homologue RecA-dépendante est peu sensible à l'éloignement des séquences homologues (Bi et Liu, 1994).

2.5.3.3 Crossing over et conversion génique

La recombinaison homologue peut aboutir à deux types d'événements : les échanges réciproques (on parle de crossing-over), et les échanges non réciproques (ou conversion génique)(pour revue Petes et Hill, 1988 ; Kowalczykowski *et al.*, 1994). Concernant les échanges non réciproques, la longueur des séquences converties peut atteindre plusieurs centaines de paires de bases (Bianchi et Radding, 1983).

Ainsi la recombinaison homologue entre deux copies ectopiques d'une répétition imparfaite pourra aboutir, par conversion génique, à la conversion de la séquence d'une des deux unités répétées, en la séquence de l'autre unité répétée. Il est à noter que les mêmes molécules recombinées pourront également être générées par la délétion d'une des deux unités répétées, et la duplication de l'autre unité répétée.

2.5.4 Impact de la recombinaison sur l'évolution des régions riches en motifs répétés

Les recombinaisons homologues et illégitimes permettent toutes deux d'expliquer les événements d'insertion ou de délétion de l'intégralité d'unités répétées, événements caractérisant l'instabilité des régions riches en répétitions. En effet, la recombinaison homologue requiert un appariement de séquences homologues. La recombinaison illégitime, quant à elle ne nécessite pas d'homologie pour avoir lieu, mais elle est beaucoup plus fréquente lorsque les séquences recombinées peuvent s'apparier. Que ce soit par recombinaison homologue ou par recombinaison illégitime, les événements d'insertion ou de délétion de l'intégralité d'unités répétées ont lieu par recombinaison ectopique, c'est-à-dire par recombinaison entre copies non alléliques (Figure 16). Les événements de recombinaison ectopiques impliquant des motifs répétés de façon directe ont pour conséquence la délétion de la région située entre les deux points de recombinaison (Figure 17 A). Les événements de

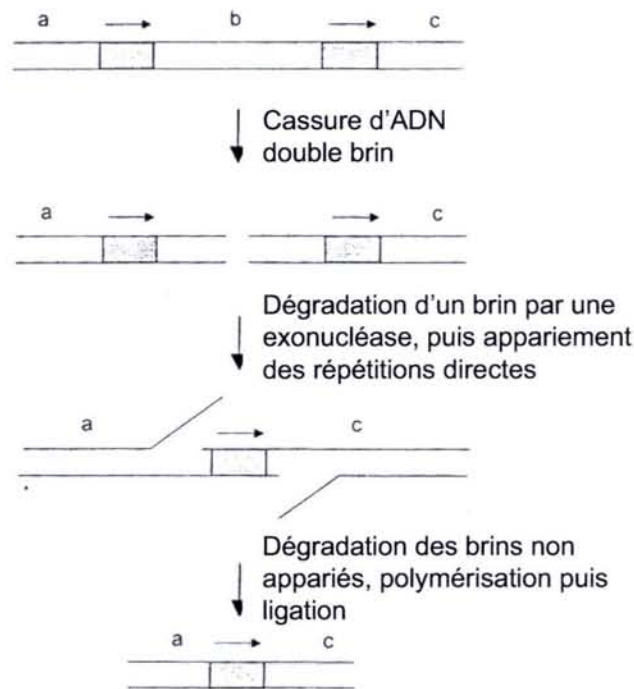


Figure 15 : Modèle de délétion « single-strand annealing » (SSA) (d'après Michel, 1999)

Les motifs répétés sont représentés par des rectangles et leur orientation est indiquée par une flèche. Les traits représentent les brins d'ADN. Les régions a et c flanquent la région contenant les motifs répétés, la région b est comprise entre les unités répétées. L'initiation a lieu par une cassure double brin de l'ADN, dans la région située entre les deux unités répétées. La dégradation d'un des brins d'ADN par les exonucléases libère les unités répétées sous la forme d'ADN simple brin, ces unités répétées peuvent ainsi s'apparier. Les extrémités d'ADN simple brin non appariées sont dégradées par des exonucléases, l'ADN est rendu entièrement double brin par polymérisation, puis les extrémités sont jointes par ligation. Ce mécanisme entraîne la délétion de la région b et de l'une des deux unités répétées.

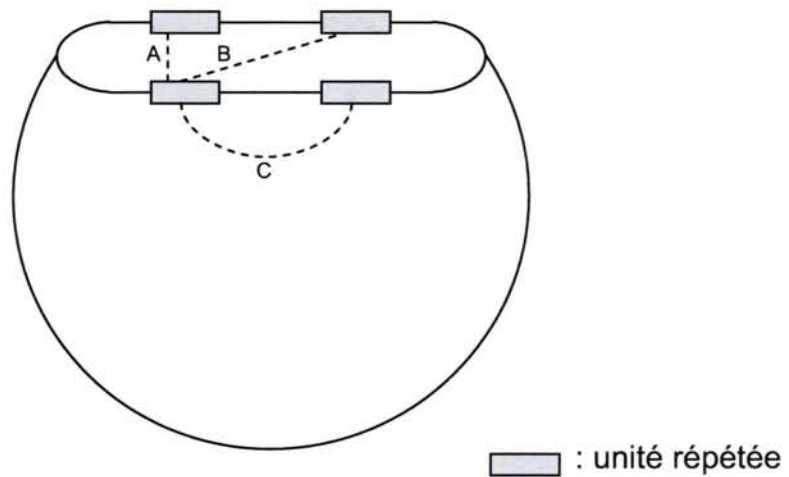


Figure 16 : Recombinaison allélique et recombinaison ectopique entre séquences répétées

Le chromosome bactérien en cours de réplication est représenté par un trait fin. Le trait pointillé A symbolise un événement de recombinaison allélique (c'est-à-dire entre deux copies alléliques), les traits pointillés B et C deux événements de recombinaison ectopique (entre deux copies non alléliques).

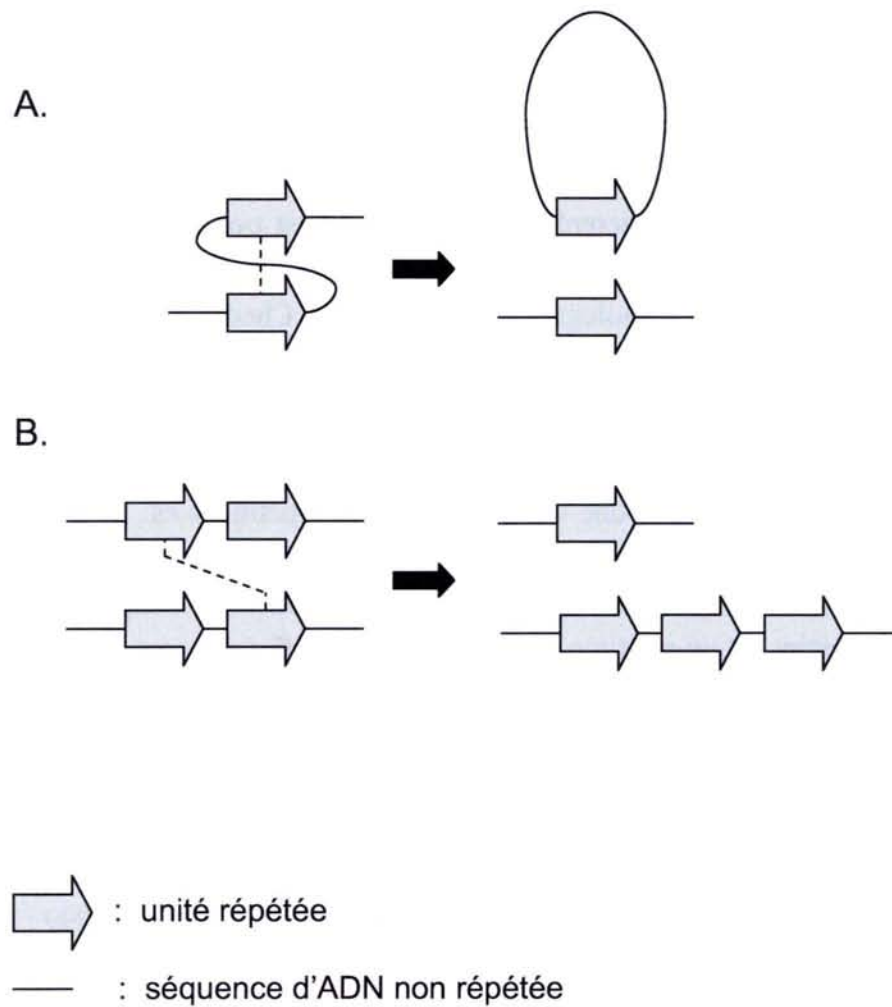


Figure 17 : Recombinaison ectopique entre séquences répétées

A. Recombinaison ectopique entre répétitions directes dispersées portées par la même molécule d'ADN

B. Recombinaison ectopique entre répétitions portées par des molécules d'ADN distinctes.

recombinaison ectopique, impliquant des motifs présents sur des brins d'ADN distincts, produisent une molécule présentant une délétion et une autre molécule présentant une duplication en tandem (Figure 17 B).

La fréquence de recombinaison, ainsi que la nature des mécanismes de recombinaison responsables de l'instabilité génétique des régions riches en motifs répétés, dépendent essentiellement de deux facteurs : la longueur des unités répétées et la distance les séparant. Pour deux motifs répétés en tandem, la fréquence de recombinaison (illégitime et homologue) augmente lorsque la taille des unités répétées augmente (Peeters *et al.*, 1988; Mazin *et al.*, 1991; Bi et Liu, 1994). Lorsque c'est la distance entre les motifs répétés qui augmente, il est vraisemblable que ce soit la recombinaison homologue qui intervienne de façon prépondérante. En effet, la recombinaison illégitime est très sensible à la distance entre les microhomologies alors que la recombinaison homologue est beaucoup moins sensible à la distance entre les régions homologues (Bi et Liu, 1994; Chedin *et al.*, 1994).

2.6 Classification des répétitions d'ADN intragéniques en deux classes

Les répétitions d'ADN peuvent avoir une localisation intragénique ou intergénique. Chez *Mycobacterium tuberculosis*, une vaste majorité des répétitions est localisée à l'intérieur des ORF (Yeramian et Buc, 1999).

Ces répétitions intragéniques peuvent être classées en deux catégories, selon que la taille des unités répétées est un multiple de trois ou non. En effet, les conséquences de la variabilité des régions riches en répétitions ne sont pas de la même nature selon la catégorie. Lorsque les unités répétées n'ont pas une taille multiple de trois nucléotides, alors la modification du nombre de motifs pourra entraîner l'inactivation de l'ORF par décalage du cadre de lecture. Le même mécanisme pourra également conduire à la réversion d'une telle inactivation. Ce passage d'un état « ON » à un état « OFF » est un des mécanismes de régulation de l'expression des gènes par variation de phase (pour revue Henderson 1999 ; Hallet 2001). En revanche, lorsque la taille des motifs répétés est multiple de trois nucléotides, et que ces motifs sont localisés dans la même phase les uns par rapport aux autres, alors l'insertion ou la délétion de motifs répétés n'induit pas d'inactivation de l'ORF par décalage du cadre de lecture. Dans ce cas, si le nombre de motifs répétés change, seule changera la séquence de la protéine codée par le gène variable.

2.7 L'avantage adaptatif est-il conféré par le caractère réitératif ou par le caractère instable des régions riches en répétitions ?

Les séquences d'ADN répétées ont la caractéristique d'être instables. Toutefois on peut se demander lequel des deux caractères confère un avantage adaptatif : le caractère réitératif ou le caractère variable ? Cette question sera abordée par la détermination des conséquences fonctionnelles du caractère réitératif d'une séquence et de sa variabilité. Des hypothèses concernant l'avantage adaptatif conféré par ces caractéristiques de séquences seront émises.

Il est à noter que les exemples présentés ici concernent une certaine catégorie de répétitions : celles localisées à l'intérieur des gènes et dont les motifs ont une taille qui est multiple de trois nucléotides.

2.7.1 Conséquences fonctionnelles du caractère réitératif des régions riches en répétitions

2.7.1.1 Fixation coopérative des protéines à leur ligand : interaction avec les matrices extracellulaires des mammifères

Les protéines nommées MSCRAMM (pour microbial surface components recognizing adhesive matrix molecules) sont des protéines de surface bactériennes interagissant avec les matrices extracellulaires des mammifères. Les matrices extracellulaires, appelées également ECM (pour extracellular matrix), sont des structures macromoléculaires localisées à la base des cellules épithéliales et endothéliales. Elles sont composées de glycoprotéines telles que le collagène, la laminine, la fibronectine et le fibrinogène (pour revue Hay, 1991).

Les cellules de *S. agalactiae* présentent à leur surface la protéine FbsA, protéine interagissant avec le fibrinogène. Le domaine de liaison au fibrinogène est constitué d'un motif, de 16 résidus d'acides aminés, répété en tandem direct (Schubert *et al.*, 2002). Il été montré que la capacité de FbsA à se fixer au fibrinogène est dépendante du nombre de motifs répétés. En effet, lorsque le nombre d'unités répétées augmente, l'affinité de la protéine FbsA pour le fibrinogène augmente (Schubert *et al.*, 2002). Il est donc probable que le caractère réitératif du domaine de liaison au fibrinogène permette à la protéine FbsA d'être liée au fibrinogène de façon stable.

D'autres protéines de liaison aux matrices extracellulaires présentent également un domaine de liaison constitué de motifs répétés. C'est le cas des protéines FnBP (pour fibronectin binding protein) interagissant avec la fibronectine (pour revue Schwarz-Linek *et al.*, 2004). Celles présentes à la surface des cellules de *S. aureus* et *S. pyogenes* ont un domaine de liaison constitué de 4 ou 5 motifs de 35 à 40 aa, répétés également en tandem direct. A l'instar de la protéine FbsA, il est probable que la multiplicité des domaines de liaison permette aux protéines FnBP de se lier de façon stable à la fibronectine.

2.7.1.2 Fixation à une structure répétée

2.7.1.2.1 Domaine d'attachement à la paroi

L'ancrage des protéines à la paroi cellulaire peut être, soit de type covalent, soit de type non covalent. Les domaines d'attachement non covalent à la paroi ont souvent une structure répétitive. C'est le cas des domaines d'attachement à la choline (domaine ChBD, pour Cholin Binding Domain) identifiés chez *S. pneumoniae* (pour revue Lopez *et al.*, 2000). La protéine LytA par exemple présente un domaine ChBD constitué de 6 exemplaires d'un motif répété en tandem direct.

La choline entre dans la composition des acides téichoïques et lipotéichoïques chez cette bactérie, eux-mêmes composants de la paroi cellulaire (voir § 1.1.1.3). Ces composés secondaires de la paroi sont des polymères et présentent en cela une structure réitérative. Ainsi, les acides téichoïques et lipotéichoïques présentent de nombreux résidus choline répartis périodiquement le long de ces polymères. Par conséquent, il est vraisemblable que la structure réitérative des domaines ChBD soit une adaptation à la structure réitérative de ces composés secondaires de la paroi avec lesquels ils interagissent.

Le même raisonnement peut être appliqué à d'autres domaines d'attachement tels que les régions riches en modules GW responsables de la liaison aux acides lipotéichoïques chez *Listeria* (pour revue Cabanes *et al.*, 2002), mais également aux domaines SHL (pour S-layer homology), domaine de liaison aux polysaccharides liés de façon covalente au peptidoglycane (Ries *et al.*, 1997; Mesnage *et al.*, 2000).

2.7.1.2.2 Interaction avec l'ADN : les protéines histone-like

Les régulateurs transcriptionnels AlgR3 et AlgP, des bactéries du genre *Pseudomonas*, présentent un domaine de liaison à l'ADN constitué d'un motif de 4 aa répété en tandem (Deretic et Konyecsni, 1990; Stover *et al.*, 2000). Cette propriété a conduit à les qualifier de protéines histone-like, la protéine histone H1 présentant également un domaine similaire de liaison à l'ADN (Deretic et Konyecsni, 1990).

Il est probable que la sélection de la structure répétée de ce type de domaine soit corrélée à la structure réitérative de l'ADN, l'ADN étant un polymère de nucléosides monophosphates.

A l'instar des protéines interagissant avec les matrices extracellulaires eucaryotes (cf § 2.7.1.1), il a été montré que l'augmentation du nombre de motifs répétés entraîne une augmentation de la constante d'association avec l'ADN (Medvedkin *et al.*, 1995). Cela suggère que la répétition en tandem des unités répétées permet d'augmenter le nombre d'unités fixant l'ADN et donc la stabilité de la fixation du domaine histone-like.

2.7.1.3 Catalyses enzymatiques réitérées permettant la biosynthèse de molécules antibiotiques

Les PKS (pour PolyKetide Synthase) de type I catalysent la biosynthèse de molécules cycliques, appelées macrolides polycycliques, précurseurs d'antibiotiques de haut poids moléculaire (pour revue Staunton et Wilkinson, 1997). Ces PKS sont généralement constituées de plusieurs polypeptides de très haut poids moléculaire se repliant en de nombreux domaines catalytiques. Ces domaines catalytiques sont organisés en modules présentant chacun 3 domaines essentiels : les domaines KS (pour KetoSynthase), AT (pour AcylTransferase) et ACP (pour Acyl Carrier Protein) (Figure 18). Ces modules peuvent également présenter un ou plusieurs des domaines KR (KetoReductase), DH (DeHydratase) et ER (EnoylReductase).

A titre d'exemple, la rapamycine est biosynthétisée par une PKS de type I qui est constituée de trois polypeptides : RAPS1, RAPS2 et RAPS3 d'une longueur respectivement de 8566, 10222 et 6260 aa (Figure 18). A l'instar des autres protéines des PKS de type I, les protéines RAPS présentent une structure répétitive. Par exemple, la protéine RAPS2 présente 6 domaines KS, AT, ACP et KR répétés de façon dispersée (Figure 18) (Schwecke *et al.*, 1995).

La biosynthèse de ces précurseurs de molécules antibiotiques présente un caractère réitératif. En effet, ces molécules cycliques sont synthétisées notamment par condensation en chaîne de molécules de propionate et d'acétate (Staunton et Wilkinson, 1997). Il est vraisemblable que la structure répétée des PKS de type I ait été sélectionnée en raison de la nature réitérative de la biosynthèse de ces molécules cycliques.

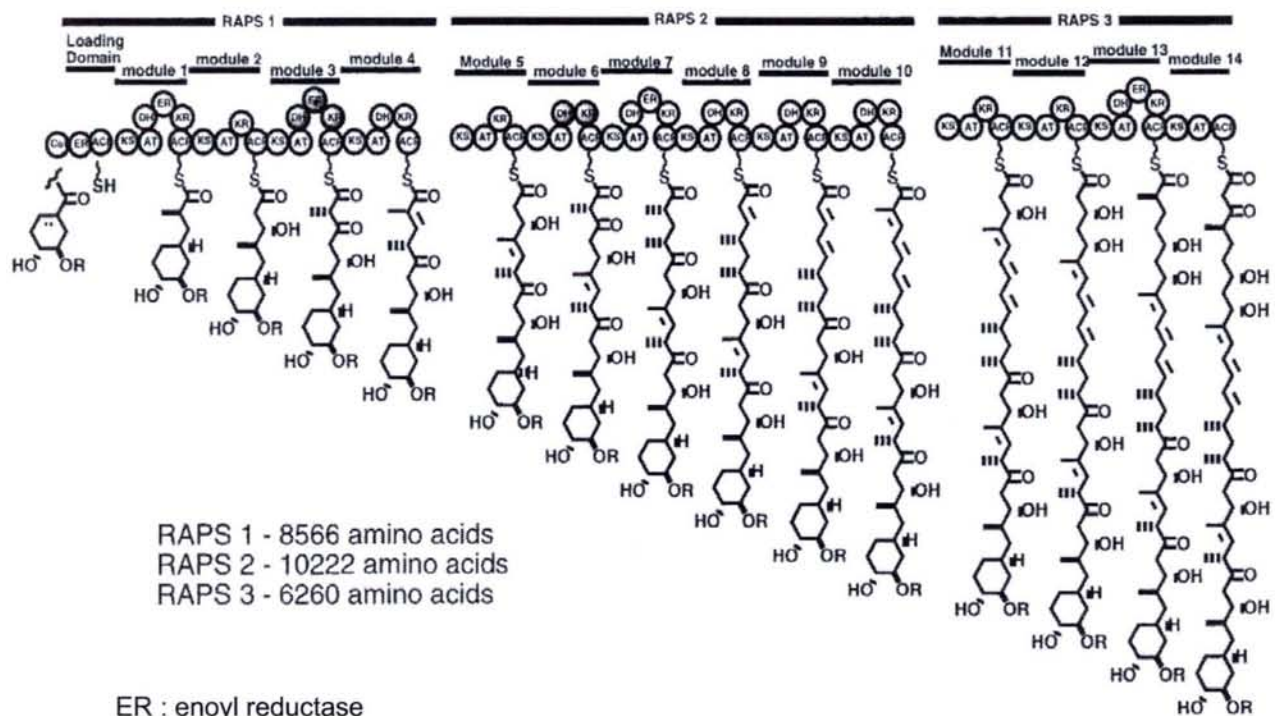


Figure 18 : Caractère réitératif de la biosynthèse du précurseur de la rapamycine et des protéines catalysant cette biosynthèse (d'après Schwecke *et al.*, 1995)

La PKS impliquée dans la biosynthèse du précurseur de la rapamycine est constituée de trois polypeptides, RAPS1 RAPS2 et RAPS3. Ces polypeptides se replient en de multiples modules, chacun symbolisé par un cercle. Au cours de cette biosynthèse, le précurseur naissant est transmis d'un module ACP à l'autre, en partant du premier module ACP de RAPS1, jusqu'au dernier module ACP de RAPS3.

2.7.2 Conséquences fonctionnelles de la variabilité des régions riches en motifs répétés

2.7.2.1 Variabilité des épitopes antigéniques

S. agalactiae est un streptocoque de groupe B pathogène de l'homme, notamment responsable d'infections chez le nouveau né pouvant conduire à la mort.

Il a été montré que la variabilité de la protéine alpha C, observée au sein d'isolats cliniques, pouvait résulter d'une pression de sélection imposée par l'hôte. En effet, la fréquence de délétion spontanée des motifs répétés du gène *bca*, codant la protéine alpha C, a été estimée à 10^{-4} , lorsqu'une souche de *S. agalactiae* est cultivée en milieu Todd-Hewitt. Cette même souche, présentant une protéine alpha C à 9 motifs répétés en tandem, a été inoculée à une souris préalablement immunisée contre cette protéine alpha C. Parmi la population de souches isolées à partir de cette souris, 30% synthétisent une protéine alpha C de masse moléculaire réduite par rapport à la souche d'origine, suggérant que les souches présentant la protéine alpha C à 9 motifs répétés sont contre-sélectionnées. De plus, les souches présentant une protéine alpha C variante seraient moins fréquemment phagocytées par des leucocytes, suggérant que la pression de sélection est imposée par le système immunitaire de l'hôte (Madoff *et al.*, 1996).

Ainsi, la protéine alpha C est une protéine de surface contenant une région riche en répétitions. La séquence protéique de cette région est codée par une séquence d'ADN génétiquement instable et par conséquent variable à haute fréquence. Cette variabilité génétique confère une variabilité antigénique permettant aux streptocoques de groupe B d'échapper au système immunitaire de leur hôte.

2.7.2.2 Variabilité de l'exposition d'un domaine à la surface des cellules

Les protéines Clf (Clumping factor) de *S. aureus* sont des protéines de surface capables d'interagir avec le fibrinogène (McDevitt *et al.*, 1994), le fibrinogène étant une glycoprotéine présente en grande concentration dans le plasma sanguin (pour revue Fuss *et al.*, 2001). Ces protéines Clf présentent en N-terminal un domaine de liaison au fibrinogène et en C-terminal un motif de type LPXTG permettant à la protéine d'être liée de façon covalente au peptidoglycane. La région située entre les deux, présente une séquence de faible complexité puisqu'elle est quasi exclusivement constituée de la répétition en tandem direct du motif DS (Ni Eidhin *et al.*, 1998). Cette région riche en répétitions permettrait au domaine de liaison au fibrinogène d'être exposé à la surface. En l'absence de cette région riche en répétitions, le domaine d'interaction avec le fibrinogène resterait enfoui au coeur de la paroi cellulaire et ne pourrait assurer son rôle (Hartford *et al.*, 1997). Cette région riche en répétitions étant variable en taille, il est possible que la variabilité de cette région entraîne une variabilité de l'exposition du domaine N-terminal de la protéine à la surface des cellules, et donc une variabilité de la fonctionnalité de la protéine Clf (Hartford *et al.*, 1997).

2.7.2.3 Modulation de l'activité du régulateur transcriptionnel CtsR d'un état « ON » à un état « OFF »

La protéine CtsR de *Listeria monocytogenes* est un répresseur transcriptionnel régulant notamment le niveau de résistance aux hautes pressions hydrostatiques, traitement utilisé en industrie agro-alimentaire pour stériliser les aliments. Il est possible de sélectionner des souches de *L. monocytogenes* variantes qui sont 1000 fois plus tolérantes aux hautes pressions hydrostatiques (Karatzas et Bennik, 2002). Les auteurs ont montré que cette résistance pouvait résulter de la délétion d'un codon glycine du gène codant CtsR. Ce codon glycine est localisé dans une région riche en répétitions, constituée de quatre codons GG(T/C) répétés en tandem, tous codant l'incorporation d'une glycine. La délétion de ce résidu glycine entraîne l'inactivation de la fonction de répresseur transcriptionnel de la protéine CtsR (Karatzas *et al.*, 2003).

Il est supposé que ce type de délétion ait lieu par glissement de fourche de réplication. Il est donc probable que cette mutation soit réversible, bien que cela ne soit pas démontré (Karatzas *et al.*, 2003). L'instabilité de la région riche en motifs répétés pourrait entraîner, de façon réversible, le passage d'un état fonctionnel à un état non fonctionnel de la protéine CtsR. Son activité serait ainsi régulée par variation de phase.

En conclusion, l'instabilité de cette région constituée de répétitions pourrait constituer un mécanisme de régulation de l'activité de la protéine CtsR.

2.7.2.4 Modulation de la spécificité d'une enzyme de restriction de type I

Les systèmes R-M de type I sont des complexes multi-enzymatiques pouvant catalyser le clivage, ou restriction, et la modification par méthylation d'une séquence d'ADN préalablement reconnue spécifiquement (pour revue Murray 2000). La reconnaissance de la séquence cible est assurée par une protéine généralement nommée HsdS (pour host specificity of DNA, specificity subunit). Les protéines HsdS présentent deux domaines, appelés TRD (pour target recognition domain), chacun reconnaissant spécifiquement une des deux composantes de la séquence cible. Ces deux domaines sont reliés par une charnière inter-domaine dont la variabilité de la taille peut conférer une variabilité de la spécificité de substrat des protéines HsdS (Price *et al.*, 1989). En effet, les enzymes de restriction *EcoR124* et *EcoR124/3* reconnaissent respectivement les sites GAA(N6)RTCG et GAA(N7)RTCG. Ces séquences cibles diffèrent uniquement par le nombre de nucléotides espaçant les deux composantes GAA et RTCG reconnues spécifiquement. Or, les gènes *hsdS* des systèmes de restriction *EcoR124* et *EcoR124/3* présentent respectivement 2 et 3 exemplaires d'un motif de 12 pb, répété en tandem direct. Les auteurs ont montré que la différence de spécificité des deux enzymes de restriction résulte de la différence de longueur de cette séquence inter-domaine (Price *et al.*, 1989).

Ainsi, la variabilité du nombre de motifs répétés peut permettre de générer de nouvelles spécificités de substrat.

2.7.2.5 *Génération de nouvelles spécificités d'hôte chez un phytopathogène*

Le gène *pthA* est impliqué dans l'avirulence de bactéries phytopathogènes appartenant au genre *Xanthomonas* (pour revue Brunings et Gabriel, 2003). Un gène d'avirulence est un gène pouvant conférer à un phytopathogène la perte de son caractère pathogène (Flor, 1971).

La région centrale du gène d'avirulence *pthA* est constituée d'un motif de 102 pb répété quasi parfaitement en tandem direct (Brunings et Gabriel, 2003). Il a été montré que l'instabilité de cette région riche en répétitions est un facteur de variabilité de la spécificité d'hôte. En effet, lorsque l'allèle *pthA* présente 17 motifs répétés, la bactérie est capable de provoquer un cancer chez le citronnier mais pas chez le haricot. La présence de cet allèle du gène *pthA* dans le génome du phytopathogène déclenche, chez le haricot, un système de défense entraînant la neutralisation du phytopathogène. En revanche, lorsque ce même allèle est délété de 6 motifs répétés, la bactérie ne peut provoquer de cancer ni chez le citronnier ni chez le haricot, car cet allèle déclenche un système de défense chez chacune de ces deux plantes (Yang et Gabriel, 1995).

Ainsi, l'instabilité d'une région riche en répétitions peut entraîner des modifications du spectre d'hôte d'une bactérie phytopathogène.

3 Création de nouvelles fonctions par réassortiment de domaines protéiques chez les bactéries

3.1 Architecture des protéines

3.1.1 Hierarchie structurale des protéines

Les protéines sont des polymères non branchés d'acides aminés dont la structure tridimensionnelle résulte de la formation de liaisons non covalentes entre les résidus d'acides aminés. Une chaîne polypeptidique atteint sa conformation fonctionnelle, dite native, par formation de structures dites secondaires, tertiaires et quaternaires.

La structure secondaire résulte du repliement tridimensionnel d'une chaîne polypeptidique en éléments de structures tels que les hélices α et les feuilletts β . Ces éléments de structure secondaire sont reliés les uns aux autres par des boucles, autres éléments de structures secondaires.

L'arrangement spatial des éléments de structures secondaires les uns par rapport aux autres constitue la structure tertiaire d'une protéine. Pour les protéines monomériques, la structure tertiaire est le plus haut niveau d'organisation.

La structure quaternaire concerne les protéines constituées de plusieurs chaînes polypeptidiques. Elle correspond à l'arrangement spatial des éléments de structures tertiaires, appelés dans ce cas sous-unités.

3.1.2 Organisation en domaines distincts de la structure tertiaire des protéines

La structure tertiaire des protéines peut être composée d'une ou plusieurs régions appelées domaines. Un domaine est codé par un fragment de gène qui sera ici dénommé « module ».

D'un point de vue structural, un domaine est une portion de polypeptide capable de se replier en une structure compacte, indépendamment des autres régions du même polypeptide (Doolittle, 1995). Une séquence formée d'une cinquantaine de résidus d'acides aminés peut se replier en un domaine dont la structure est suffisamment stable. Toutefois, de plus petits segments polypeptidiques peuvent constituer un domaine structural si la structure comprend des éléments stabilisateurs tels que des ions métalliques ou des ponts disulfures (Doolittle, 1995). Dans une protéine multi-domaine, les domaines sont connectés entre eux par des fragments polypeptidiques appelés charnière (ou linker). Généralement, un domaine assure une fonction élémentaire. Cette fonction peut consister, par exemple, à catalyser une réaction, interagir avec d'autres molécules, structurer, ou encore transporter des molécules d'un compartiment cellulaire à un autre.

La glutamine méthyltransférase HemK d'*E. coli* est un exemple représentatif de cette architecture des protéines. Elle présente deux domaines, chacun assurant une fonction : un domaine de fixation du substrat, la glutamine, en N-terminal et un domaine catalytique méthyltransférase en C-terminal (Figure 19)(Yang *et al.*, 2004). L'association d'un domaine reconnaissant la glutamine à un domaine catalysant une réaction de méthylation confère à la protéine son activité de glutamine méthyltransférase.

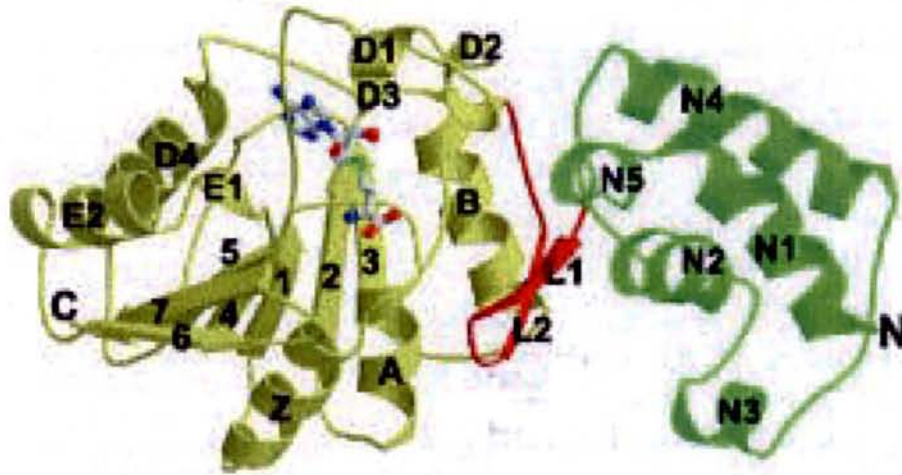


Figure 19 : Structure de la protéine HemK (d'après Yang *et al.*, 2004)

Représentation schématique de la structure tridimensionnelle de la protéine HemK. Les rubans représentent les hélices α et les flèches les brins β . Chaque élément de structure secondaire est nommé par une lettre, un chiffre, ou les deux. La localisation des extrémités N-terminale (N) et C-terminale (C) est indiquée. Le domaine N-terminal, reconnaissant le substrat, est représenté en vert. Le domaine C-terminal, assurant la catalyse enzymatique, est représenté en jaune. Celui-ci est complexé à une molécule de S-adenosyl-L-homocysteine. La charnière joignant les deux domaines l'un à l'autre est représentée en rouge.

3.2 Création expérimentale de nouvelles fonctions par construction de nouvelles combinaisons de domaines

3.2.1 Introduction

Si la fonction des protéines multi-domaine résulte de l'association des domaines les constituant, il semble alors vraisemblable que la création de nouvelles associations de domaines puisse générer de nouvelles fonctions. Cette hypothèse a été testée par construction de protéines chimériques et analyse de leur activité.

3.2.2 L'ATCase et l'OTCase

Houghton *et al.* (1989) ont utilisé comme modèle d'étude deux transcarbamoylases : l'aspartate transcarbamoylase (ATCase) et l'ornithine transcarbamoylase (OTCase). Ces enzymes présentent deux domaines : un domaine catalytique et un domaine de fixation du substrat. Les auteurs ont construit un gène chimérique en fusionnant le module de gène codant le domaine catalytique de l'OTCase avec le module de gène codant le domaine de spécificité de substrat de l'ATCase. Cet échange conduit à une modification de la spécificité de substrat : l'activité OTCase est remplacée par une activité ATCase. Toutefois, l'ATCase et l'OTCase présentent toutes deux un domaine catalytique en commun. Qu'en est-il lorsque l'échange de domaines est réalisé entre enzymes ne présentant pas de domaine assurant une fonction commune ?

3.2.3 CPL7 et LytA

D'autres expériences de construction de chimères ont été réalisées, en particulier avec des gènes codant des hydrolases de la paroi bactérienne (Diaz *et al.*, 1990; 1991; Croux *et al.*, 1993). Dans l'une d'elles, le gène chimérique a été construit à partir de gènes ne présentant pas d'identité nucléotidique entre eux (Diaz *et al.*, 1991). Ces gènes codent LytA, une amidase de *S. pneumoniae*, et CPL7, une muramidase du phage Cp-7 (phage infectant les cellules de *S. pneumoniae*). Ainsi, ces deux enzymes catalysent l'hydrolyse de différentes liaisons du peptidoglycane. De plus, la fixation de LytA à la paroi cellulaire est dépendante de la choline, un composant de la paroi de *S. pneumoniae*, et ne l'est pas pour la protéine phagique CPL7, suggérant que les deux protéines sont fixées à des composés différents de la paroi (Garcia *et al.*, 1990).

Chacune de ces protéines présente deux domaines : un domaine catalytique à l'extrémité N-terminale et un domaine de fixation à la paroi à l'extrémité C-terminale. LytA présente un domaine amidase et un domaine constitué de motifs ChBD (choline binding domain) de fixation aux résidus cholines de la paroi (pour revue Lopez *et al.*, 2000). CPL7 présente, quant à elle, un domaine muramidase et un domaine vraisemblablement impliqué dans la fixation à un autre composant de la paroi cellulaire que la choline.

Un gène chimérique constitué de la région 5' du gène codant CPL7 et de la région 3' du gène codant LytA a été construit. Ce gène chimérique code une protéine présentant le domaine catalytique muramidase de CPL7 ainsi que le domaine d'attachement de LytA. La protéine chimérique arbore une activité chimérique : c'est une muramidase dépendante de la choline (Diaz *et al.*, 1991). Ces résultats montrent qu'une nouvelle fonction peut être générée par réassortiment de domaines.

3.3 Des réassortiments de domaines ont été sélectionnés au cours de l'évolution

Les transcarbamoylases ATCase et OTCase présentent un domaine apparenté, le domaine transcarbamoylase, et un domaine de spécificité de substrat divergent, tant d'un point de vue de la fonction que d'un point de vue de la séquence. Deux hypothèses peuvent expliquer la divergence entre les domaines de spécificité de substrat. Dans la première, ces deux domaines ont un ancêtre commun et auraient divergé par accumulation de mutations ponctuelles. Dans la deuxième, un réarrangement de gènes aurait engendré la formation d'un gène chimérique codant une protéine présentant une nouvelle association de domaines.

Dans le cas de la famille des protéines HsdS, l'hypothèse de l'évolution par réassortiments de domaines peut être favorisée (pour revue Murray, 2000) (Dybvig *et al.*, 1998; O'Sullivan *et al.*, 2000). Les protéines HsdS entrent dans la composition des systèmes de restriction de type I. Elles assurent la reconnaissance de la séquence cible et présentent deux domaines, appelés TRD (pour target recognition domain), chacun reconnaissant spécifiquement une des deux composantes de la séquence d'ADN cible. Il est possible avec ces protéines de créer de nouvelles spécificités de substrat par génération de nouvelles combinaisons de domaines TRD (Nagaraja *et al.*, 1985). En effet, une chimère générée à partir des gènes codant les protéines HsdS des systèmes de restriction *Sty*SPI et *Sty*LTIII, reconnaît une séquence d'ADN chimérique (Figure 20).

Des systèmes naturels de création de nouvelles spécificités par réassortiment de domaines ont été identifiés (Dybvig *et al.*, 1998; O'Sullivan *et al.*, 2000). Par exemple, les plasmides circulaires pAH33 et pAH82 de *L. lactis* présentent chacun un gène codant des protéines HsdS reconnaissant des séquences d'ADN différentes. Un variant de cette souche de *L. lactis* présentant un unique plasmide pAH90 a été isolé. L'existence de ce plasmide résulte de la co-intégration de pAH33 et de pAH82 par recombinaison homologue entre les gènes *hsdS*. Les régions ayant été le siège de cet événement de recombinaison homologue sont les régions codant la charnière joignant les domaines TRD. La conséquence de cet événement de recombinaison homologue est la formation de deux gènes chimériques, chacun codant une protéine présentant une nouvelle combinaison de domaines TRD. Il a été montré que ces gènes chimériques confèrent de nouvelles spécificités de séquence d'ADN (O'Sullivan *et al.*, 2000).

3.4 Caractère versatile des domaines

3.4.1 Construction de banques de données permettant d'établir un répertoire de combinaisons de domaines

La compréhension de l'évolution des protéines par réassortiment de domaines est en partie issue des informations provenant de l'analyse de banques de données. Ces banques de données ont été construites à partir de l'ensemble des données de biologie structurale et de génomique (pour revue Vogel *et al.*, 2004).

Plus précisément les données de structures tridimensionnelles de protéines ont permis d'établir une banque de familles de domaines protéiques nommées SCOP (pour structural classification of proteins) (Murzin *et al.*, 1995). Dans cette banque, un domaine est défini

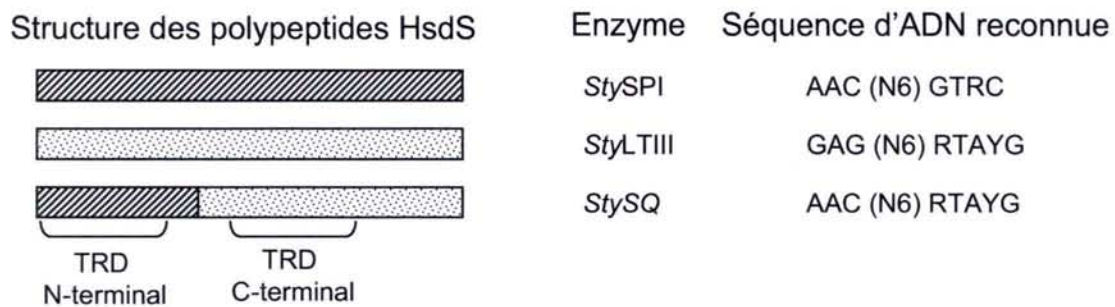


Figure 20 : Génération d'une enzyme de restriction de type I présentant une nouvelle spécificité par réassortiment de domaines (d'après Murray, 2000)

Les fragments polypeptidiques originaires des protéines HsdS des enzymes de restriction *Sty*SPI et *Sty*LTIII sont respectivement représentés en hachuré et grisé.

comme une unité évolutive indépendante, capable de former un domaine structural. Ce domaine peut former une protéine présentant un seul domaine, ou alors faire partie d'une protéine multi-domaine. Les domaines sont regroupés en familles sur les critères cumulés d'identité de séquence, d'identité structurale et d'identité fonctionnelle. Il est à noter que dans cette banque, les domaines appartenant à la même famille ont des fonctions apparentées et donc non nécessairement identiques (Murzin *et al.*, 1995).

Cette banque de domaine SCOP a été utilisée pour annoter les protéines issues de l'annotation des génomes. Cette annotation de masse des protéines a ainsi généré d'autres banques de données telles que SUPERFAMILY (Madera *et al.*, 2004). Cette dernière a été analysée dans le but de déterminer les caractéristiques du répertoire de combinaisons de familles de domaines.

3.4.2 Caractéristiques du répertoire de combinaisons de familles de domaines

L'analyse du répertoire de combinaisons de familles de domaines suggère que 65% des protéines bactériennes présentent au moins deux domaines (Teichmann *et al.*, 1998; Apic *et al.*, 2001). Ce pourcentage est très inférieur à celui des organismes eucaryotes dont plus de 80% des protéines sont des protéines multi-domaines (Apic *et al.*, 2001).

L'analyse de ce répertoire montre également que certains domaines sont préférentiellement répétés en tandem au sein d'une même protéine. Toutefois, la proportion de protéines bactériennes présentant des domaines dupliqués en tandem est faible (4%). Il en est de même de la proportion de familles de domaines dont les domaines peuvent être dupliqués en tandem au sein d'une même protéine (14%) (Apic *et al.*, 2001).

Apic *et al.* (2001) ont analysé le répertoire de domaine dans le but d'estimer le caractère versatile des domaines. La versatilité est, selon Apic *et al.* (2001), la propriété qu'a un domaine d'une famille donnée d'être associé à des domaines appartenant à de nombreuses familles différentes. Cette analyse montre que le caractère versatile est très variable selon les familles (Apic *et al.*, 2001). En effet, les domaines d'une grande majorité de familles sont peu versatiles, les domaines ne sont majoritairement associés qu'à des domaines appartenant à une ou deux familles. En revanche, les domaines d'une minorité de famille sont au contraire très versatiles. Le domaine P-loop nucleotide triphosphate hydrolase serait le plus versatile de tous ; il est rencontré dans 47 combinaisons différentes. Les domaines de cette famille catalysent l'hydrolyse d'ATP ou de GTP et peuvent avoir une fonction de kinase ou de transférase. Les domaines Rossmann sont également très versatiles : ils ont été rencontrés dans 29 combinaisons différentes. Ces domaines sont capables de réduire ou d'oxyder le cofacteur NAD(P)(H), et de transmettre le pouvoir réducteur ou oxydant produit par la réaction.

Il est à noter qu'un domaine qualifié de peu versatile par Apic *et al.* (2001) ne signifie pas qu'il évolue peu par réassortiment de domaines. En effet, les réassortiments peuvent être restreints aux domaines appartenant à la même famille. Etant donné que deux domaines d'une même famille de la banque SCOP peuvent avoir des fonctions, certes apparentées, mais différentes, un réassortiment impliquant des domaines appartenant à une même famille peut engendrer la création d'une nouvelle fonction.

3.4.3 Origine du caractère disparate de la versatilité des domaines

3.4.3.1 Contraintes sur la recombinaison des séquences d'ADN

Les réarrangements de modules de gène, codant des domaines protéiques, sont facilités par l'existence des introns chez les organismes eucaryotes dont les gènes ont une structure exon/intron. En effet, les bornes des exons ont tendance à coïncider avec les bornes des domaines protéiques (Liu et Grigoriev, 2004). Cette coïncidence entre les bornes des exons et des domaines serait propice aux réassortiments de domaines. En effet, le réassortiment de domaines protéiques nécessite dans ce cas le réassortiment d'exons (ou exon shuffling). Or, la fonction d'intron impose peu de contraintes de séquence. Ainsi, les introns pourraient être des régions favorables à l'accueil de bornes de réarrangements responsables de réassortiments d'exons, et par conséquent de réassortiment de domaines (Gilbert, 1978; Patthy, 1991; 1994).

En revanche, les organismes dont les gènes ne présentent pas d'intron ne peuvent pas bénéficier de cet avantage. Chez ces organismes, les bornes des réarrangements aboutissant à des réassortiments de domaines doivent être situées à l'intérieur des régions codantes. La contrainte qui en découle est que le réarrangement doit respecter la phase de lecture. Cette contrainte pourrait être un des facteurs à l'origine du plus faible pourcentage de protéines multi-domaines chez les organismes procaryotes par rapport aux organismes eucaryotes (Apic *et al.*, 2001).

3.4.3.2 Certaines fonctions élémentaires sont requises en grand nombre, d'autres pas

Les domaines les plus versatiles sont des domaines impliqués dans l'hydrolyse de l'ATP ou le GTP et dans l'utilisation du pouvoir redox du NAD(P)(H). Ces fonctions sont nécessaires au déroulement d'un très grand nombre de réactions catalysées au sein de la cellule. C'est d'ailleurs probablement la raison pour laquelle ces domaines, en plus d'être les plus versatiles, sont les plus abondants (Apic *et al.*, 2001). En revanche, il est vraisemblable que la fonction de certains domaines ne soit requise que dans un petit nombre d'associations. Ainsi, le caractère versatile des domaines pourrait être disparate car certaines fonctions élémentaires sont requises plus fréquemment que d'autres.

3.4.3.3 Nécessité d'une co-adaptation des domaines

Une mutation ponctuelle située dans un gène et qui est désavantageuse pour son maintien peut être, soit maintenue, soit contre-sélectionnée. Cette mutation peut entre autres être maintenue par mutation suppressive. Dans le cas d'une protéine multi-domaine, la modification de l'activité d'un domaine due à une mutation peut être compensée par une mutation affectant l'activité d'un autre domaine. En effet, la protéine SmNifA de *Sinorhizobium meliloti* présente trois domaines : un domaine de liaison à l'ADN, un domaine d'activation de la transcription et un domaine de régulation. Son domaine de fixation à l'ADN interagit faiblement avec sa séquence cible par rapport aux protéines NifA des autres organismes (Juarez *et al.*, 2000). Il a été montré que cette caractéristique résulte d'une mutation ponctuelle d'un résidu conservé chez les autres protéines NifA. Toutefois, la faible capacité du domaine de liaison à se fixer à l'ADN est compensée par une très haute efficacité du domaine d'activation par rapport aux autres protéines NifA. Cette différence d'efficacité

est due à une mutation d'un résidu d'acide aminé du domaine d'activation. Les auteurs suggèrent que le domaine de fixation à l'ADN et le domaine d'activation ont co-évolué et que le résultat de cette co-évolution ait été le maintien d'une activité globale constante de la protéine, bien que l'activité de chacun des domaines ait fluctué (Juarez *et al.*, 2000).

Compte tenu de cette co-évolution des domaines, il est vraisemblable que la génération d'une nouvelle combinaison de domaines ne soit pas directement productive et que la fonctionnalité de la protéine chimérique nécessite une co-adaptation des domaines nouvellement associés. La nécessité d'une co-adaptation de ces domaines pourrait être un facteur favorisant la contre-sélection de la combinaison nouvellement formée et pourrait expliquer le faible caractère versatile de certains domaines.

4 Objectifs

L'objectif initial de cette thèse visait à caractériser l'implication de gènes de *Streptococcus thermophilus* dans la défense contre le stress oxydant. *S. thermophilus* est une bactérie non pathogène utilisée en industrie agroalimentaire pour la production de produits laitiers, à partir desquels toutes les souches connues ont été isolées. Sa température optimale de croissance est 42°C et elle fermente les carbohydrates pour produire de l'acide lactique (Schleifer, 1986). Lorsqu'elle est cultivée en milieu lait l'abaissement du pH, résultant de cette production d'acide lactique, est responsable de la précipitation des caséines du lait, entraînant la formation d'un gel appelé coagulum.

C'est une bactérie Gram positive dont le génome est constitué d'un unique chromosome circulaire de 1,8 Mb d'ADN constitué d'un faible pourcentage en base G+C (39%). La comparaison du génome de deux souches de *S. thermophilus*, les souches CNRZ1066 et LMG18311, révèle un faible polymorphisme (0,15%), suggérant que leur ancêtre commun vivait il y a entre 3000 et 30000 ans. L'émergence de *S. thermophilus* coïncide avec le moment où l'homme aurait commencé à produire des laitages fermentés, il y a environ 7000 ans (Bolotin *et al.*, 2004).

Au commencement de ces travaux de thèse, nous disposions de mutants, obtenus par mutagenèse insertionnelle aléatoire présentant potentiellement une sensibilité accrue au stress superoxyde par rapport à la souche sauvage. Durant les expériences de confirmation de cette sensibilité, nous avons identifié un mutant présentant de longues chaînes de cellules.

Bien que ce mutant ne présente pas de sensibilité au stress superoxyde, sa caractérisation permettait toutefois d'aborder un tout autre thème d'intérêt : la ségrégation des cellules de *S. thermophilus*. Des gènes impliqués dans la ségrégation cellulaire sont connus chez les autres coques se développant sous la forme de chaînes. Chez *S. pneumoniae*, la ségrégation cellulaire est conférée par les gènes *lytA* et *lytB* (Sanchez-Puelles *et al.*, 1986; Garcia *et al.*, 1999), et chez *L. lactis* par *acmA* (Buist *et al.*, 1995). Or l'analyse du génome de *S. thermophilus* ne révèle pas d'homologue de ces gènes. Ainsi, la caractérisation de ce mutant de *S. thermophilus* ouvrait la possibilité d'identifier un nouveau gène impliqué dans la ségrégation cellulaire.

La section résultat est composée de trois parties :

- la première partie est consacrée à la caractérisation de gènes impliqués dans la défense contre le stress oxydant. Les résultats y seront présentés succinctement et le lecteur est invité à se reporter à l'article 1 figurant en annexe pour plus de détails.

- la seconde partie, fera état de l'ensemble des données obtenues sur le gène *cse*. Seront notamment exposés les résultats prouvant l'implication de ce gène dans la ségrégation cellulaire. L'étude des caractéristiques génétiques de *cse*, c'est-à-dire ses caractères chimérique, variable et mosaïque, y seront présentés en détails.

- dans la troisième partie, seront présentés les résultats obtenus suite aux expériences visant à déterminer si la méthode de sélection de mutants de la longueur des chaînes, méthode mise au point chez *L. lactis* (Mercier *et al.*, 2002), est applicable à *S. thermophilus*.

Matériel et méthodes

La plupart du matériel et des méthodes utilisés dans ce travail sont décrits dans les articles 1 et 2 figurant en annexe. Ceux qui n'y sont pas décrits sont présentés ici.

1 Construction de mutants par remplacement allélique

La méthode consiste à remplacer un allèle du génome de *S. thermophilus* par un allèle mutant grâce à deux événements de recombinaison homologue situés de part et d'autre des régions à remplacer. L'allèle mutant, cloné dans le plasmide pGh9, est flanqué des régions destinées à être la cible du double événement de recombinaison homologue (Figure 21).

La souche de *S. thermophilus* devant subir le remplacement allélique est transformée avec le plasmide contenant la construction nécessaire (voir les articles 1 et 2 pour le détail des constructions). Les transformants sont sélectionnés à température permissive (30°C) en présence d'érythromycine (5 µg/ml).

Puis, un des transformants est ensemencé au 1/100^{ème} en milieu M17 liquide supplémenté par 2 µg/ml d'érythromycine, incubé pendant 2 heures 30 minutes à 30°C puis 2 heures 30 minutes à 42°C et enfin étalé sur milieu M17 solide supplémenté par 2 µg/ml d'érythromycine et incubé à 42°C en jarre anaérobie. La température de 42°C étant restrictive pour le plasmide, ceci permet de sélectionner des événements de simple recombinaison, donnant lieu à l'intégration du plasmide au sein du chromosome, au locus ciblé.

Afin d'accomplir un remplacement allélique, des événements de double recombinaison, dérivant d'un simple recombinant, sont sélectionnés par la perte de la résistance à l'érythromycine. Pour cela, un clone ayant subi l'intégration du plasmide par un événement unique de recombinaison homologue est ensemencé au 1/1000^{ème} en milieu M17 liquide non supplémenté en érythromycine et incubé pendant 24 heures à 30°C, afin de favoriser l'excision du plasmide par recombinaison homologue. Les cellules sont étalées sur milieu M17 solide non supplémenté en érythromycine, et incubées en jarre à 42°C. Parmi ces clones, ceux qui ont perdu la résistance à l'érythromycine sont sélectionnés par réplique.

Ces doubles recombinants peuvent, soit avoir recouvré la séquence originale, soit avoir intégré l'allèle porté par le plasmide. L'ultime sélection des clones ayant effectivement subi le remplacement allélique est effectuée par analyse moléculaire.

2 Analyses phénotypiques

2.1 Stress oxydant

2.1.1 Courbe de survie des cellules en présence de ménadione

Afin de tester la sensibilité au stress oxydant d'une souche, la survie des cellules en présence d'une gamme de concentrations en ménadione a été étudiée. Pour cela, une culture cellulaire en phase exponentielle de croissance ($DO_{600} = 0,6$) est aliquotée par 1 ml, et chaque aliquot est incubé, pendant 3 heures, avec une concentration donnée de ménadione (de 0 à 20 mg/ml). Les suspensions cellulaires sont ensuite diluées jusqu'à la dilution appropriée,

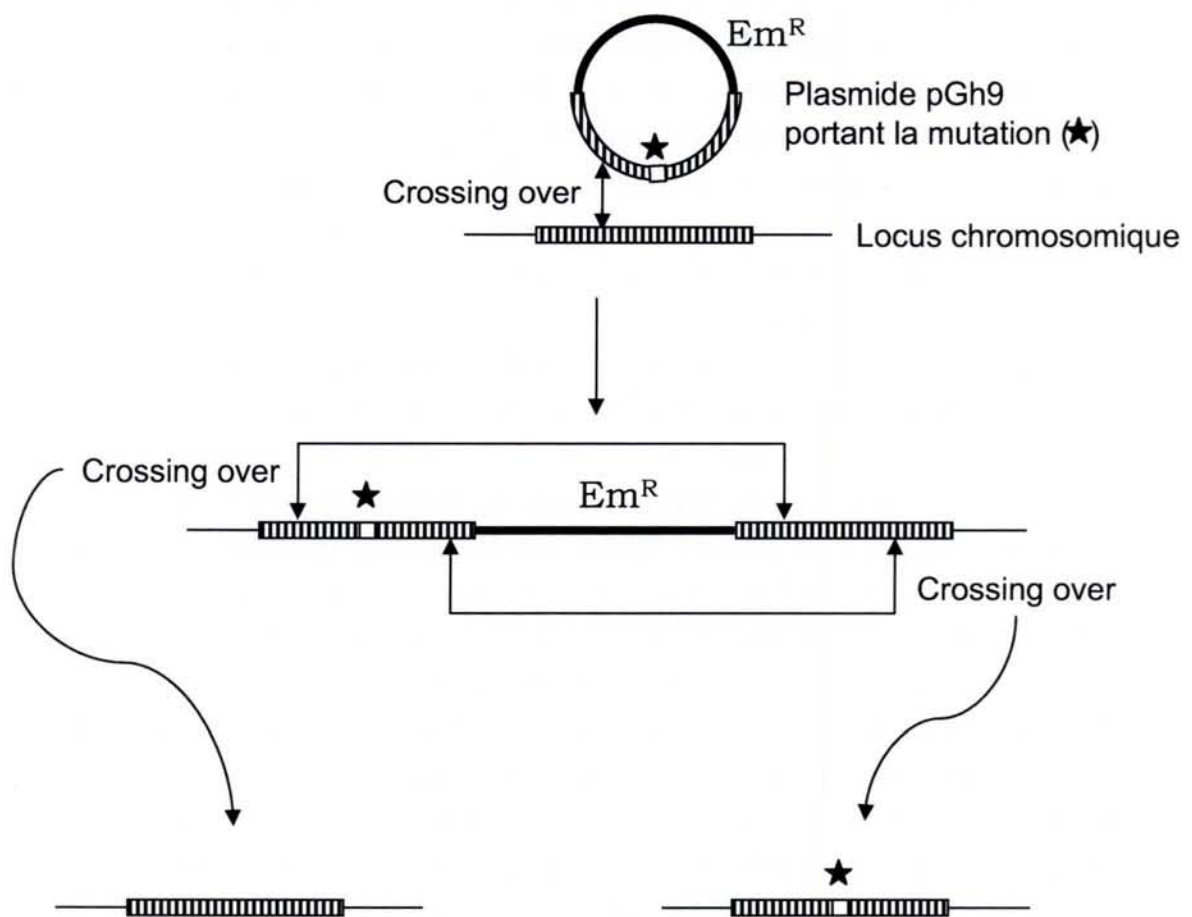


Figure 21 : Représentation schématique de la méthode de mutagenèse dirigée par remplacement d'allèles

étalées sur milieu TPPY gélosé, puis incubées pendant 20 heures en jarre anaérobie. Le pourcentage de survie est calculé en rapportant le nombre de colonies obtenu après traitement, à celui obtenu en absence d'agent oxydant. Concernant les mutants issus de la collection de mutants d'insertion, chaque clone est traité en parallèle à un clone issu de cette même collection (le clone 2C3) afin d'identifier d'éventuels artefacts. Le clone 2C3 a été choisi comme témoin car il ne présente ni résistance ni sensibilité particulière au stress oxydant.

2.1.2 Test de diffusion sur boîte

Une culture cellulaire est ensemencée au $1/100^{\text{cmc}}$ en milieu TPPY liquide à partir d'une culture en milieu lait elle-même ensemencée la veille. Une fois en phase exponentielle de croissance ($DO_{600} = 0,6$), un volume de 200 μl de cette culture est ajouté à 4 ml de TPPY Top agar en surfusion. Ce mélange est homogénéisé, puis déposé sur une boîte de Pétri contenant 20 ml de milieu TPPY. Une fois la gélose solidifiée, un disque imbibé de 10 μl d'une solution d'agent oxydant est déposé à sa surface et la boîte est incubée pendant 20 heures à 42°C.

Un gradient de concentration est généré par diffusion de l'agent oxydant, à partir du disque imbibé. Ainsi, suite à ce traitement, il se délimite autour du disque, sur un tapis confluent de cellules, une zone d'inhibition de la croissance.

2.2 Adhésion à une surface en verre

Les tests d'adhésion ont été réalisés comme l'ont décrit Mercier *et al.* (2002). Des cellules cultivées sur la nuit en milieu M17 ont été collectées par centrifugation, lavées trois fois avec du milieu M17 frais, puis resuspendues dans du milieu M17 frais à une $DO_{600\text{nm}}$ finale de 0,1. Les lames de verre, chacune placée dans une boîte de Pétri, sont incubées en présence de 30 ml de suspension bactérienne, à 20°C pendant 4 h. Ensuite les lames sont rincées délicatement 5 fois avec de l'eau distillée stérile, puis colorées avec une solution dans l'eau d'acridine orange (0,01%). Les cellules colorées sont visualisées et photographiées grâce à un microscope à épifluorescence (Leica DMLB, Tokio, Japon), équipé d'un objectif 10X. Les images sont analysées avec le logiciel IMAGETOOL afin d'estimer le pourcentage de cellules recouvrant la lame de verre.

2.3 Dosage d'exopolysaccharides

Les exopolysaccharides (EPS) produits par *S. thermophilus* sont extraits selon la méthode de Robson *et al.* (1974) et dosés selon la méthode de Dubois *et al.* (1951). Les souches de *S. thermophilus* sont cultivées dans 25 ml de milieu lait durant une nuit. Ensuite les protéines sont précipitées grâce à une solution de TCA (20%), puis séparées par centrifugation à 20000 g pendant 20 min. Le surnageant est prélevé et un 1 volume d'acétone à 4°C est ajouté afin de précipiter les EPS durant la nuit à 4°C. Les EPS sont séparés de la phase liquide par centrifugation à 20000 g pendant 20 min, puis solubilisés dans 500 μl d'eau ultra pure pendant 2 h à température ambiante. Ensuite, 1 volume d'une solution de TCA (20%) est ajouté afin de précipiter les protéines résiduelles. Les protéines précipitées sont séparées de la phase liquide contenant les EPS par centrifugation à 20000 g pendant 20 min. Les EPS sont à nouveau précipités avec de l'acétone à 4°C,

pendant 15 min à 4°C. Les EPS sont séparés de la phase liquide par centrifugation à 20000 g pendant 20 min, puis solubilisés dans 1 ml d'eau pendant 2 h à température ambiante. Afin d'estimer la concentration en EPS, 50 µl de phénol sont ajoutés à 1 ml de la solution d'EPS diluée au ¼. Puis 5 ml d'acide sulfurique à 98% sont ajoutés. L'ensemble est vortexé puis incubé à température ambiante pendant 10 min, incubé à 30°C pendant 10 min, puis la DO_{490nm} est lue. Une gamme étalon est réalisée selon le même protocole avec des solutions de glucose (concentration : 10, 20, 50, 80, 100, 150 mg/l) afin de pouvoir estimer la concentration en EPS, en équivalent glucose.

RESULTATS

1 Identification du gène *ese* par caractérisation de gènes impliqués dans la défense contre le stress oxydant

La majeure partie des résultats présentés dans ce chapitre est détaillée dans l'article 1 figurant en annexe.

1.1 Isolement de onze mutants sensibles à la ménadione

Des travaux antérieurs visaient à identifier des gènes impliqués dans la défense contre le stress oxydant chez *S. thermophilus* CNRZ368 (thèse d'Annabelle Thibessard, 2002). A cette fin, une banque de mutants de *S. thermophilus* avait été construite par mutagenèse insertionnelle aléatoire. Cette mutagenèse avait été réalisée grâce au plasmide pGh9:ISS1, plasmide à réplication thermosensible (Maguin *et al.*, 1996) capable de se transposer aléatoirement dans le génome de *S. thermophilus* (Thibessard *et al.*, 2002a).

Une collection de 2112 mutants ainsi obtenue avait été criblée sur le critère de sensibilité et de résistance à la ménadione, agent générateur d'ions superoxydes. Ce criblage avait permis d'isoler onze mutants présentant potentiellement une sensibilité accrue au stress superoxyde.

1.2 Identification de dix mutants sensibles au paraquat

Mon travail de thèse a débuté par la confirmation de la sensibilité au stress superoxyde de ces onze mutants. En effet, la sensibilité des mutants à la ménadione pouvait résulter soit des dommages occasionnés par les ions superoxydes dont elle est à l'origine, soit d'une toxicité secondaire de la ménadione. Afin de tester ces hypothèses, le niveau de sensibilité au paraquat, un autre agent générateur d'ions superoxydes, a été estimé pour chacun de ces mutants. Les résultats montrent que tous les mutants, sauf le mutant 16D10, sont sensibles au paraquat (cf article 1 Figure 1). Ainsi, dix des onze mutants sont à la fois sensibles à la ménadione et au paraquat ; ces mutants sont donc sensibles au stress superoxyde.

Les régions flanquant les loci interrompus dans le génome de ces mutants ont été clonées puis séquencées. Les analyses de séquences avaient montré que le plasmide s'était intégré au sein d'une ORF (thèse Annabelle Thibessard, 2002). Les ORF interrompues dans ces mutants sont ainsi impliquées dans la défense contre le stress superoxyde, puisque leur inactivation conduit à une sensibilité à ce stress. Ces gènes ont par la suite été nommés *ossA* à *ossJ*, pour oxydative stress sensitive. Les analyses *in silico* des gènes *oss* ont permis de les classer en cinq catégories sur le critère de leur fonction putative :

- la structure de la paroi et la morphologie cellulaire : les gènes *ossA*, *ossB* et *ossC* codent respectivement des protéines homologues de MreD, RodA et PBP2b. Les mutants de ces gènes présentent des cellules de forme sphérique, celles de la souche

sauvage étant de forme ovoïde. Ces mutants présentent également de plus longues chaînes de cellules par rapport à la souche sauvage.

-la synthèse des exopolysaccharides : le gène *ossD* code la protéine CpsX potentiellement impliquée dans l'export d'exopolysaccharides.

-la maturation des ARNt : les gènes *ossE* et *ossF* codent respectivement des protéines homologues de la tRNA guanine transglycosylase (TGT) et de TrmA. Ces deux protéines modifient des ARNt post-transcriptionnellement.

-le métabolisme du fer : le gène *ossH* code une protéine homologue de la protéine FatD, une perméase d'un ABC transporteur de Fer. Les gènes *ossI* et *ossJ* codent respectivement des protéines homologues de SufD et IscU, deux protéines impliquées dans la biosynthèse et/ou la réparation des centres [Fe-S].

-les gènes de fonction inconnue : le gène *ossG*.

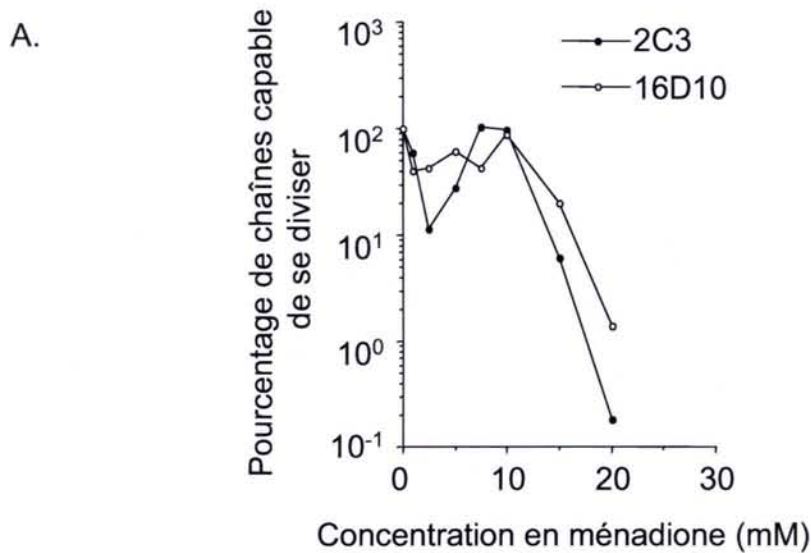
1.3 Niveau de sensibilité du mutant 16D10 au stress oxydant

Lors des tests de sensibilité au paraquat, le mutant 16D10 contrairement aux autres n'a pas révélé de niveau de sensibilité significativement différent de celui de la souche témoin. Ce résultat pouvait signifier que ce mutant est sensible à un effet secondaire de la ménadione. Toutefois, une autre hypothèse pouvait être que ce mutant est sensible, non pas au stress superoxyde, mais au stress peroxyde. Il a en effet été montré qu'un stress superoxyde engendre également un stress peroxyde, le peroxyde d'hydrogène étant produit par dismutation de radicaux superoxydes (Fridovich, 1983). Par conséquent, le niveau de sensibilité à d'autres agents oxydants, ainsi que le niveau de sensibilité à la ménadione, en guise de vérification, a été estimé. Les autres agents testés ont été le cumène hydroperoxyde et le peroxyde d'hydrogène, deux agents générateurs de stress peroxyde.

Les résultats montrent que le mutant présente un niveau de sensibilité à ces différents stress, y compris celui occasionné par la ménadione, qui n'est pas significativement différent de celui de la souche témoin (Figure 22). Ainsi, bien que le mutant 16D10 ait été sélectionné pour sa sensibilité à la ménadione lors du criblage de la banque de mutants, aucune des conditions de stress testées n'a permis de révéler de sensibilité au stress oxydant. Trois hypothèses peuvent expliquer ces résultats : le mutant 16D10 est un « faux positif » et n'est pas impliqué dans la défense contre le stress oxydant, les conditions expérimentales n'ont pas permis de révéler sa sensibilité, ou alors au cours des sous-clonages de la souche mutante, un mutant suppresseur du phénotype de sensibilité au stress superoxyde a été sélectionné.

1.4 Le mutant 16D10 présente de longues chaînes

Lors de l'étude du comportement du mutant 16D10 vis-à-vis du stress oxydant, les souches bactériennes ont été cultivées en milieu liquide non agité. Une observation au microscope photonique à contraste de phase a révélé que les chaînes de cellules de la souche mutante sont plus longues que celles de la souche sauvage (Figure 23). Ce résultat suggérait que le locus interrompu dans le génome de ce mutant est impliqué dans la ségrégation cellulaire. Or le génome de *S. thermophilus* ne présente pas de gène homologue des gènes *lytA* et *lytB* de *S. pneumoniae* et du gène *acmA* de *L. lactis*, gènes impliqués dans la ségrégation cellulaire



B.

Souches	Cumène hydroperoxyde 0,53M	Peroxyde d'hydrogène		
		10M	1M	0,1M
2C3	2,9	3,7	2,8	1,2
16D10	3,0	3,9	3,0	1,3

Figure 22 : Niveau de sensibilité du mutant 16D10 à différents stress oxydants

Le niveau de sensibilité de la souche mutante 16D10 à différents stress a été comparé à celui de la souche témoin 2C3. La souche 2C3 présente au sein de son génome une copie du plasmide pGh9:ISS1, plasmide qui avait été utilisé pour effectuer la mutagenèse insertionnelle aléatoire. Il avait été montré que cette souche ne présente pas de comportement différent de celui de la souche sauvage vis-à-vis du stress oxydant. Elle constitue ainsi un témoin négatif de l'influence du plasmide sur le comportement de *S. thermophilus* vis-à-vis du stress oxydant.

A. Estimation de la capacité à se diviser suite à un stress occasionné par la ménadione en milieu liquide.

B. Estimation du niveau d'inhibition de croissance en milieu gélosé, en présence d'agent oxydant, par la méthode de diffusion sur boîte (voir matériels et méthodes). Les valeurs présentées dans le tableau correspondent à des diamètres d'inhibition de croissance exprimés en cm.

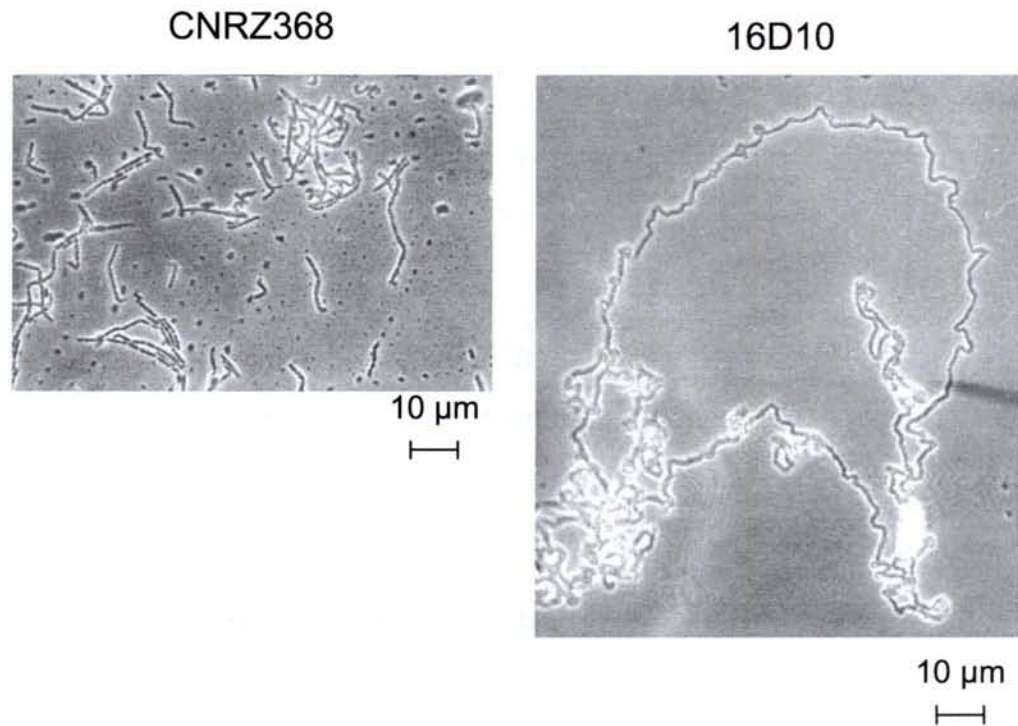


Figure 23 : Phénotype des chaînes de cellules de la souche mutante 16D10 et de la souche sauvage CNRZ368

Les souches bactériennes ont été cultivées en milieu TPPY liquide non agité jusqu'en milieu de phase exponentielle ($DO_{600nm}=0,6$). Les chaînes ont été photographiées en microscopie à contraste de phase à un grossissement de 1000.

chez ces organismes (Sanchez-Puelles *et al.*, 1986; Buist *et al.*, 1995; Garcia *et al.*, 1999). Le mutant 16D10 ouvrait ainsi la possibilité d'identifier un nouveau gène impliqué dans la ségrégation cellulaire.

2 Le gène *cse* est chimérique, variable et mosaïque et code une protéine extracellulaire impliquée dans la ségrégation cellulaire chez *Streptococcus thermophilus*

Une partie des résultats présentés dans ce chapitre est également présentée dans l'article 2 figurant en annexe.

2.1 Le gène *cse* est impliqué dans la ségrégation cellulaire

2.1.1 Le locus interrompu dans le génome du mutant 16D10 est l'ORF *cse* codant potentiellement une protéine de surface

L'analyse de la séquence du locus interrompu au sein du génome du mutant 16D10 avait révélé que le plasmide pGh9:ISS1 interrompt une ORF de fonction inconnue (thèse d'Annabelle Thibessard, 2002). Cette ORF a été nommée *cse* pour cell segrégation.

L'ORF *cse*, de 1386 pb, est précédée d'un site putatif de fixation de ribosomes (Figure 24). Un promoteur constitué d'une boîte -10 étendue et d'une boîte -35 a également été identifié en amont de l'ORF *cse*. Un terminateur de transcription potentiel constitué d'une séquence pouvant adopter une structure tige-boucle et d'une succession de quatre résidus thymine, a été identifié 18 pb en aval du codon stop de l'ORF *cse*. Ces données de séquences suggèrent que l'ORF *cse* pourrait être transcrite de façon monocistronique.

Cette ORF code une protéine de 461 aa présentant potentiellement, à son extrémité N-terminale, un peptide signal clivable, suggérant que Cse serait extracellulaire (Figure 24). L'analyse de la séquence de Cse avec le programme HMMTOP a permis de détecter deux hélices transmembranaires potentielles, la première allant du résidu 71 au résidu 95, et l'autre du résidu 380 au résidu 402. En revanche, lorsque cette même séquence est analysée avec les programmes TMHMM, DAS et PRED-TMR2, aucune hélice transmembranaire n'est prédite. Il est donc probable que la protéine Cse ne présente pas de région transmembranaire.

Une interrogation de la banque de données CDD (Conserved Domain Database), avec le logiciel BLAST, a révélé que la région comprise entre les résidus 53 et 96 s'alignerait de façon fiable (expect=5e-08) avec 95,7 % du domaine LysM. La région qui est comprise entre les résidus 347 et 460 s'alignerait de façon fiable (expect=2e-19) avec 93,7 % du domaine CHAP (pour cysteine, histidine-dependent amidohydrolase/peptidase). Le domaine LysM est un domaine d'attachement à la paroi cellulaire (Steen *et al.*, 2003), et le domaine CHAP pourrait présenter une activité hydrolase du peptidoglycane dans les protéines de surface (Bateman et Rawlings, 2003; Heilmann *et al.*, 2003; Rigden *et al.*, 2003).

Les résultats des recherches d'identités avec les séquences présentes dans les banques de données sont présentés § 2.8.

Ainsi, le gène interrompu au sein du génome du mutant 16D10 a été nommé cse. Il coderait une protéine extracellulaire liée à la paroi cellulaire et impliquée dans l'hydrolyse du peptidoglycane.

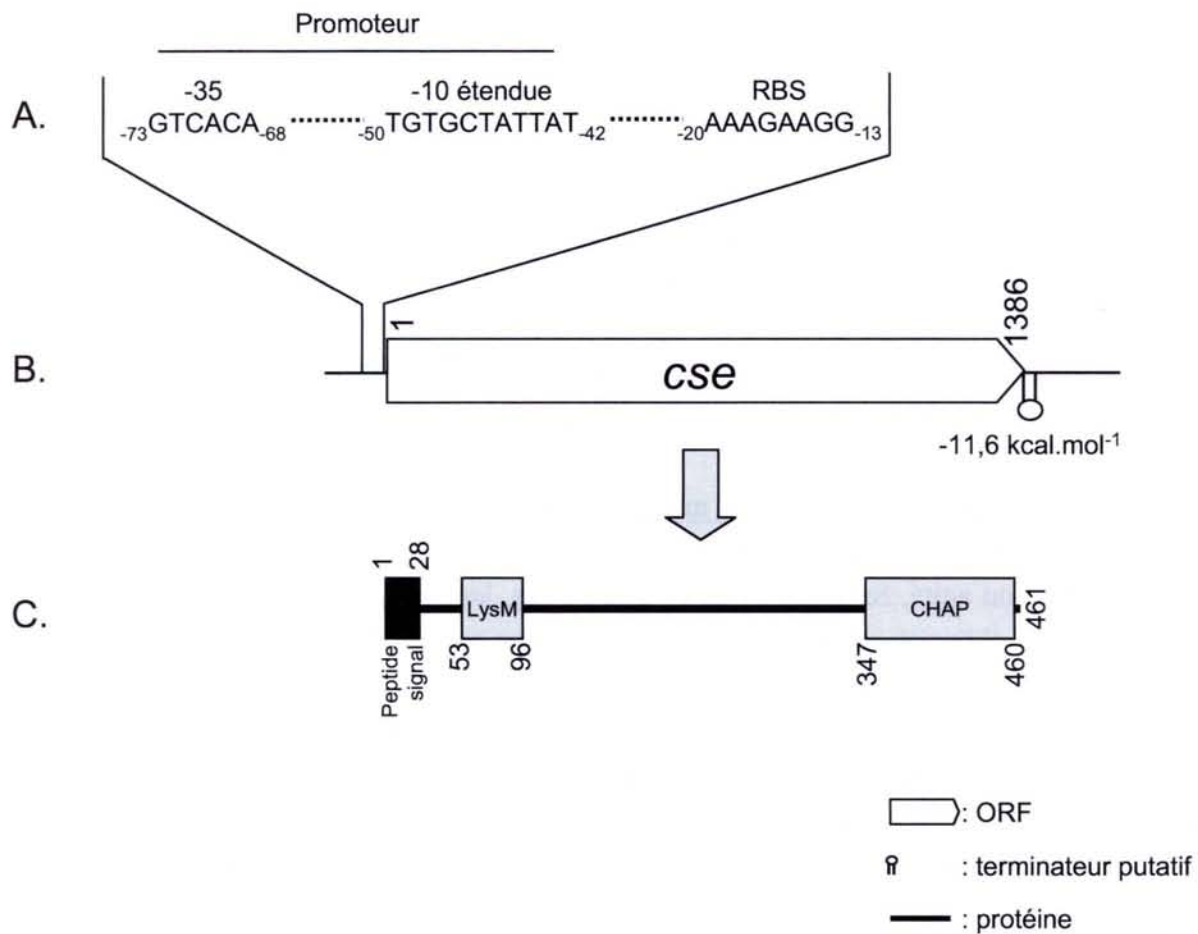


Figure 24 : Représentation schématique du locus *cse* et de la protéine *Cse*

La position 1 de la séquence nucléotidique a été arbitrairement attribuée au premier nucléotide de l'ORF *cse*.

A. Séquence partielle de la région située en amont de l'ORF *cse* contenant le promoteur putatif et le site putatif de fixation du ribosome RBS (ribosome binding site). Les pointillés remplacent des séquences non représentées.

B. Représentation schématique du locus *cse*.

C. Représentation schématique de la protéine *Cse* et de ses domaines putatifs.

2.1.2 Construction d'un mutant présentant une délétion de l'ORF *cse*

Afin de confirmer l'implication de *cse* dans la ségrégation cellulaire, un mutant, présentant une délétion en phase de la quasi totalité de l'ORF *cse*, a été construit chez la souche *S. thermophilus* CNRZ368. La délétion est effectuée en phase dans le but d'inactiver le gène *cse* en minimisant les risques d'affecter, par effet polaire, l'expression des gènes environnants.

Afin de construire ce mutant, deux produits de PCR, flanquant la région de l'ORF *cse* à déléter, ont été clonés dans le plasmide pGh9. La construction résultante, qui a été vérifiée par séquençage, a été introduite dans *S. thermophilus* CNRZ368 afin de sélectionner le remplacement de l'allèle *cse* sauvage par l'allèle mutant au locus chromosomique (voir Matériels et Méthodes). La souche CNRZ368- Δ *cse* sélectionnée présente un allèle mutant constitué des trois premiers et quatre derniers codons de l'ORF *cse*, eux-mêmes flanquant un site de restriction *Eco*RI (Figure 25). La présence de ce site de restriction résulte de la procédure de clonage des fragments de PCR : il a été ajouté dans le but d'augmenter l'efficacité de la réaction de ligation des produits de PCR l'un avec l'autre lors de la construction du plasmide. La séquence de ce site étant GAATTC, celle-ci sera traduite en la séquence Glu-Phe et n'induirait ainsi pas de décalage du cadre de lecture.

2.1.3 Longueur des chaînes du mutant CNRZ368- Δ *cse*

Les souches sauvage CNRZ368 et mutante CNRZ368- Δ *cse* ont été cultivées 20 heures en milieu TPPY non agité. Suite à 20 heures de cultures, les cellules sont en phase stationnaire prolongée. Des observations au microscope photonique ont révélé que les chaînes de cellules de la souche CNRZ368- Δ *cse* sont plus longues que celles de la souche sauvage CNRZ368 (Figure 26 A et B). Le nombre de cellules par chaîne a été compté.

Les résultats de comptabilisation du nombre de cellules par chaîne peuvent être représentés de deux manières. Dans les deux cas, en abscisse est représenté le nombre de cellules par chaîne et en ordonnée est représenté soit le pourcentage de chaînes soit le pourcentage de cellules. La représentation du pourcentage de cellules en ordonnée a été préférée car celle-ci présente l'avantage de rendre compte de l'activité de ségrégation cellulaire. En effet, prenons l'exemple d'une population théorique de 100 chaînes de cellules constituées de deux longueurs différentes : 2 et 100 cellules. Cette population est constituée de 50 chaînes de 2 cellules et de 50 chaînes de 100 cellules, soit 50% de chaînes dans chacune des catégories (Figure 27 A). Cela représente 100 (soit 2%) et 5000 (soit 98%) de cellules incluses respectivement dans des chaînes de 2 et 100 cellules. Deux graphiques ont été représentés à partir de ces données (Figure 27 B et C). Contrairement à la représentation graphique présentée Figure 27 B, celle présentée Figure 27 C rend compte du nombre de liens intercellulaires non rompus, et donc de l'activité de ségrégation cellulaire. En effet, une chaîne de deux cellules présente un unique lien intercellulaire, tandis qu'une chaîne de 100 cellules en présente 99.

Les souches de *S. thermophilus* ont été cultivées sans agitation, afin d'éviter la cassure des chaînes. Les cellules ont été ensuite suspendues délicatement par inversion des tubes. Puis, le nombre de cellules par chaîne a été comptabilisé par observation au microscope photonique à contraste de phase.

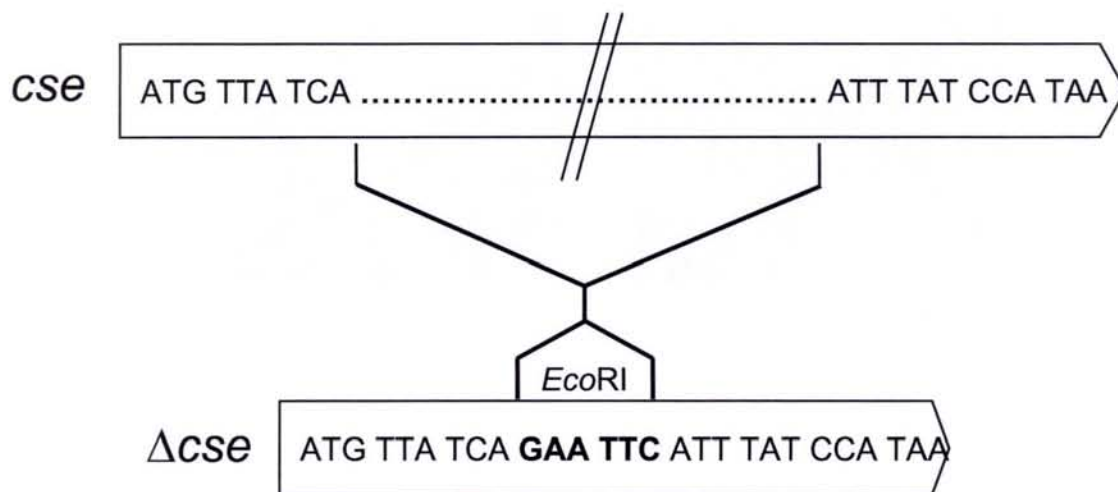


Figure 25 : Représentation schématique de la délétion en phase de la quasi totalité de l'ORF *cse*

Les flèches représentent les ORF, les deux traits obliques indiquent que l'ORF est partiellement représentée. Les pointillés remplacent des séquences non représentées

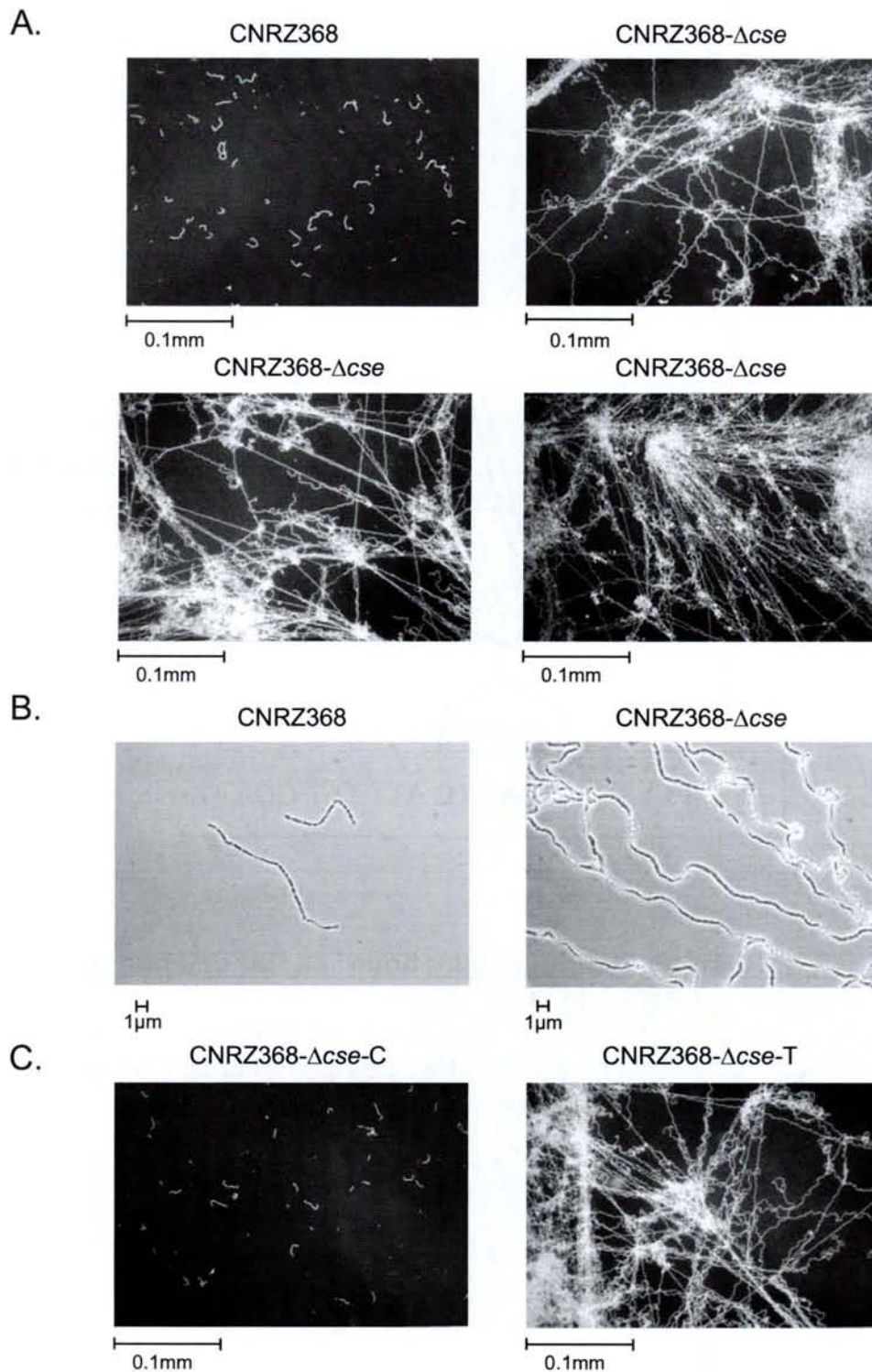


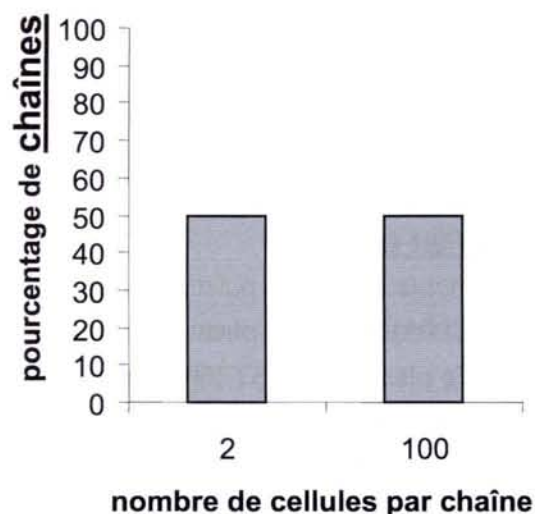
Figure 26 : Phénotype des chaînes de cellules de la souche sauvage, de la souche mutante CNRZ368- Δ cse et de celles en dérivant, en milieu TPPY

Les souches bactériennes ont été cultivées en milieu TPPY liquide non agité pendant 20h à 42°C. Les chaînes ont été photographiées à des grossissements de 100 (A. et C.) et de 1000 (B.).

A.

Nombre de cellules par chaîne	Nombre de chaînes	Nombre de cellules	Pourcentage de chaînes	Pourcentage de cellules
2	50	100	50	2
100	50	5000	50	98

B.



C.

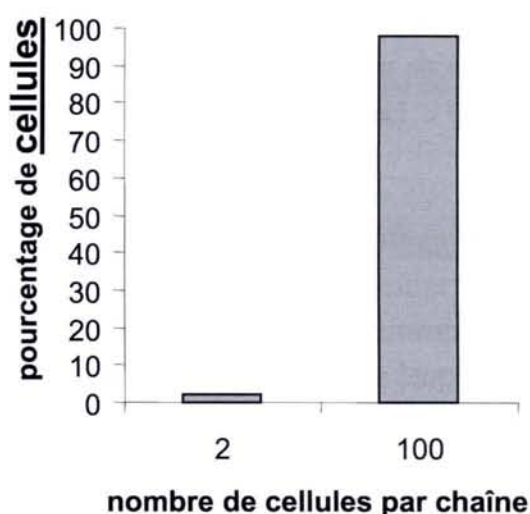


Figure 27 : Exemple théorique de représentations de résultats issus de la comptabilisation du nombre de cellules par chaîne au sein d'une population

A. Tableau rassemblant les valeurs relatives à une comptabilisation théorique de 100 chaînes de cellules.

B. Graphique représentant le pourcentage de chaînes en fonction du nombre de cellules par chaîne.

C. Graphique représentant le pourcentage de cellules en fonction du nombre de cellules par chaîne.

Les résultats montrent qu'en phase stationnaire, 15 % des chaînes de la souche sauvage CNRZ368 présentent plus de 100 cellules par chaîne, alors que 100 % des chaînes de la souche mutante CNRZ368- Δcse se situent dans cette même catégorie (Figure 28 A). Il est à noter que les chaînes de la souche mutante étant fortement emmêlées (Figure 26 A) le nombre de cellules par chaîne n'est généralement que partiellement comptabilisable. C'est pourquoi la dernière catégorie de chaînes présentées Figure 28 regroupe les chaînes présentant un nombre de cellules supérieur à 100. Toutefois, le plus grand fragment de chaîne mutante ayant été comptabilisé est constitué de 440 cellules, alors qu'aucune chaîne de la souche sauvage ne présente plus de 140 cellules.

Le nombre de cellules par chaîne a également été compté en phase exponentielle (Figure 28 B). Cette comptabilisation a révélé que la délétion de *cse* entraîne également un phénotype de longues chaînes en phase exponentielle de croissance, montrant que le phénotype est observé quel que soit la phase de croissance.

Le phénotype de la souche mutante a également été analysé en milieu lait ainsi qu'en milieu M17 (Figure 29 et 30). La souche CNRZ368- Δcse présente également un phénotype de chaînes longues, suggérant que l'expression du phénotype n'est pas dépendante du milieu de culture.

L'observation des cellules à un grossissement de 1000 en microscopie à contraste de phase n'a pas révélé de modifications majeures de la morphologie cellulaire suite à la délétion de *cse* (Figure 31). Le gène *cse* ne jouerait ainsi pas de rôle majeur dans la morphogénèse cellulaire.

2.1.4 Complémentation par ajout d'une copie sauvage de *cse*

Le vecteur utilisé pour réaliser l'expérience de complémentation est le plasmide intégratif pNST260+, construit récemment au laboratoire (Pavlovic, 2004). C'est un plasmide dérivé de pGh9, dans lequel a été cloné le locus *int-attI* de ICES σ 1. Ce plasmide pNST260+ est capable d'intégrer le chromosome de la souche *S. thermophilus* CNRZ368 au site *attR* d'ICES σ 1, un élément intégratif résident (Burrus *et al.*, 2000). Le gène *int* qu'il porte, code une intégrase putative catalysant des événements de recombinaison site-spécifique entre le site *attI*, porté par le plasmide, et le site *attR* chromosomique.

L'ORF *cse*, ainsi que son promoteur putatif et son terminateur potentiel, ont été clonés dans le plasmide pNST260+. La construction résultante pNST260+::*cse* a été introduite dans *S. thermophilus*, et des événements d'intégrations dans le chromosome ont ensuite été sélectionnés. Le plasmide ayant une répllication thermosensible, cette sélection est opérée par culture des souches à 42°C, température à laquelle le plasmide ne peut pas se répliquer de façon autonome, et en présence d'érythromycine. Une expérience analogue a été réalisée en parallèle avec le plasmide pNST260+ sans l'insert, en guise de témoin. La localisation des plasmides au site *attR* d'ICES σ 1 a été contrôlée par hybridation ADN-ADN (résultat non montré). La souche CNRZ368- Δcse présentant le plasmide pNST260+ intégré dans le chromosome a été nommée CNRZ368- Δcse -T, et celle présentant le plasmide pNST260+::*cse* a été nommée CNRZ368- Δcse -C.

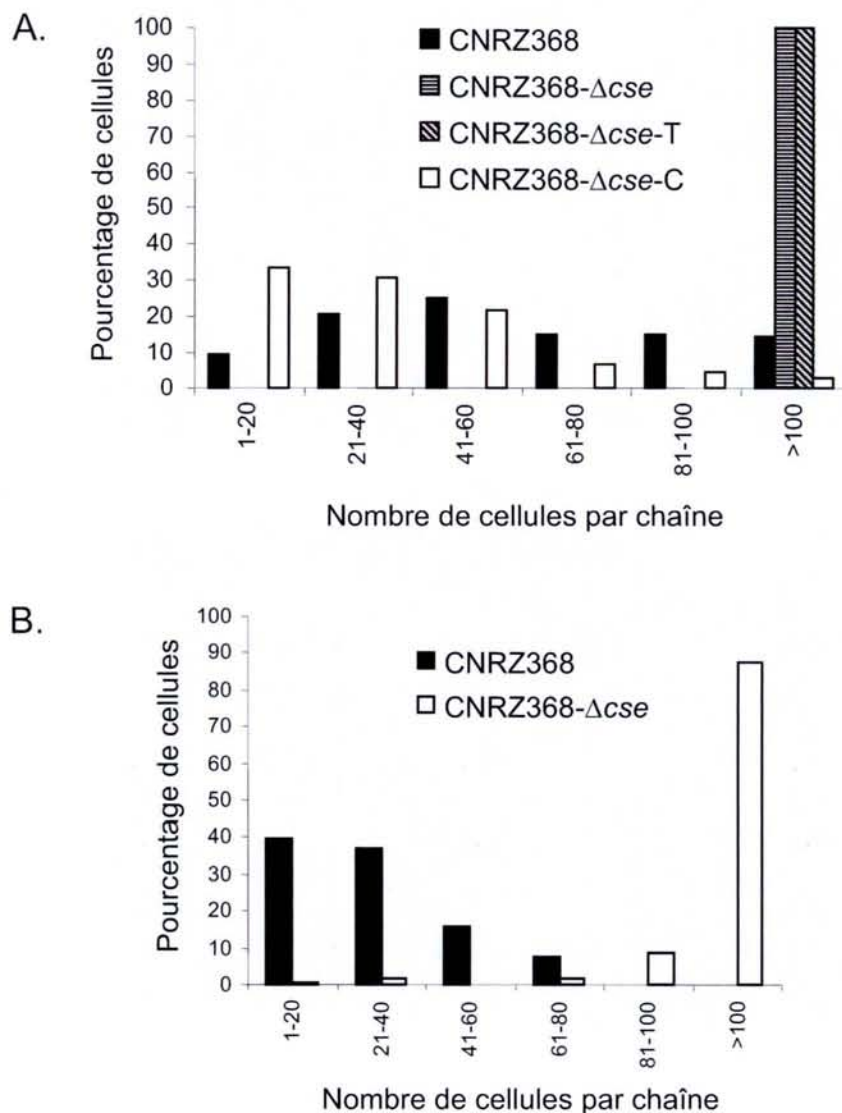
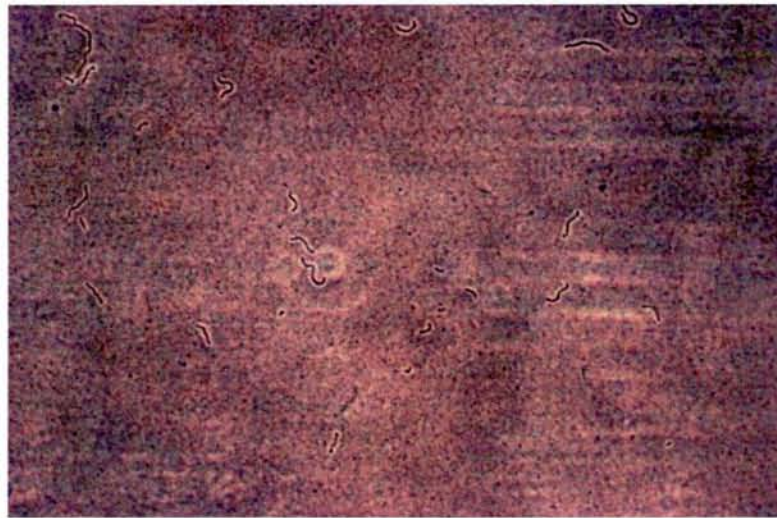


Figure 28 : Graphiques représentant l'influence de *cse* sur la ségrégation cellulaire

Les souches bactériennes ont été cultivées en milieu TPY liquide non agité. Au total, au moins 1000 cellules ont été comptées pour chacune des souches, à partir de trois cultures indépendantes, en phase stationnaire prolongée (20 heures d'incubation)(A.), ainsi qu'en phase exponentielle, à DO=0,2 (B.). Les cellules ont été comptées par observation au microscope photonique à contraste de phase à un grossissement de 1000.

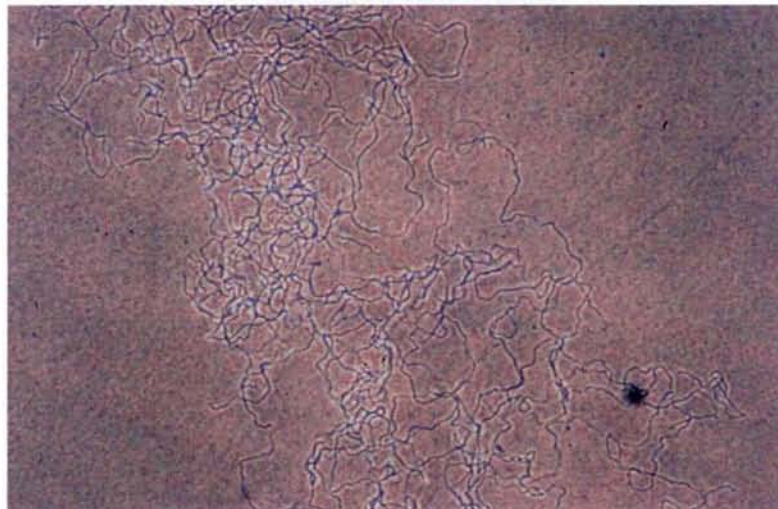
A.

CNRZ368



H
10µm

CNRZ368-Δcse



H
10µm

B.

CNRZ368

CNRZ368-Δcse



H
1µm



H
1µm

Figure 29 : Phénotype des chaînes de cellules de la souche sauvage et de la souche mutante CNRZ368-Δcse en milieu lait

Les chaînes ont été photographiées en microscopie à contraste de phase à des grossissements de 200 (A.) et de 1000 (B.), en phase exponentielle de croissance (pH = 5,7).

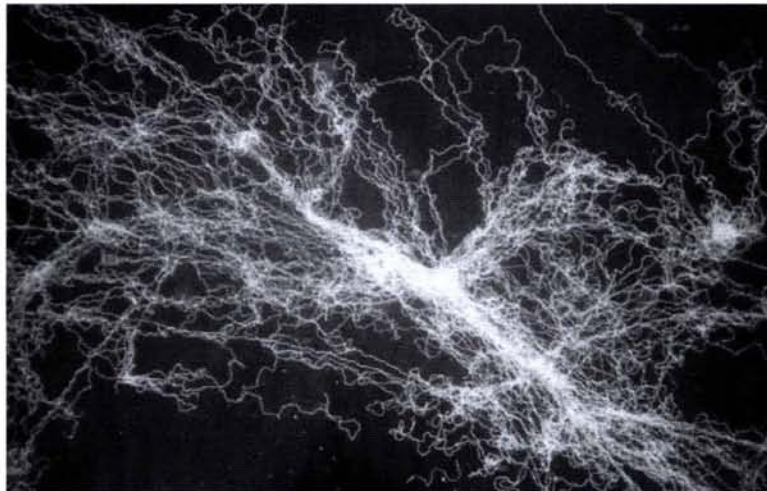
A.

CNRZ368



0.1mm

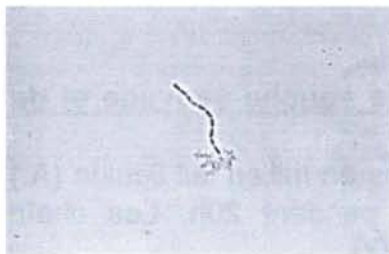
CNRZ368- Δ cse



0.1mm

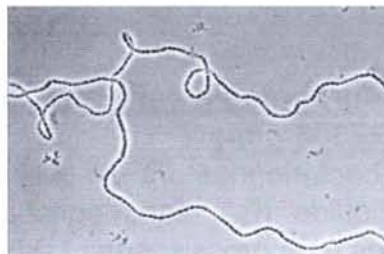
B.

CNRZ368



H
1μm

CNRZ368- Δ cse



H
1μm

Figure 30 : Phénotype des chaînes de cellules de la souche sauvage et de la souche mutante CNRZ368- Δ cse en milieu M17

Les chaînes ont été photographiées en microscopie à contraste de phase à des grossissements de 100 (A.) et de 1000 (B.), en phase stationnaire de croissance (20 h de culture).

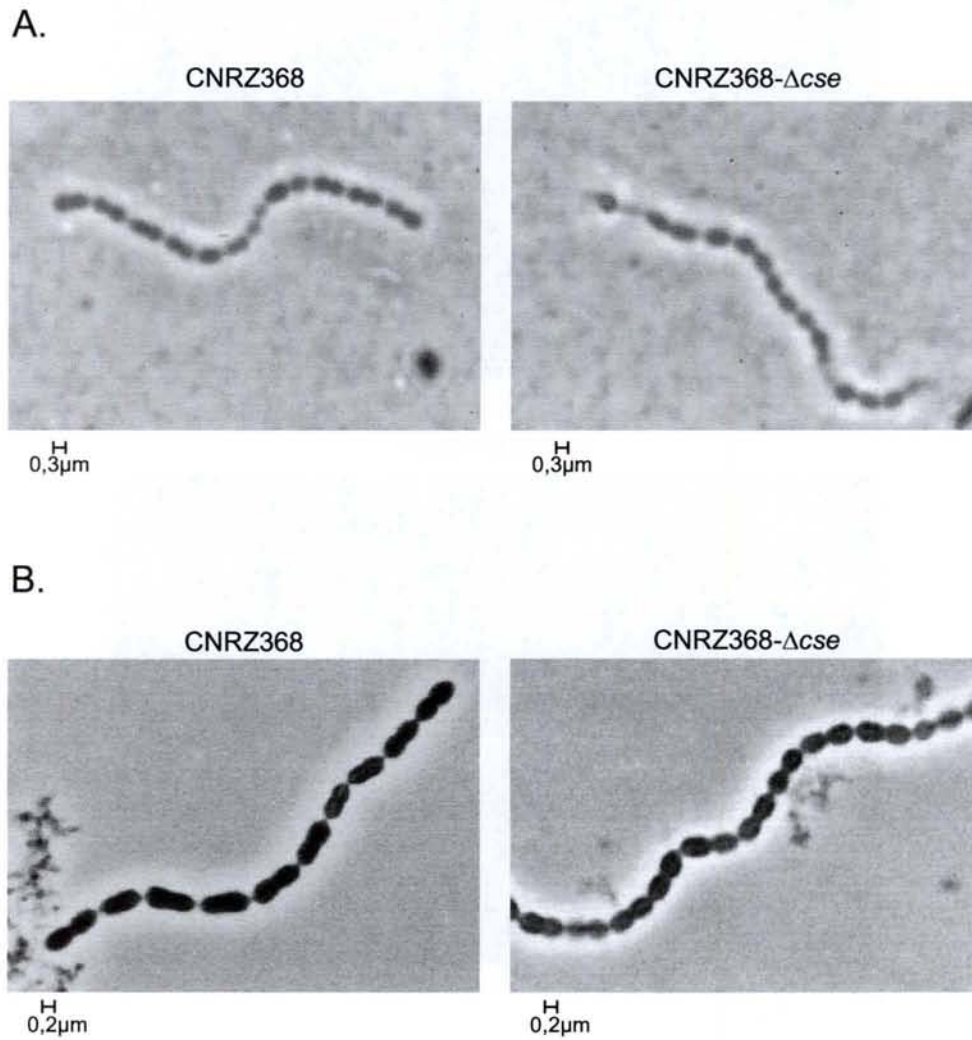


Figure 31 : Morphologie cellulaire de la souche sauvage et de la souche mutante CNRZ368-Δcse

Les souches bactériennes ont été cultivées en milieu lait liquide (A.) ainsi qu'en milieu M17 liquide (B.), sans agitation pendant 20h. Les chaînes ont été photographiées à un grossissement de 1000.

2.1.5 Complémentation du phénotype de chaînes longues de la souche CNRZ368- Δ cse

Les chaînes de cellules de la souche CNRZ368- Δ cse-C ont été observées au microscope photonique (Figure 26 C) et leur nombre de cellules a été compté (Figure 28 A). Les résultats montrent que l'ajout de *cse* dans le génome de la souche CNRZ368- Δ cse permet de restaurer le phénotype sauvage.

L'observation des chaînes de cellules (Figure 26 C), ainsi que le comptage du nombre de cellules par chaîne de la souche témoin CNRZ368- Δ cse-T (Figure 28 A), n'ont pas révélé de différence par rapport à la souche CNRZ368- Δ cse, montrant, comme attendu, que le plasmide pNST260+ n'a pas d'influence sur la longueur des chaînes de cellules.

Par conséquent, l'ensemble de ces résultats montre que le gène cse est impliqué dans la ségrégation cellulaire chez S. thermophilus.

2.2 Influence de cse sur la croissance en milieu liquide et en milieu gélosé

La croissance des souches mutante CNRZ368- Δ cse et sauvage a été suivie en fonction du temps. Cultivées en milieu TPPY, le temps de doublement en phase exponentielle des souches sauvage et mutante est respectivement de 18 min et de 22 min. Ce résultat montre que la délétion de *cse* a peu ou pas d'influence sur la capacité des cellules à se diviser. La souche sauvage CNRZ368 et la souche mutante CNRZ368- Δ cse entrent en phase stationnaire respectivement à une valeur de DO_{600nm} de 0,66 +/- 0,074 et de 0,50 +/- 0,029. Il semblerait que la souche mutante entre en phase stationnaire à une valeur de DO_{600nm} légèrement inférieure à celle de la souche sauvage.

L'observation des flacons de culture a révélé que la souche mutante se caractérise par une floculation beaucoup plus marquée comparativement à la souche sauvage (Figure 32 A). Cette floculation est suivie d'une sédimentation précoce des cellules de la souche mutante, comparativement à la souche sauvage (Figure 32 B).

Les différentes souches ont été cultivées 20 heures, en aérobiose, en milieu TPPY gélosé. L'observation des colonies a révélé que celles de la souche mutante CNRZ368- Δ cse ont une morphologie différente de celles de la souche sauvage CNRZ368 (Figure 33). En effet, les colonies de la souche mutante sont plus grandes et moins opaques que celles de la souche sauvage. Ce résultat est confirmé par l'expérience de complémentation.

2.3 Influence de cse sur l'adhésion des cellules à une surface non biotique

Il a été montré que la longueur des chaînes pouvait avoir une influence sur la capacité des lactocoques et des streptocoques à adhérer à des surfaces abiotiques (Mercier *et al.*, 2002). Plus les chaînes de cellules sont longues et plus la capacité à adhérer diminue.

La capacité à adhérer à une surface en verre a été évaluée pour la souche sauvage et les souches mutantes de *S. thermophilus* (collaboration avec S. Kulakauskas, URLGA, INRA Jouy-en-Josas). Des cellules de chacune de ces souches suspendues en milieu M17 ont été incubées 4 h à 20°C dans une boîte de Pétri contenant une lame de verre. Les lames ont ensuite été lavées avec de l'eau stérile afin de retirer les cellules n'ayant pas adhéré à la surface de la lame. Les cellules ont ensuite été colorées à l'acridine orange, un agent

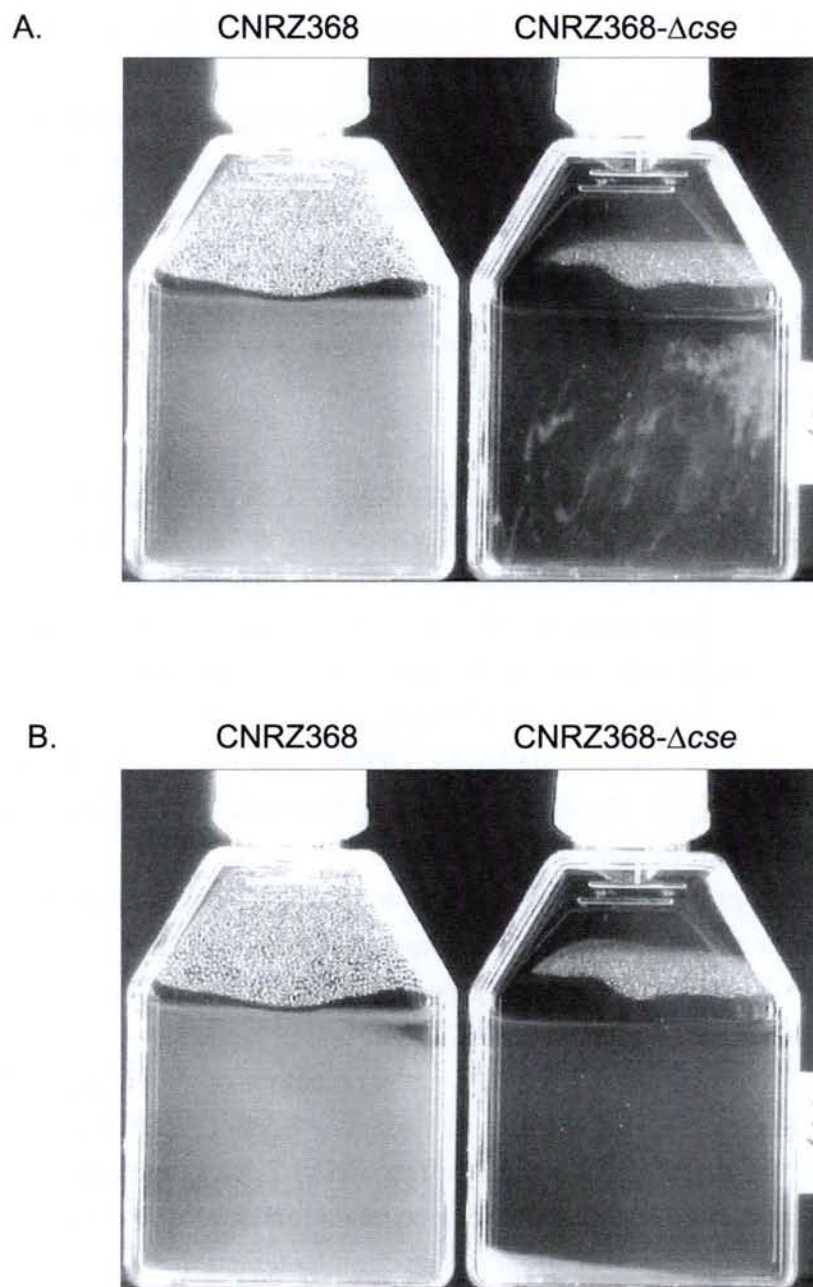


Figure 32 : Croissance en milieu liquide de la souche sauvage et de la souche mutante CNRZ368- Δ cse

Photographies de flacons de culture non agité en milieu M17 liquide de souches de *S. thermophilus* à 10 h (A.) et 15 h (B.) d'incubation à 42°C. L'ensemencement a été effectué à la surface du milieu de culture.

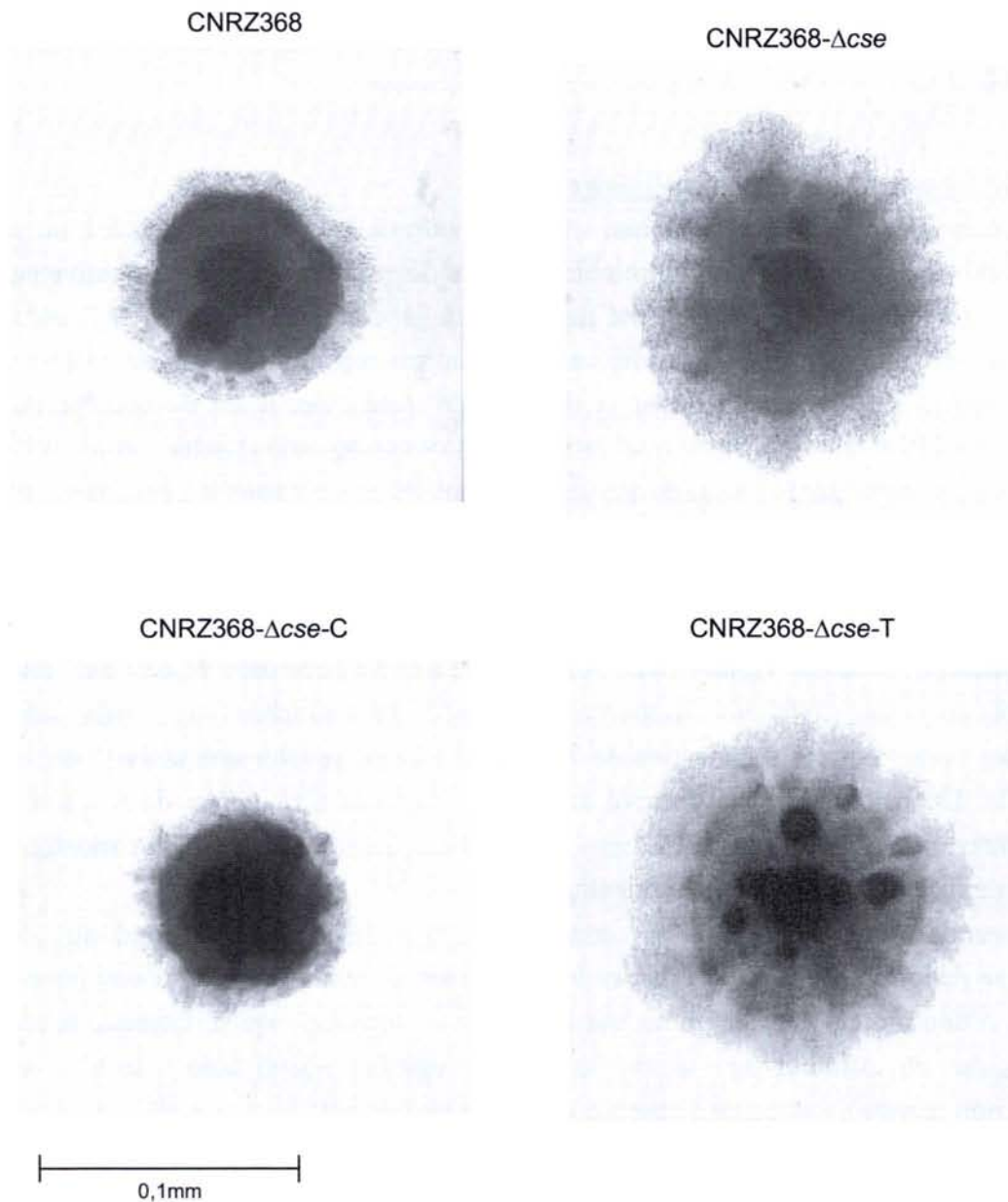


Figure 33 : Morphologie des colonies de la souche sauvage, de la souche mutante CNRZ368- Δ cse et de celles en dérivant

Les souches bactériennes ont été cultivées sur milieu TPY gélosé, en aérobiose, à 42°C pendant 20 heures. Les colonies ont été photographiées à un grossissement de 100 au microscope photonique.

intercalant, et les lames ont été photographiées au microscope à fluorescence. Le pouvoir adhésif est estimé en pourcentage de surface recouverte par les cellules.

Les résultats montrent que la souche CNRZ368- Δcse présente un pouvoir adhésif significativement inférieur à celui de la souche sauvage (Figure 34). Ce résultat a été confirmé par l'expérience de complémentation.

Le gène cse de S. thermophilus est ainsi impliqué dans la ségrégation cellulaire. Son absence entraîne également une modification de la morphologie coloniale, ainsi qu'une diminution de la capacité à adhérer à une surface en verre.

2.4 Cse est une protéine extracellulaire

Les analyses de séquence suggèrent que Cse pourrait être exportée grâce à un peptide signal localisé à son extrémité N-terminale. Afin de déterminer si Cse est effectivement une protéine extracellulaire, l'ORF *cse* a été fusionnée au gène rapporteur $\Delta_{sp}nuc$ de *S. aureus*. Ce gène code une nucléase dont l'activité enzymatique est repérable en milieu TBD-agar. Ce milieu contient de l'ADN, de l'agar et du bleu de Toluidine. Il est de couleur bleue en présence d'ADN intact, et vire au rose lorsque l'ADN est dégradé (Lachica *et al.*, 1971). Le gène $\Delta_{sp}nuc$ est porté par le plasmide autorépliquatif pFUN et n'est sous la dépendance d'aucun promoteur (Poquet *et al.*, 1998).

L'intégralité de l'ORF *cse* excepté son codon stop, ainsi que la région contenant son promoteur putatif situé en amont, ont été clonées dans le plasmide pFUN afin de générer une fusion traductionnelle *cse- $\Delta_{sp}nuc$* . Cette construction a été clonée dans *E. coli* puis introduite dans *S. thermophilus* CNRZ368. Après avoir cultivé les deux souches transformées sur milieu gélosé, les boîtes ont été recouvertes de TBD-agar afin de révéler une activité nucléasique éventuelle. Un halo rose a été observé autour des colonies d'*E. coli* et de *S. thermophilus* transformées par le plasmide pFUN::*cse* (résultat non montré). Ce résultat montre que la protéine de fusion Cse- Δ_{sp} Nuc est extracellulaire.

Le plasmide pFUN:: $\Delta_{sp}cse$ a été construit en parallèle : il correspond au plasmide pFUN::*cse* dépourvu de la région codant le peptide signal putatif de Cse. Il doit permettre la synthèse d'une protéine de fusion ne possédant pas de signal d'export. Lorsque le plasmide pFUN:: $\Delta_{sp}cse$ est introduit dans *E. coli* ou *S. thermophilus*, aucun halo rose n'est observé (résultat non montré), indiquant l'absence de nucléase extracellulaire. Ce résultat montre que la région du gène *cse* codant le peptide signal putatif est nécessaire à la localisation extracellulaire de la protéine de fusion Cse- Δ_{sp} Nuc.

L'ensemble de ces résultats montre que Cse est une protéine extracellulaire, et que sa localisation extracellulaire est conférée par son peptide signal.

2.5 La protéine Cse est-elle une hydrolase du peptidoglycane ?

2.5.1 Hypothèse

Les analyses de séquences indiquent que la protéine Cse présente en C-terminal un domaine CHAP. Or le domaine CHAP de la protéine Aae de *Staphylococcus epidermidis*

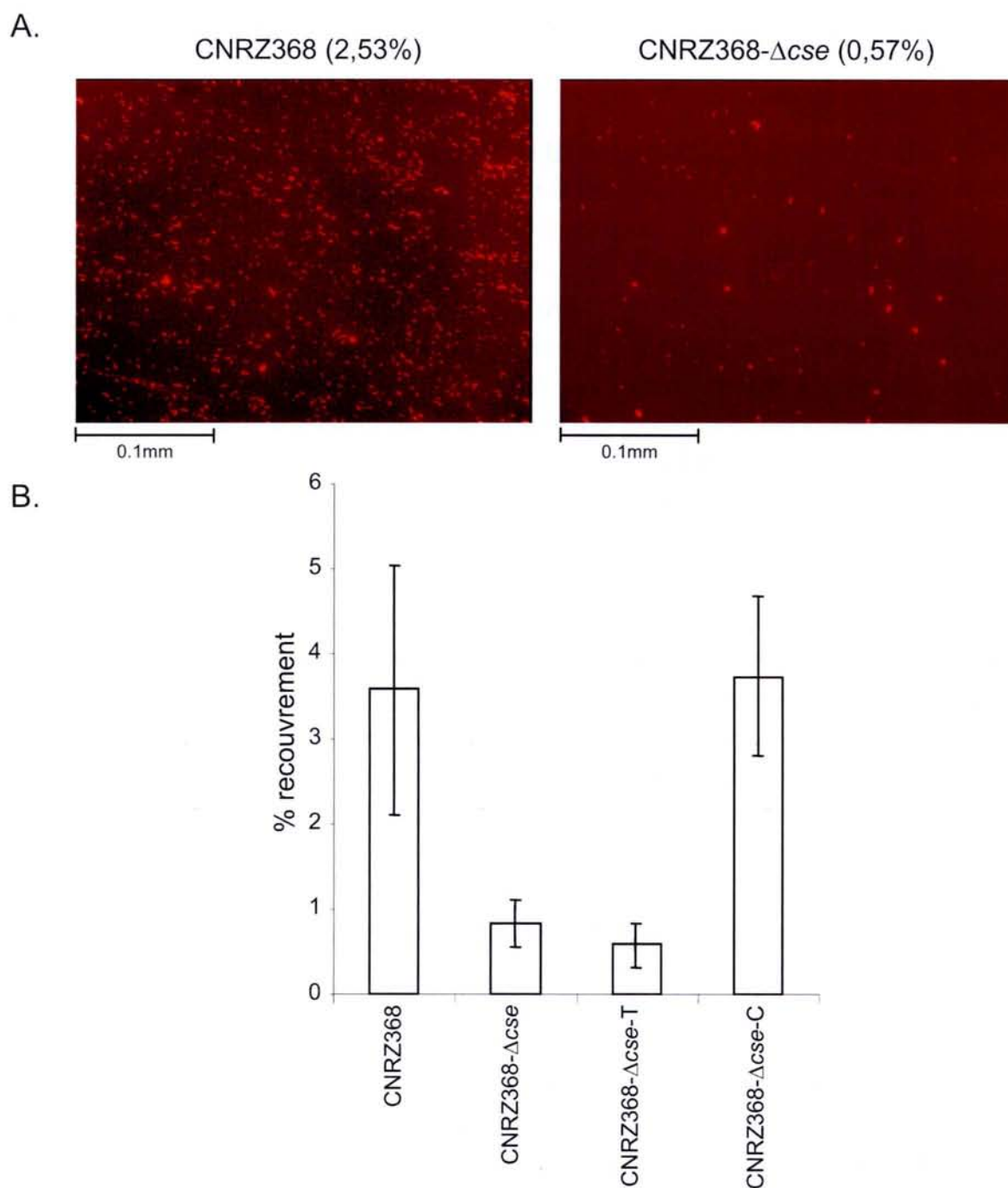


Figure 34 : Influence de *cse* sur le pouvoir adhésif des chaînes de cellules sur une surface en verre

Les lames de verre ont été photographiées en microscopie à fluorescence (A.) afin de calculer le pourcentage de la surface recouverte par les cellules à l'aide de l'analyseur d'image IMAGETOOL (B.). Trois expériences indépendantes ont été réalisées pour chacune des souches bactériennes. Pour chaque expérience d'adhésion, le pourcentage de recouvrement a été calculé à partir de 20 clichés photographiques.

présente une activité hydrolase du peptidoglycane (Heilmann *et al.*, 2003). Par ailleurs, il est décrit dans la littérature que la ségrégation cellulaire résulte d'activité hydrolase de la paroi détectable par zymographie chez d'autres bactéries (Wuenscher *et al.*, 1993; Buist *et al.*, 1995; Ishikawa *et al.*, 1998; Qin *et al.*, 1998; Ohnishi *et al.*, 1999; Carroll *et al.*, 2003). Il est donc raisonnable d'envisager que la protéine Cse présente une activité hydrolase du peptidoglycane.

2.5.2 Activité lytique d'extraits protéiques bruts de *S. thermophilus*

Cette hypothèse a été testée par zymographie, méthode permettant de visualiser l'activité enzymatique d'une protéine après migration électrophorétique. Les protéines extraites des souches CNRZ368 et CNRZ368- Δcse ont été séparées par électrophorèse en gel de polyacrylamide en présence de SDS (SDS-PAGE). Le gel de polyacrylamide contient des cellules autoclavées de *S. thermophilus* qui rendent le gel opaque. Lors de l'étape de renaturation, qui suit l'électrophorèse, si une hydrolase du peptidoglycane se renature de façon native, elle pourra potentiellement entraîner la lyse des cellules incluses dans le gel et ainsi provoquer un éclaircissement local du gel.

L'activité d'une protéine d'environ 30 kDa est observable dans la piste correspondant à la souche CNRZ368, sous la forme d'une zone d'éclaircissement (Figure 35). Une zone d'éclaircissement correspondant à une protéine de taille équivalente est observable dans la piste correspondant à la souche CNRZ368- Δcse (Figure 35). Un test d'activité similaire a été réalisé avec des cellules de *Micrococcus lysodeikticus* comme substrat à la place de celles de *S. thermophilus*. Dans ce cas, aucune bande de lyse n'a été observée (résultat non montré). Ainsi, aucune différence n'a été observée en zymographie entre la souche sauvage et la souche Δcse . La quantité de la protéine Cse pouvant être insuffisante pour que son activité hydrolase du peptidoglycane soit détectée, il a été envisagé de surproduire la protéine Cse dans *E. coli*.

2.5.3 Activité lytique de fragments de Cse surproduits chez *E. coli*

L'ORF *cse* a été clonée dans le plasmide pET15b, afin de créer une fusion traductionnelle avec une queue hexa-histidine localisée à l'extrémité N-terminale de la protéine Cse. Le plasmide résultant a été nommé pET15b::*cse*. Après 4 h d'induction de l'expression du gène de fusion avec de l'IPTG, les protéines ont été extraites puis séparées par SDS-PAGE. La comparaison du profil protéique de la souche portant le plasmide pET15b::*cse* avec celui de la souche portant le plasmide sans insert n'a pas révélé de différence, indiquant que la protéine Cse est peu ou pas surproduite chez *E. coli* (résultat non montré).

Des résultats similaires ont été obtenus avec la protéine Acmb de *L. lactis* (Huard *et al.*, 2003). La protéine entière Acmb n'étant pas surproduite chez *E. coli*, les auteurs ont surproduit des fragments de cette protéine. Une approche similaire a été envisagée pour surproduire Cse. L'ORF *cse* a été clonée en deux fragments : un fragment codant la région de la protéine Cse comprise entre les résidus d'acides aminés 1 à 183 (comprenant le domaine LysM) et un autre fragment codant la région allant des résidus d'acides aminés 184 à 461 (comprenant le domaine CHAP) (Figure 36). Ces deux fragments sont produits efficacement chez *E. coli* (Figure 35 A).

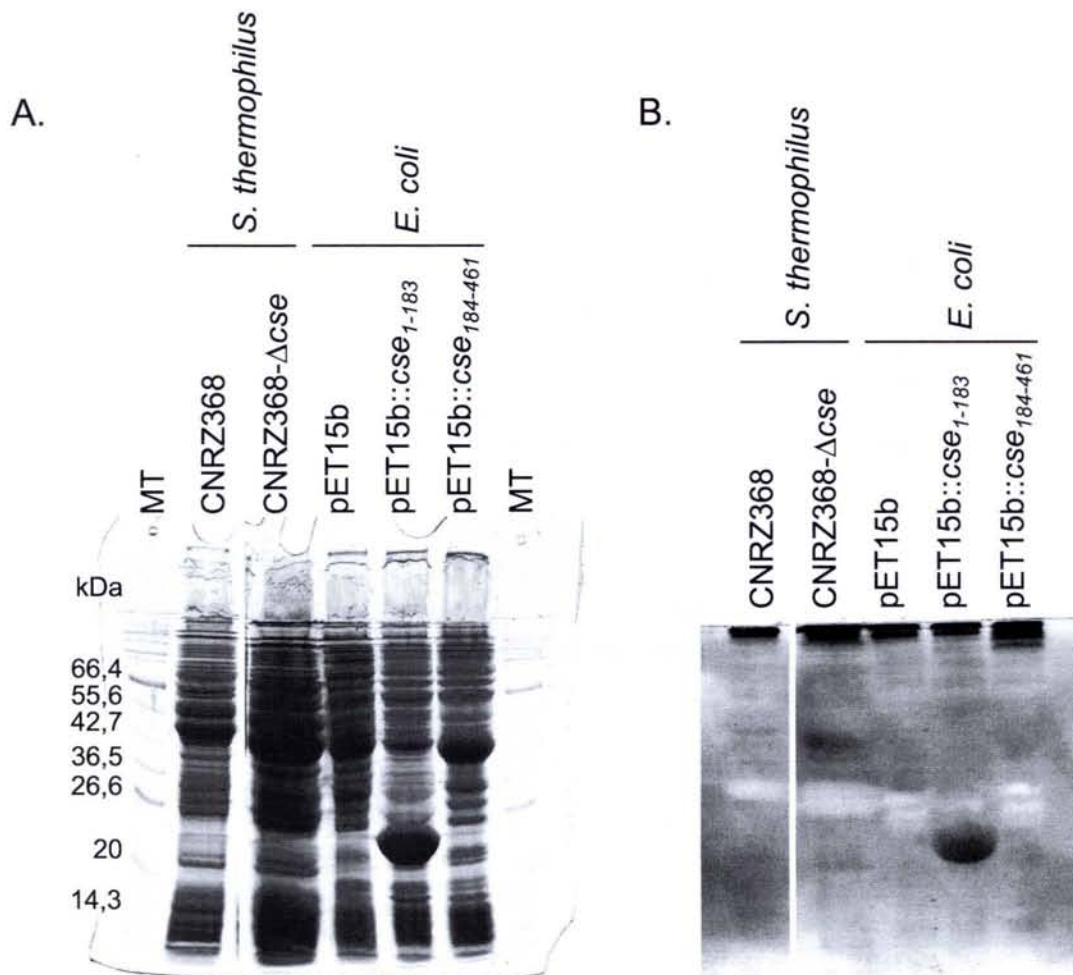


Figure 35 : Test d'activité hydrolase de la paroi par zymographie

A.et B. SDS-PAGE d'extraits protéiques bruts de souches de *S. thermophilus* et de clones recombinants de *E. coli* BL21(DE3). Ces deux SDS-PAGE ont été réalisés en parallèle dans la même cuve.

A. Les protéines ont été colorées par du bleu de coomassie.

B. Le gel de polyacrylamide contient des cellules autoclavées de *S. thermophilus*. Suite à une étape de renaturation des protéines, les cellules ont été colorées au bleu de méthylène. Les zones d'éclaircissement résultent de la lyse des cellules de *S. thermophilus*.

MT : marqueur de taille.

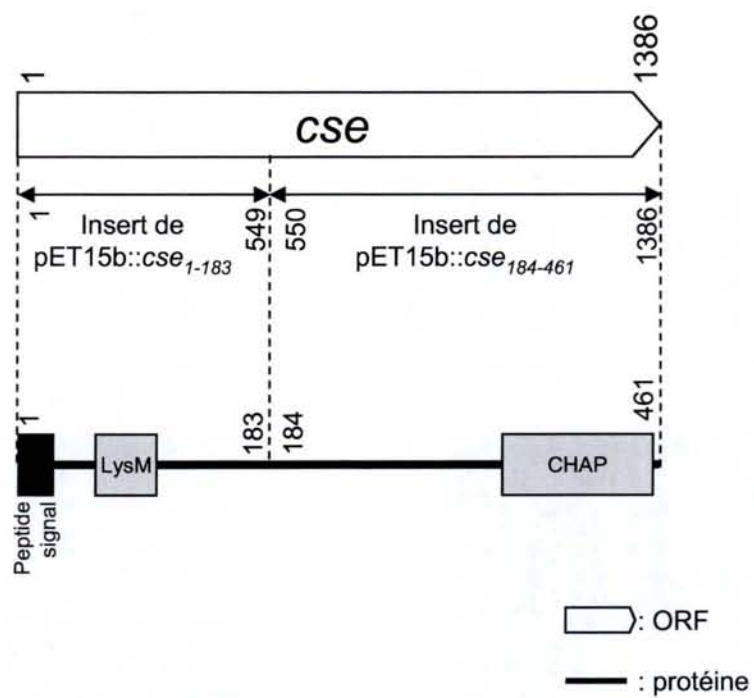


Figure 36 : Représentation schématique des fragments de l'ORF *cse* clonés dans le vecteur de surexpression pET15b

Deux zymogrammes ont été réalisés avec de extraits bruts des clones surproduisant les fragments de Cse : un avec des cellules de *M. lysodeikticus* et un autre avec des cellules de comme substrat *S. thermophilus*. Aucune bande de lyse n'a été observée avec les cellules de *M. lysodeikticus* (résultat non montré). Lorsque les cellules de *S. thermophilus* sont utilisées comme substrat, trois bandes de lyse, résultant de l'activité de protéines d'environ 23, 26 et 30 kDa, sont observables dans les pistes correspondant aux souches portant le plasmide pET15b sans insert et le plasmide pET15b::*cse*₁₈₄₋₄₆₁ (Figure 35 B.). Ces bandes de lyse résultent vraisemblablement d'activités d'hydrolyse de la paroi portées par des protéines de *E. coli*. Ces bandes ne sont pas observées dans la piste correspondant à la souche portant le plasmide pET15b::*cse*₁₋₁₈₃. Il est probable que l'absence de ces bandes soit due à un défaut de migration des protéines (Figure 35 A.). Ainsi, aucune bande de lyse attribuable à un fragment de la protéine Cse n'a été observée dans les deux cas (Figure 35 B.).

2.5.4 Complémentation avec *cse* d'un phénotype de longues chaînes résultant d'une déficience en une hydrolase du peptidoglycane

Une autre approche, cette fois-ci indirecte, a été envisagée (thèse de Séverine Layec). Elle visait à déterminer si la protéine Cse était capable de compléter un phénotype de longues chaînes résultant d'une déficience en une hydrolase du peptidoglycane, tel que le mutant $\Delta acmA$ de *L. lactis*. La protéine AcmA étant une hydrolase du peptidoglycane responsable de la ségrégation cellulaire chez *L. lactis* (Buist *et al.*, 1995), le mutant $\Delta acmA$ présente de longues chaînes de cellules, à l'instar du mutant Δcse .

Le mutant $\Delta acmA$ a été transformé avec le plasmide pNST260+::*cse*, plasmide précédemment utilisé pour compléter le phénotype de longues chaînes du mutant Δcse de *S. thermophilus*. Aucune différence de longueur de chaîne n'a été observée en présence du plasmide pNST260+::*cse* (résultats non montré), suggérant que la présence de Cse de *S. thermophilus* n'entraîne pas l'hydrolyse du peptidoglycane de la souche *L. lactis* $\Delta acmA$.

2.6 La région *var-cse* du gène *cse* est variable et riche en répétitions

2.6.1 Mise en évidence de la variabilité de la région *var-cse*

La séquence du gène *cse* de la souche *S. thermophilus* CNRZ368 a été alignée avec celle de la souche *S. thermophilus* LMG18311 (séquence fournie par P. Hols, UCL Belgique). Les 600 premières et les 346 dernières paires de bases des deux allèles *cse* présentent respectivement 94 % et 96 % d'identité. En revanche, leur région centrale ne présente que 61 % d'identité (Figure 37). Cette région a été nommée *var-cse*.

La région centrale du gène *cse* de 18 souches de *S. thermophilus* a été amplifiée par PCR. Les oligonucléotides choisis recouvrent des régions qui, d'une part flanquent la zone de divergence, et d'autre part sont parfaitement conservées entre les souches CNRZ368 et LMG18311. La comparaison des profils électrophorétiques révèle une variabilité importante de la taille des produits de PCR obtenus (pour exemple, Figure 38). Les produits de PCR de 16 souches ont été séquencés, la séquence de la souche LMG18311 nous a été fournie par P. Hols, celle de la souche LMD-9 était accessible sur le site http://genome.jgi-psf.org/draft_microbes/strth/strth.home.html. Les séquences ont été alignées à l'aide du logiciel CLUSTALW. L'alignement de ces séquences révèle d'une part un polymorphisme de

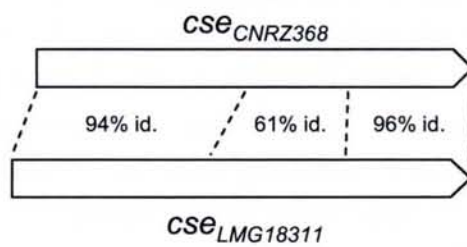


Figure 37 : Hétérogénéité du niveau d'identité entre les allèles *cse* des souches *S. thermophilus* CNRZ368 et LMG18311

Les ORF sont représentées par des flèches. Trois régions, délimitées par des traits interrompus fins, sont discernables sur le critère d'identité (id.).

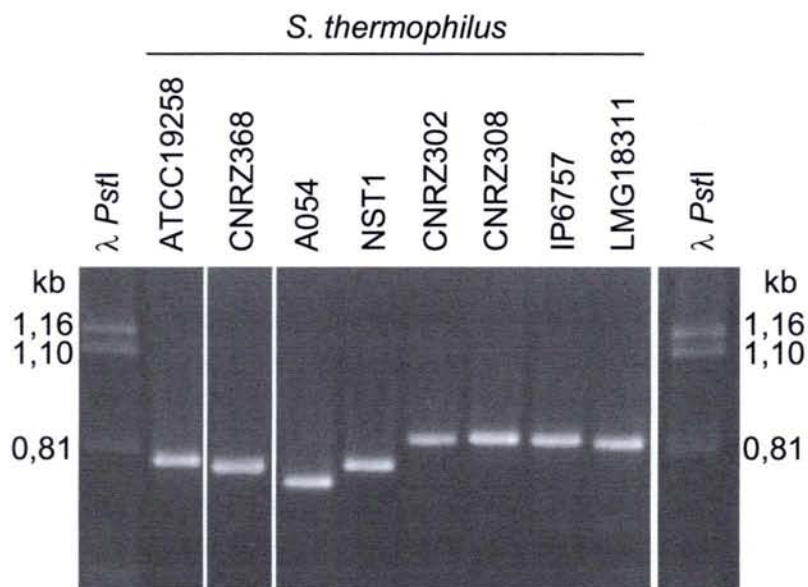


Figure 38 : Profils électrophorétiques des produits d'amplification de la région centrale du gène cse

La région centrale du gène *cse* de différentes souches de *S. thermophilus* a été amplifiée par PCR à l'aide des amorces Forvar (5'-CTGTAGTAGCAGAATCTAAC-3') et Revvar2 (5'-GCACTAGCAATCCAGTCTT-3'). L'ADN du bactériophage λ digéré par l'enzyme de restriction *Pst*I est utilisé comme marqueur de taille.

taille puisque les séquences ont une taille variant de 386 à 542 pb, soit une variation de 156 pb, et d'autre part un polymorphisme de séquence important (figure 39).

En conclusion, la région *var-cse* du gène *cse* arbore un haut niveau de variabilité à l'échelle intraspécifique.

2.6.2 La région *var-cse* est riche en motifs répétés

Une simple analyse visuelle de la séquence de la protéine Cse des souches CNRZ368 et LMG18311 a permis de constater que leur région *var-cse* est riche en motifs répétés (pour exemple, Figure 40). Ce caractère répétitif a été analysé plus avant.

La séquence des allèles *cse*_{CNRZ368} et *cse*_{LMG18311} a été analysée par construction d'un DOT PLOT, mode d'analyse d'une séquence permettant de détecter des répétitions internes (Figure 41). Le DOT PLOT est une matrice de comparaison de deux séquences (différentes ou identiques). Chacune des séquences à comparer occupe un des axes du graphique. La matrice de comparaison est construite en faisant glisser une fenêtre, de taille définie par l'utilisateur, le long des deux séquences. Lorsque les séquences incluses dans les fenêtres présentent le niveau d'identité défini par le manipulateur, un point est dessiné sur le graphique. Ainsi, lorsqu'une séquence est comparée à elle-même une diagonale est dessinée du fait que les deux séquences comparées sont identiques. Tout point sortant de la diagonale révèle l'existence d'une répétition interne.

Les DOT PLOT des deux allèles *cse* présentent tous les deux des groupes de points de forme carrée dans la région *var-cse*, signifiant que la région *var-cse* de ces allèles présente une région riche en motifs répétés (Figure 41). Une conséquence de cette richesse en motifs répétés est que cette région présente une séquence de faible complexité d'après le logiciel SEG (résultat non montré). Une région est de faible complexité lorsque sa composition en nucléotides est biaisée par rapport à la moyenne globale. La même analyse a ensuite été effectuée avec les autres séquences *var-cse* et montre qu'elles présentent toutes une région riche en motifs répétés dont la séquence est de faible complexité (résultat non montré).

2.6.3 Structure et organisation des motifs répétés de la région *var-cse*

Les séquences nucléiques des différents allèles *var-cse* ont été analysées dans le but de détecter des motifs répétés. Le résultat de l'analyse de la région *var-cse*_{CNRZ368} est présenté en détail à titre d'exemple dans la Figure 40. Cette analyse révèle que les régions *var-cse* sont presque exclusivement constituées de motifs répétés.

Les répétitions sont directes et les unités répétées peuvent aussi bien être disposées en tandem comme de façon dispersée (Figure 40). Dans l'ensemble des séquences analysées, 11 motifs répétés différents ont été identifiés (Figure 42). Les motifs détectés ont tous une taille (en nucléotide) multiple de trois (Figure 42 A) et sont positionnés dans la même phase les uns par rapport aux autres. La conséquence directe en est que chaque motif nucléotidique répété correspond à un motif protéique répété. Il est à noter que la séquence consensus de 8 de ces motifs répétés, nommés A, C, D, E, F, H, I et J, est dégénérée.

L'alignement de la séquence des différents allèles *var-cse* révèle que la divergence entre ces séquences peut être due à la présence de motifs répétés en plus ou en moins (Figure 43). En effet, la séquence *var-cse* des souches CNRZ368, NST1 et CNRZ385 présente 3 motifs D

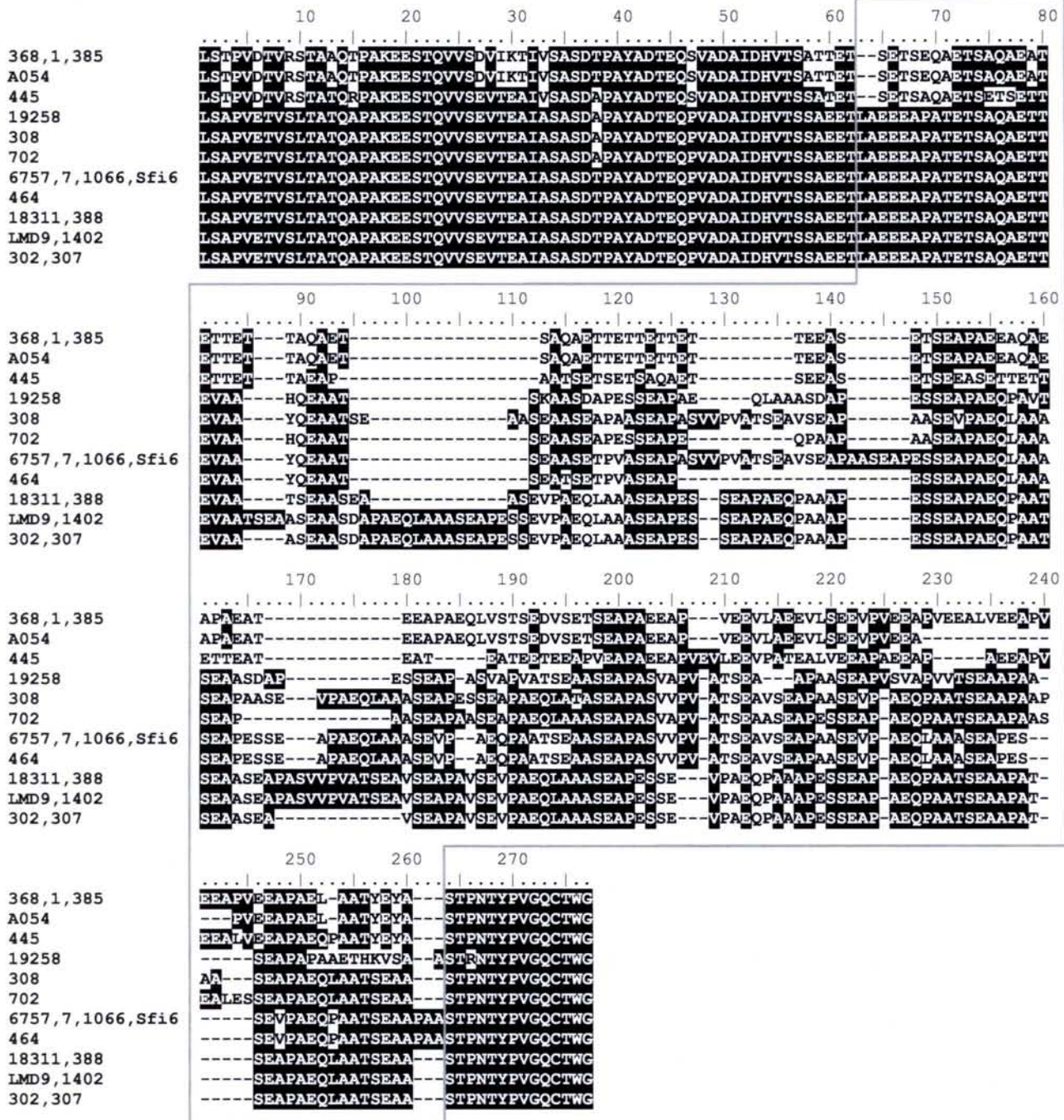


Figure 39 : Alignement des séquences comprenant la région Var-Cse de différentes souches de *S. thermophilus*

Les produits de PCR obtenus avec les amorces Forvar et Revvar2 ont été séquencés, les séquences ont été alignées à l'aide du logiciel CLUSTALW. Les souches CNRZ368 (368), NST1 (1), CNRZ385 (385), A054, CNRZ445 (445), ATCC19258 (19258), CNRZ308 (308), CNRZ702 (702), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, CNRZ464 (464), LMG18311 (18311), CNRZ388 (388), LMD-9, CNRZ1402 (1402), CNRZ302 (302), CNRZ307 (307) ont été analysées. La région Var-Cse est encadrée.

561 TGAACAATCAGTTGCAGATGCTATTGATCATGTTACTTCAGCAACAAC**TGAAACATCTGAAACATCT**GAA
E Q S V A D A I D H V T S A T T **E T S E T S** E

631 CAAGCT**GAAACATCA****GCACAAGCT****GAAGCAACT****GAAACAAC****TGAAACAAC****TGCACAAGCT****GAAACATCT**G
Q A **E T S A Q A** E A T **E T T E T T** A Q A **E T S**

701 CACAAGCT**GAAACAAC****TGAAACAAC****TGAAACAAC****TGAAACAAC****TGAAGAGGCATCA****GAAACATCA****GAAGC**
A Q A **E T T E T T** E E A S **E T S E A**

771 ACCAGCTGAA**GAAGCACAAGCT****GAAGCACCAGCAGAA****GCAACTGAA****GAAGCTCCAGCAGAA****CAACTAGTA**
P A E E **A Q A E A P A E** A T E **E A P A E** Q L V

841 TCTACATCAGAAGATGTATCA**GAAACATCA****GAAGCACCAGCTGAA****GAAGCACCAGTTGAA****GAAGTACTAG**
S T S E D V S **E T S E A P A E** **E A P V E** **E V L**

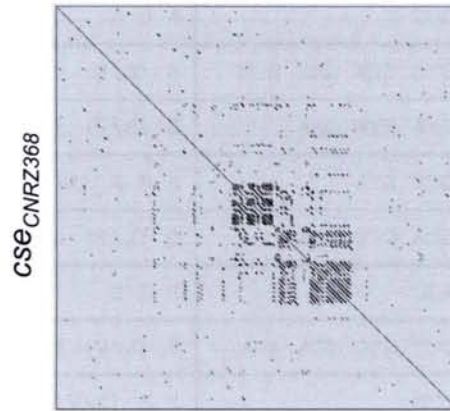
911 CTGAAG**GAAGTACTATCTGAA****GAAGTACCAGTTGAA****GAAGCACCAGTTGAA****GAAGCACTAGTTGAA****GAAGC**
A E **E V L S E** **E V P V E** **E A P V E** **E A L V E** **E A**

981 ACCAGTTGAA**GAAGCACCAGTTGAA****GAAGCACCAGCAGAA****CTAGCAGCTACATACGAATATGCATCAACT**
P V E **E A P V E** **E A P A E** **L A A T Y E Y A S T**

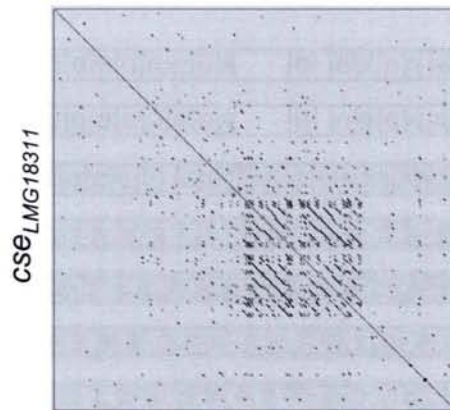
Figure 40 : Contenu en motifs répétés de la région centrale de l'allèle

***cse*_{CNRZ368}**

La séquence protéique est indiquée en-dessous de la séquence nucléique. Les lettres en italique ne font pas partie de la région *var-cse*. Chaque motif répété est encadré, le nom du motif répété est indiqué par une lettre capitale au-dessus de l'encadré.



*CSE*_{CNRZ368}



*CSE*_{LMG18311}

Figure 41 : Détection de la région riche en répétitions des allèles *cse*

Les DOT PLOT ont été construits par comparaison d'une séquence avec elle-même. La fenêtre de comparaison a été fixée à une longueur de 9 nucléotides, et la tolérance envers la divergence a été fixée à 1 nucléotide.

A.

nom	Consensus des motifs nucléotidiques ^a	Consensus des motifs protéiques
A	GAA ACA WCW	E T (S/T)
B	GCA CAA GCT	A Q A
C	TCA GAD GYD CCW GMD RDM	S (E/D) (A/V) P (A/E) (E/S/V)
D	GAA GYW CYA KYW GAA	E (A/V) (P/L) (A/V/S) E
E	TCA GAA GCT GYA	S E A (A/V)
F	CAA CYA GCA GCD RCA	Q (L/P) A A (A/T)
G	CCA GAG AGC	P E S
H	GTA GYA CCT GTC GCA ACA	V (A/V) P V A T
I	CCR GCT DYW	P A (T/V/A/S)
J	GAA GMA ACT	E (A/E) T
K	GAA GAG GCA TCA	E E A S

^a W : A/T ; R : A/G ; Y : C/T ; M : A/C ; K : G/T ; D : A/G/T

B.

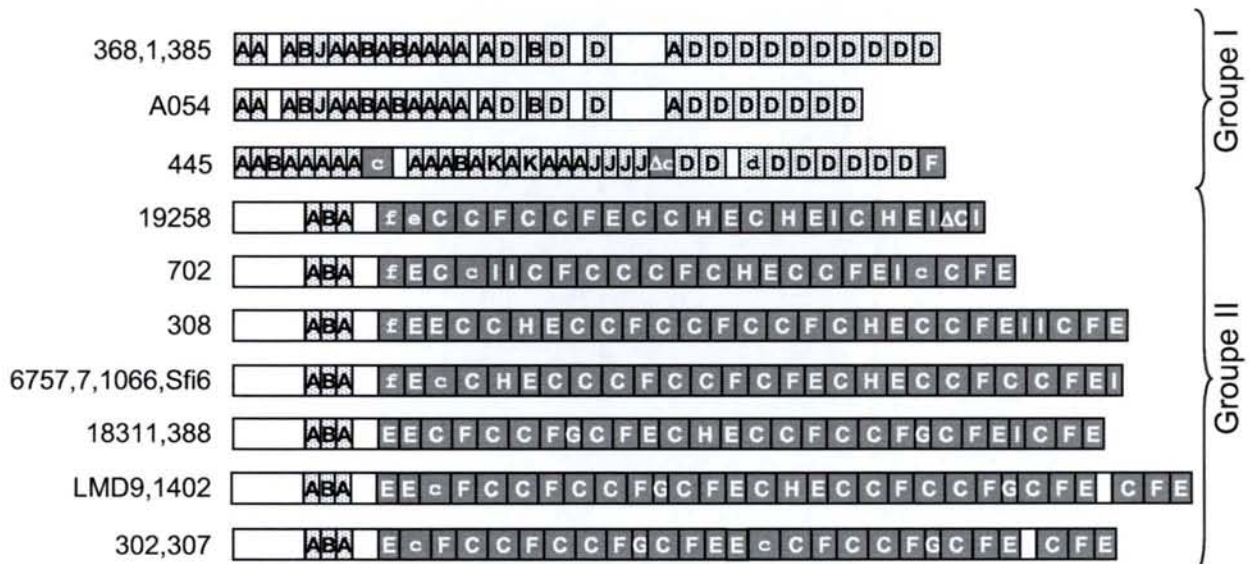


Figure 42 : Contenu en motifs répétés des allèles *var-cse*

A. Tableau répertoriant la séquence consensus des différents types de motifs répétés identifiés dans la séquence des allèles *var-cse* et de leur produit.

B. Représentation schématique des régions *var-cse* des souches CNRZ368 (368), NST1 (1), CNRZ385 (385), A054, CNRZ445 (445), ATCC19258 (19258), CNRZ702 (702), CNRZ308 (308), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, LMG18311 (18311), CNRZ388 (388), LMD-9, CNRZ1402 (1402), CNRZ302 (302) et CNRZ307 (307). Les motifs répétés sont représentés sous la forme de rectangles contenant le nom de l'unité répétée. Les unités répétées de séquence conforme à la séquence consensus sont en lettres majuscules, celles qui ne le sont pas, sont en lettres minuscules. Les motifs caractéristiques du groupe I sont représentés par des rectangles gris clair et ceux du groupe II en gris foncé.

1) D D D D

368.1.385 TACTATCTGAAGAAGTACCAGTTGAAGAAGCACCAGTTGAAGAAGCACTAGTTGAAGAAGCACCAGTTGA
 V L S E E V P V E E A P V E E A L V E E A P V E
 A054 TACTATCTGAAGAAGTACCAGTTGAAGAAGCACCAGTTGAA
 V L S E E V P V E E A P V E

D D

368.1.385 AGAAGCACCAGTTGAAGAAGCACCAGCAGAATAGCAGCTACATACGAATATGCATCAACTCCTAATACA
 E A P V E E A P A E L A A T Y E Y A S T P N T
 A054 GAAGCACCAGCAGAATAGCAGCTACATACGAATATGCATCAACTCCTAATACA
 E A P A E L A A T Y E Y A S T P N T

2) A B A E

308 ACTAAACATCTECACAAGCTAAACAACTAAAGTAGCAGCATACCAAGAAGCAGCAACATCAGAAGCTG
 T E T S A Q A E T T E V A A Y Q E A A T S E A
 6757.7.1066.Sfi6 ACTAAACATCTECACAAGCTAAACAACTAAAGTAGCAGCATACCAAGAAGCAGCAACATCAGAAGCTG
 T E T S A Q A E T T E V A A Y Q E A A T S E A

E C C H

308 CATCAGAAGCTGCATCAGAACACCAGCTGCATCAGAAGCTCCTGCTAGCTAGTACCTGTGCGCAACATC
 A S E A A S E A P A A S E A P A S V V P V A T S
 6757.7.1066.Sfi6 CATCAGAAGCTGCATCAGAACACCAGCTGCATCAGAAGCTCCTGCTAGCTAGTACCTGTGCGCAACATC
 A S E T P V A S E A P A S V V P V A T S

E C C C

308 AGAAGCTGTATCAGAAGCACCAGCTGCA
 E A V S E A P A A TCAGAAGTGCCAGCAGAACAACTA
 6757.7.1066.Sfi6 AGAAGCTGTATCAGAAGCACCAGCTGCATCAGAAGCACCAGAGAGTCAGAAGTGCCAGCAGAACAACTA
 E A V S E A P A A S E A P E S S E A P A E Q L

3) C F E C

LMD9.1402 AGCTCAGAAGCGCCAGCAGAACAACCAGCAGCAACATCAGAAGCTGCATCAGAAGCTCCTGCTAGCGTAG
 S S E A P A E Q P A A T S E A A S E A P A S V
 302.307 AGCTCAGAAGCGCCAGCAGAACAACCAGCAGCAACATCAGAAGCTGCATCAGAAGCT
 S S E A P A E Q P A A T S E A A S E A

H E C C F

LMD9.1402 TACCTGTGCAACATCAGAAGCTGTATCAGAAGCACCAGCTGTATCAGAAGTTGCCAGCAGAACAACTAGC
 V P V A T S E A V S E A P A V S E V P A E Q L A
 302.307 GTATCAGAAGCACCAGCTGTATCAGAAGTTGCCAGCAGAACAACTAGC
 V S E A P A V S E V P A E Q L A

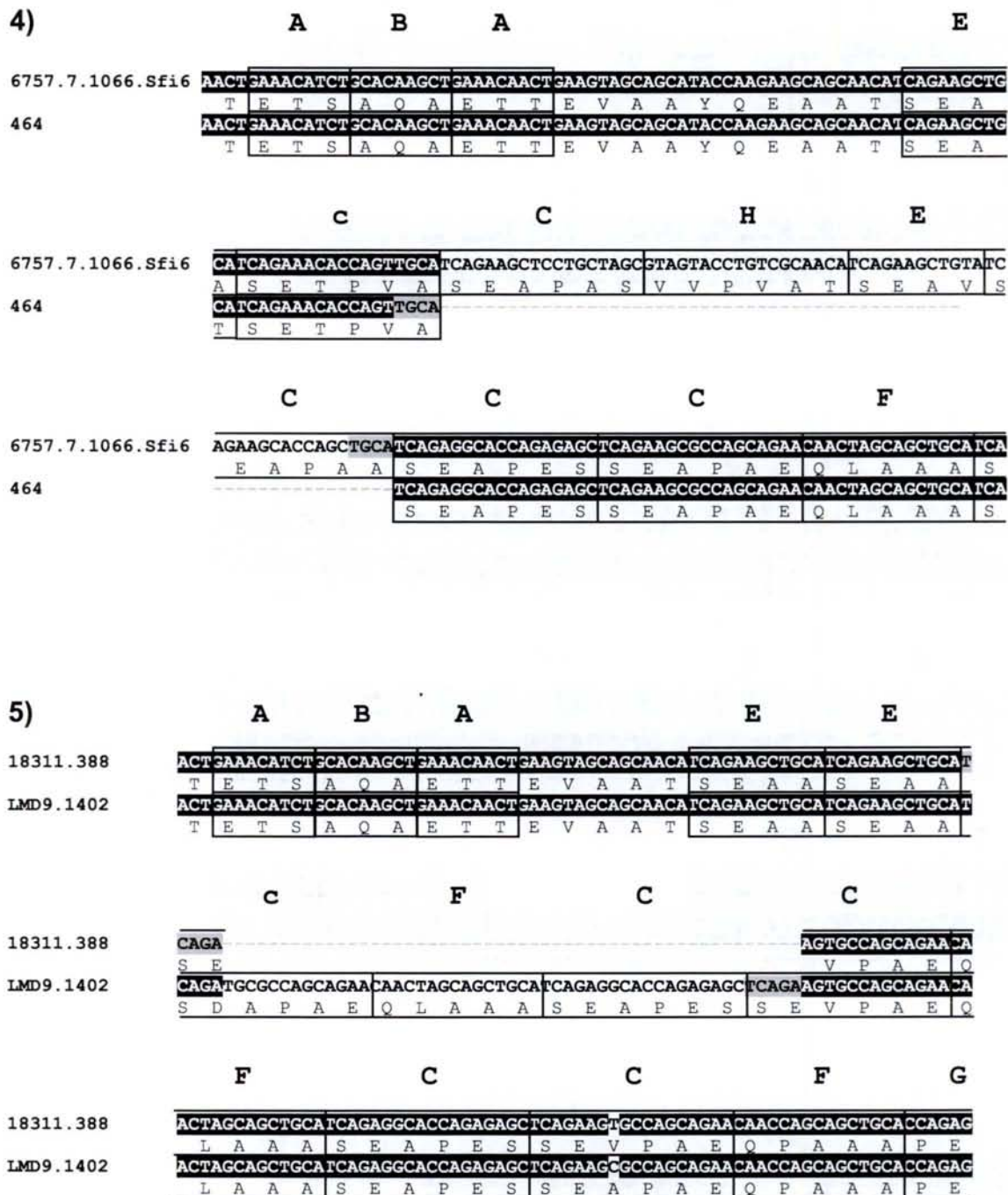


Figure 43 : Mise en évidence d'insertions et de délétions de motifs répétés

Les séquences nucléotidiques des régions *var-cse* des différentes souches de *S. thermophilus* ont été alignées grâce au logiciel CLUSTALW. Seuls les fragments d'alignements fiables et permettant de visualiser une divergence caractéristique d'événements d'insertion et/ou de délétion d'ADN sont présentés. Dans chacun des cas, deux alignements sont possibles concernant le positionnement des trous. Un seul des deux alignements possibles est ici représenté, les séquences devant être alignées dans l'autre alignement possible sont surlignées en gris. Les résultats obtenus avec les séquences *var-cse* des souches CNRZ368 (368), NST1 (1), CNRZ385 (385), A054, CNRZ308 (308), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, LMD-9, CNRZ1402 (1402), CNRZ302 (302), CNRZ307 (307), CNRZ464 (464), LMG18311 (18311), CNRZ388 (388) sont ainsi présentés. Les séquences nucléotidiques identiques sont surlignées en noir. En-dessous de chaque séquence nucléotidique est présentée la séquence protéique correspondante. Les motifs répétés sont encadrés et leur nom est indiqué au-dessus de l'encadré.

en plus par rapport à celle de la souche A054. De façon analogue, la séquence *var-cse* de la souche CNRZ308 présente un motif E en plus et un motif C en moins par rapport à celle des souches IP6757, CNRZ7, CNRZ1066 et Sfi6. D'autres cas similaires sont également présentés à la Figure 43.

Les séquences des motifs ont été comparées entre elles (Figure 44). La séquence codant l'exemplaire SEAPAA du motif C présente une région identique, à un nucléotide près, à la séquence codant l'exemplaire PAA du motif I. Egalement, la séquence codant l'exemplaire SEAPAE du motif C présente également une région identique, à un nucléotide près, à celle codant l'exemplaire EAPAE du motif D. En conclusion, certains motifs répétés sont similaires entre eux.

Les unités répétées peuvent être constituées d'une succession de petits motifs répétés. La séquence *var-cse* des souches LMG18311 et CNRZ388 présente une duplication interne de 140 pb quasi parfaite puisqu'il n'y a qu'un nucléotide de divergence (Figure 45 A). Cette séquence contient la succession de motifs CFCCFGCFE (Figure 45 B). Cette situation a également été détectée dans la séquence *var-cse* d'autres souches de *S. thermophilus* (Figure 45 B).

Il doit être noté que malgré la variabilité de la région Var-Cse, la composition en résidus d'acides aminés est conservée entre les allèles. En effet, tous les allèles sont riches en résidus alanine, glutamate, proline, sérine et thréonine (Tableau 1).

En conclusion, la région var-cse est quasi exclusivement constituée de motifs répétés, quel que soit l'allèle analysé. De nombreux motifs répétés différents ont été identifiés. De plus la comparaison des allèles entre eux montre qu'ils peuvent différer les uns des autres par la présence ou l'absence de l'intégralité de motifs répétés.

2.7 cse est un gène mosaïque

Une simple observation visuelle des DOT PLOT obtenus avec les séquences *var-cse* montre que les souches peuvent être classées en deux groupes. Le groupe I comprend les allèles des souches CNRZ368, NST1, CNRZ385, A054 et CNRZ445, et le groupe II comprend les allèles des souches ATCC19258, CNRZ702, CNRZ308, IP6757, CNRZ7, CNRZ1066, Sfi6, LMG18311, CNRZ388, LMD9, CNRZ1402, CNRZ302 et CNRZ307. Le DOT PLOT des allèles *var-cse* du groupe I présente deux carrés distincts et à proximité l'un de l'autre alors que celui du groupe II n'en présente qu'un (pour exemple, Figure 41). Ce résultat montre que l'agencement des motifs répétés le long de la séquence est différent dans ces deux groupes d'allèles.

Sur le critère de composition en motifs répétés, les séquences des différents allèles *var-cse* peuvent également être classées en ces deux mêmes groupes (Figure 42 B). Le groupe I contient 100 % des motifs J et K, 74 % des motifs A et 59 % des motifs B. Le groupe II quant lui contient 100 % des motifs E, G, H et I et 98 % des motifs C et F. Ainsi, les séquences du groupe I ont qualitativement une composition en motifs répétés différente de celles du groupe II.

Nom du motif	Séquence en acides aminés	Séquence en nucléotides
C:	SE PAA	TCAGAAGCA CCAGCTGCA
I:	PAA	CCAGCTGCT
C:	SE PES	TCAGAGGCA CCAGAGAGC
G:	PES	CCAGAGAGC
C:	SEAPA	TCAGAAGCT CCTGCTAGC
E:	SEAA	TCAGAAGCT GCA
E:	SEAV	TCAGAAGCT GTA
C:	SE APAE	TCA GAAGCGCCAGCAGAA
D:	EAPAE	GAAGCACCAGCAGAA

Figure 44 : Similarités entre motifs répétés

Les motifs répétés similaires ont été alignés. Les acides aminés ou les nucléotides identiques entre les deux séquences comparées sont indiqués en caractère gras.

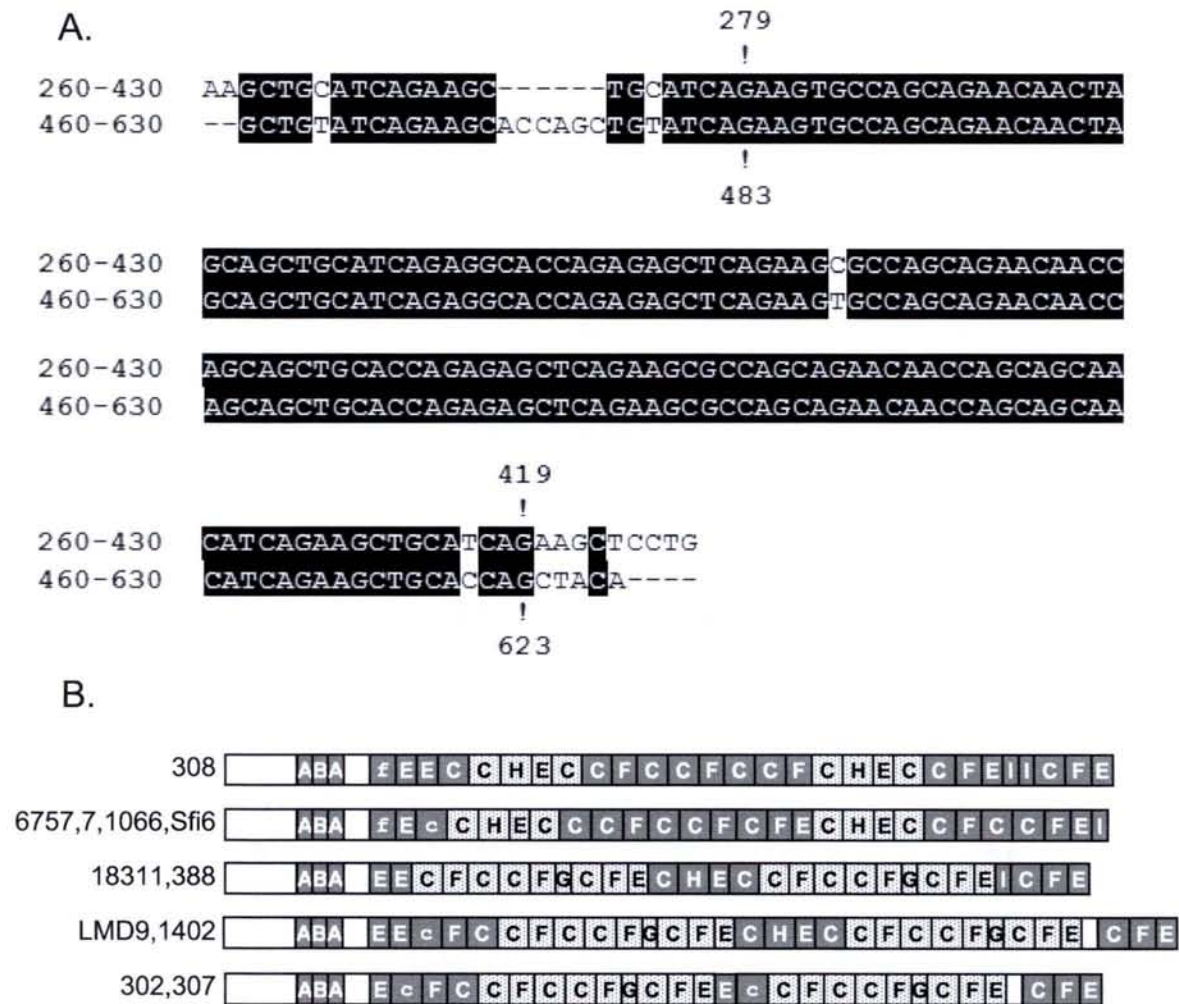


Figure 45 : Répétitions de successions de motifs répétés dans les allèles *var-cse*

A. Alignement de deux régions distinctes des allèles *var-cse* de la souche LMG18311 : les régions comprises entre les nucléotides 260 à 430 et les nucléotides 460 à 630. La position de chacune des unités répétées de 140 pb est indiquée au-dessus ou en-dessous des séquences.

B. Représentation schématique des régions *var-cse* des souches CNRZ308 (308), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, LMG18311 (18311), CNRZ388 (388), LMD-9, CNRZ1402 (1402), CNRZ302 (302), CNRZ307 (307). Les motifs répétés sont représentés sous la forme de rectangles contenant le nom de l'unité répétée. Les unités répétées de séquence conforme à la séquence consensus (séquences consensus définies figure 42) sont en lettres majuscules, celles qui ne le sont pas sont en lettres minuscules. Les motifs répétés inclus dans les grandes unités répétées sont en gris clair, ceux qui ne le sont pas en gris foncé.

	Ala	Glu	Pro	Ser	Thr
368,1,385	18,3	22,4	6,9	9,1	16,4
A054	18,5	21,0	6,3	9,8	17,6
445	18,5	23,2	7,0	9,7	18,1
19258	28,4	14,0	9,9	14,0	9,5
702	30,8	17,2	10,0	13,1	8,6
308	30,5	16,1	9,9	12,8	8,6
6757,7,1066,Sfi6	27,4	16,7	10,2	13,9	9,0
18311,388	28,0	17,6	10,9	13,4	9,2
LMD9,1402	28,2	17,3	10,8	13,5	8,5
302,307	28,3	17,6	10,1	13,1	8,2
Protéome	7,3	6,8	3,3	4,2	5,7

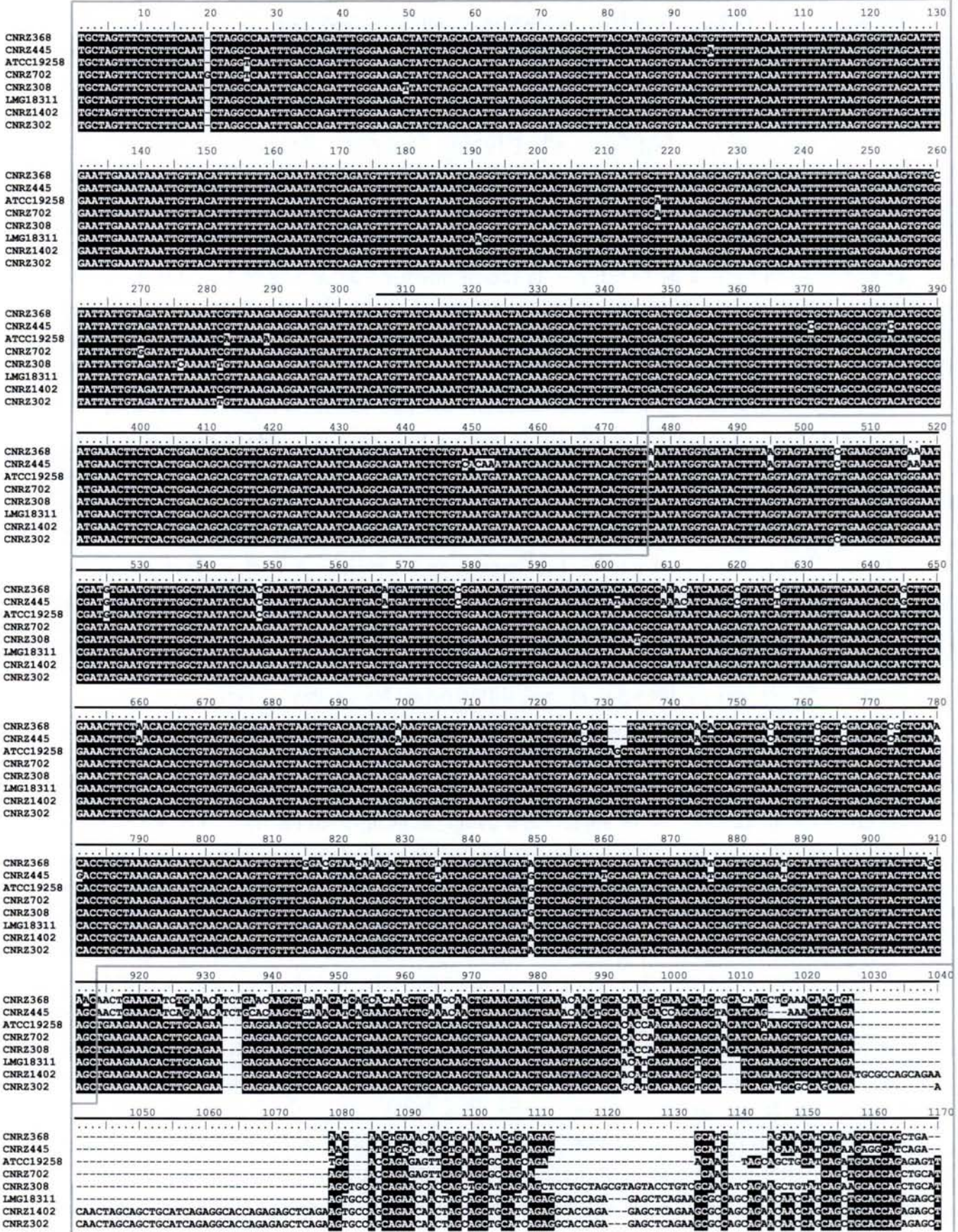
Tableau 1 : Enrichissement de la région Var-Cse en certains résidus d'acides aminés

Le contenu (en %) en certains acides aminés dans la région Var-Cse des souches CNRZ368 (368), NST1 (1), CNRZ385 (385), A054, CNRZ445 (445), ATCC19258 (19258), CNRZ702 (702), CNRZ308 (308), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, LMG18311 (18311), CNRZ388 (388), LMD-9, CNRZ1402 (1402), CNRZ302 (302) et CNRZ307 (307) est indiqué. Le contenu en ces mêmes acides aminés dans le protéome de *S. thermophilus* est également présenté.

Cette scission en ces deux mêmes groupes peut également être réalisée lorsque les séquences localisées en amont de la région *var-cse* sont comparées entre elles (Figure 39). En effet, cette région présente au total 16 positions polymorphes. Sept d'entre elles (les positions 3, 6, 9, 10, 33, 47 et 60) sont caractérisées par deux allèles, avec l'un des deux allèles exclusivement commun aux séquences des souches du groupe I, et l'autre aux séquences des souches du groupe II. Dans le but de caractériser plus finement cette particularité, le reste du locus *cse* des souches CNRZ445, ATCC19258, CNRZ702, CNRZ308, CNRZ1402 et CNRZ302 a été amplifié par PCR avec les amorces *fcompl* et *rcmpl*, et les produits de PCR obtenus ont ensuite été séquencés. Les séquences en résultant, ainsi que celles des souches CNRZ368 et LMG18311 déjà disponibles, ont ensuite été alignées (Figure 46). Les séquences comprennent l'intégralité de l'ORF *cse*, la région intergénique en amont de l'ORF et les 24 paires de bases en aval. L'analyse de la région *var-cse*, comprise entre les positions 914 et 1500, et nommée subdivision C dans la Figure 46, a été présentée dans le paragraphe précédent.

L'analyse de l'alignement des séquences révèle que les régions comprises entre les positions 477 et 913 (nommée région B), et les positions 1501 et 1729 (nommée région D), permet également de scinder les souches en les groupes I et II, sur le même critère de divergence ponctuelle (Figure 46). En revanche, cette distinction n'est plus possible lorsque l'alignement des régions nommées A et E, flanquant la région comprenant B, C et D, est analysée. C'est sur ce critère que les frontières entre les régions A et B, et D et E, ont été définies. Plus précisément, la frontière entre A et B a été positionnée juste en amont de la position 477, première position polymorphe différenciant les souches du groupe I des souches du groupe II. La frontière entre D et E a quant à elle été positionnée juste en aval de la position 1729, dernière position polymorphe différenciant les souches de la même façon.

La subdivision de ce locus en 5 parties se justifie également par les résultats de la construction de matrice d'identité de séquences. Une matrice d'identité des séquences disponibles a été construite pour les régions A, B, C, D et E (Figure 47). Ensuite, les résultats de la comparaison des séquences appartenant au même groupe ont été moyennés et un écart-type a été calculé (lorsque qu'il y a plus de une donnée) afin de construire une autre matrice d'identité de séquences (Figure 48). Dans la région B, les séquences du groupe I présentent entre elles 97% d'identité et les séquences du groupe II présentent entre elles $99 \pm 0,35\%$ d'identité. En revanche, les séquences du groupe I présentent avec celles du groupe II $92 \pm 0,77\%$ d'identité. Ainsi, dans cette région du locus *cse*, les identités sont plus élevées entre séquences du même groupe qu'entre séquences de groupes différents. Le même résultat est obtenu lorsque les séquences comprises dans les régions C (*var-cse*) et D sont analysées (Figure 48). En revanche ce n'est plus vrai pour les régions A et E. En effet, dans la région A, les séquences du groupe I présentent entre elles 98,3% d'identité, celles du groupe II $99 \pm 0,40\%$ et celles du groupe I présentent $99 \pm 0,76\%$ d'identité avec les séquences du groupe II. Il y a en effet peu ou pas de différence significative entre ces trois valeurs. Un résultat analogue est obtenu lorsque les séquences de la région E sont comparées entre elles. Par conséquent, la distinction des séquences en deux groupes n'est plus possible dans les régions A et E. Ce ne sont donc que les séquences comprises dans les régions B, C et D qui peuvent être classées en deux groupes.



A

B

C

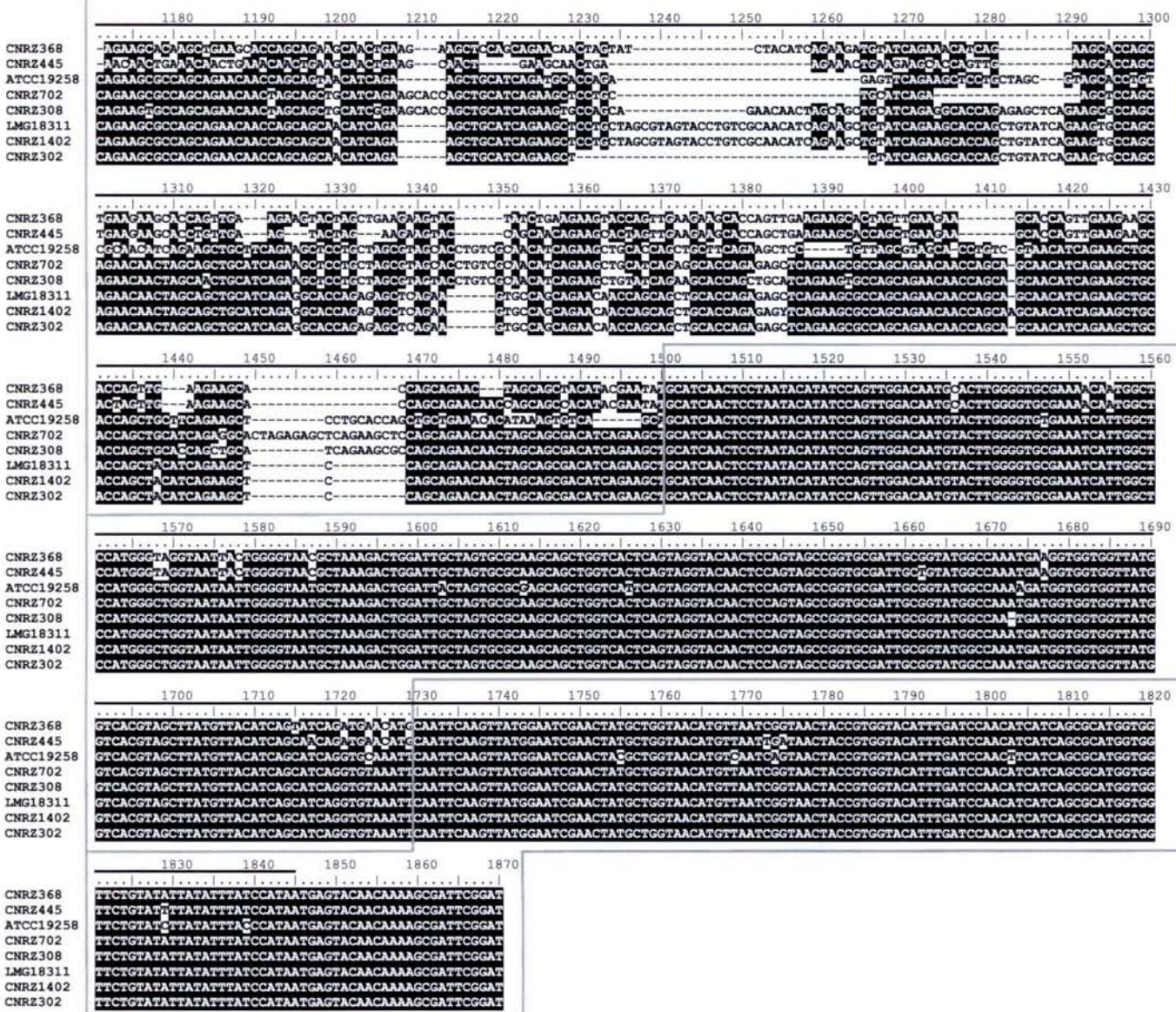


Figure 46 : Alignement de séquences de loci *cse* de différentes souches de *S. thermophilus*

Les produits de PCR obtenus avec les amorces fcompl et rcompl à partir des ADN génomiques des souches CNRZ445, ATCC19258, CNRZ702, CNRZ308, CNRZ1402 et CNRZ302, ont été séquencés. Les séquences obtenues, ainsi que celles des souches CNRZ368 et LMG18311 déjà disponibles, ont été alignées avec le logiciel CLUSTALW. L'alignement montre que le locus *cse* est une mosaïque de 5 parties (de A à E), délimitées par les encadrés gris. L'ORF *cse* est indiquée par un trait noir fin.

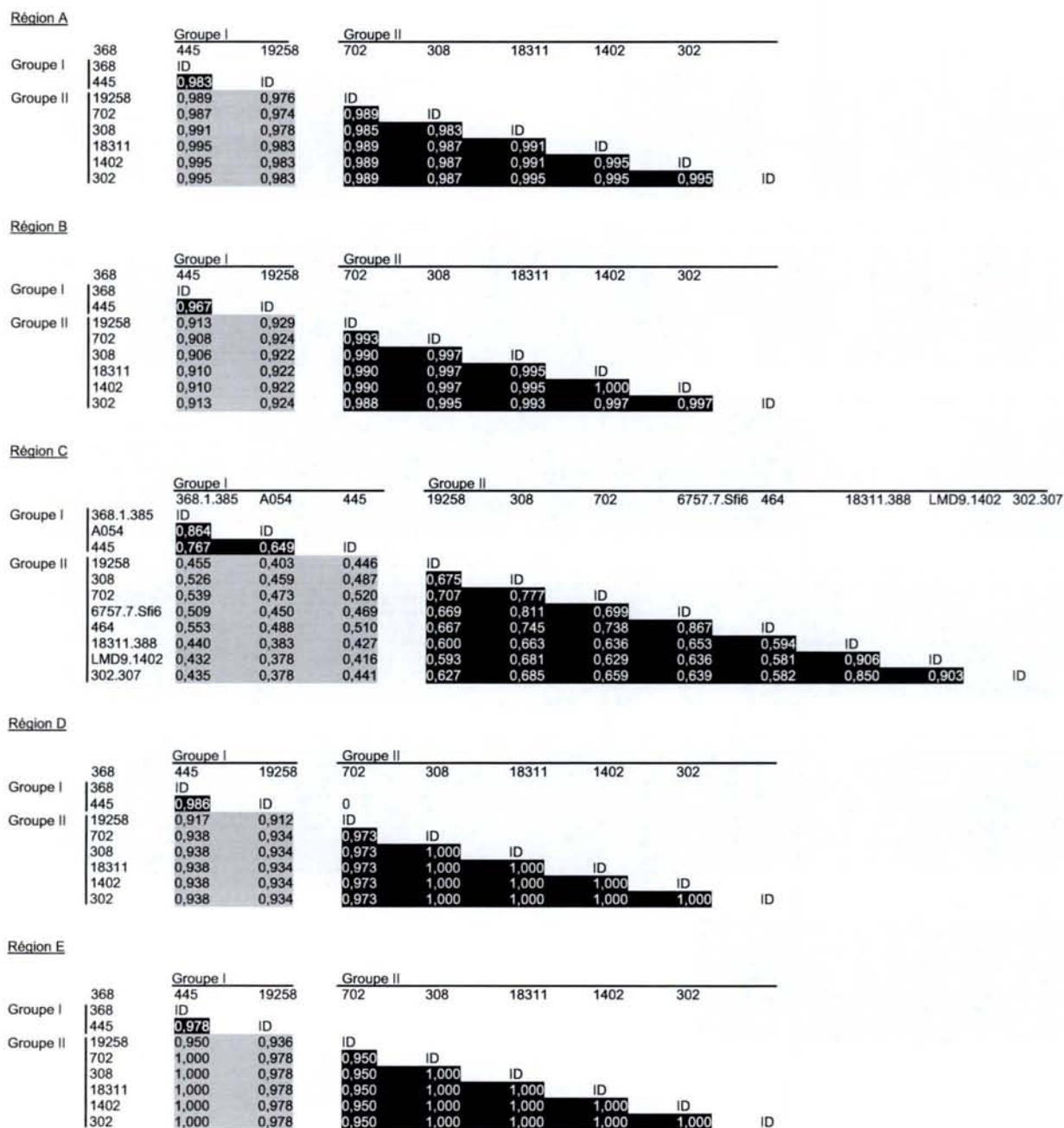


Figure 47 : Matrice d'identités de séquences des allèles cse

Les séquences du locus cse ont été subdivisées en les parties A, B, C, D et E (voir Figure 46). Les séquences ont ensuite été alignées avec le logiciel CLUSTAL W, et une matrice d'identités de séquences a ensuite été construite à partir de chaque alignement avec le programme Sequence Identity Matrix du logiciel Bioedit. Les résultats surlignés en noir correspondent aux comparaisons des séquences appartenant au même groupe, et ceux surlignés en gris, aux comparaisons des séquences d'un groupe avec celles de l'autre groupe.

<u>Région A</u>		
	Groupe I	Groupe II
Groupe I	98	
Groupe II	99±0,76	99±0,40

<u>Région B</u>		
	Groupe I	Groupe II
Groupe I	97	
Groupe II	92±0,77	99±0,35

<u>Région C</u>		
	Groupe I	Groupe II
Groupe I	76±11	
Groupe II	46±5,0	70±9,6

<u>Région D</u>		
	Groupe I	Groupe II
Groupe I	98	
Groupe II	93±0,86	99±1,3

<u>Région E</u>		
	Groupe I	Groupe II
Groupe I	98	
Groupe II	98±2,1	98±2,4

Figure 48 : Matrices d'identités des séquences des groupes I et II

Les séquences du locus *cse* ont été subdivisées en les parties A, B, C, D et E (voir figure 46). Les résultats présentés figure 47 ont été moyennés, afin de construire les matrices d'identités de séquences. La comparaison des séquences du groupe I entre elles dans les régions A, B, D et E n'aboutissant qu'à une valeur d'identité, la valeur représentée dans ces matrices n'est donc pas une moyenne mais la valeur brut.

L'ensemble de ces résultats montre que le locus *cse* est mosaïque (Figure 49). Les frontières entre les régions A et B, et D et E, étant vraisemblablement localisées à l'intérieur de l'ORF *cse*, l'ORF *cse* serait elle-même mosaïque.

2.8 Le gène *cse* est chimérique

Le génome de la souche *S. thermophilus* LMG18311 a été analysé à l'aide du programme BLASTN avec la séquence nucléotidique comprenant l'ORF *cse*_{LMG18311} et ses régions flanquantes. Les résultats révèlent qu'une région de 619 bp, nommée HRC1 (pour Homologous Region, Copy 1) (Figure 50), présente 93 % d'identité avec une autre région du génome appelé HRC2 (pour Homologous Region, Copy 2). HRC1 comprend les dernières 350 pb de *cse*, la région intergénique localisée en aval de *cse*, et les premières 58 pb d'*orf1* qui suivent (Figure 50). HRC2 comprend les dernières 350 pb de l'ORF *pcsB*, la région intergénique localisée en aval de *pcsB*, et les premières 58 pb de l'ORF *rppk* qui suivent. *Orf1* code une protéine de fonction inconnue, et *rppk* une ribose-phosphate pyrophosphokinase putative, enzyme de biosynthèse de phosphoribosyldiphosphate (ou PRPP), un précurseur des nucléotides puriques et pyrimidiques (Eriksen *et al.*, 2002).

La partie C-terminale de la protéine Cse présente 94 % d'identité avec la partie C-terminale de PcsB. La protéine PcsB joue un rôle important dans la morphologie cellulaire chez *S. agalactiae*, *S. mutans* et *S. pneumoniae* (Chia *et al.*, 2001b; Reinscheid *et al.*, 2001; Ng *et al.*, 2003; Reinscheid *et al.*, 2003; Ng *et al.*, 2004). L'interrogation de la banque de données CDD avec les protéines Cse et PcsB suggère que les régions homologues de ces deux protéines contiennent un domaine CHAP. Le domaine CHAP (cysteine histidine dependent aminohydrolase/peptidase) présente une activité glutathionylspermidine amidase chez *E. coli* (Bollinger *et al.*, 1995). Ce domaine est notamment rencontré dans des protéines extracellulaires dans lesquelles il pourrait être impliqué dans l'hydrolyse du peptidoglycane (Bateman et Rawlings, 2003; Heilmann *et al.*, 2003; Rigden *et al.*, 2003).

Des interrogations des banques de données ont révélé l'existence de protéines orthologues de PcsB dans les genres *Streptococcus* et *Lactococcus*. En revanche, aucune protéine présentant des similarités significatives sur toute la longueur de la protéine Cse n'est présente dans les banques de données.

La partie N-terminale de Cse présente 51 % d'identité avec la protéine SIP (Surface Immunogenic Protein) de *S. agalactiae*, protéine de fonction inconnue (Figure 50). Les résultats de l'interrogation de la banque de données CDD avec les protéines Cse et SIP suggèrent que les régions homologues de ces deux protéines contiennent un domaine LysM. Le domaine LysM, caractéristique de protéines de surface, est un domaine d'attachement à la paroi cellulaire (Steen *et al.*, 2003). Ce domaine est probablement responsable de la localisation préférentielle de la protéine SIP aux pôles cellulaires (Rioux *et al.*, 2001; Steen *et al.*, 2003).

La partie C-terminale de la protéine SIP présente 54 % d'identité avec la partie C-terminale de la protéine Orf1 (Figure 50), protéine codée par le gène situé en aval du gène

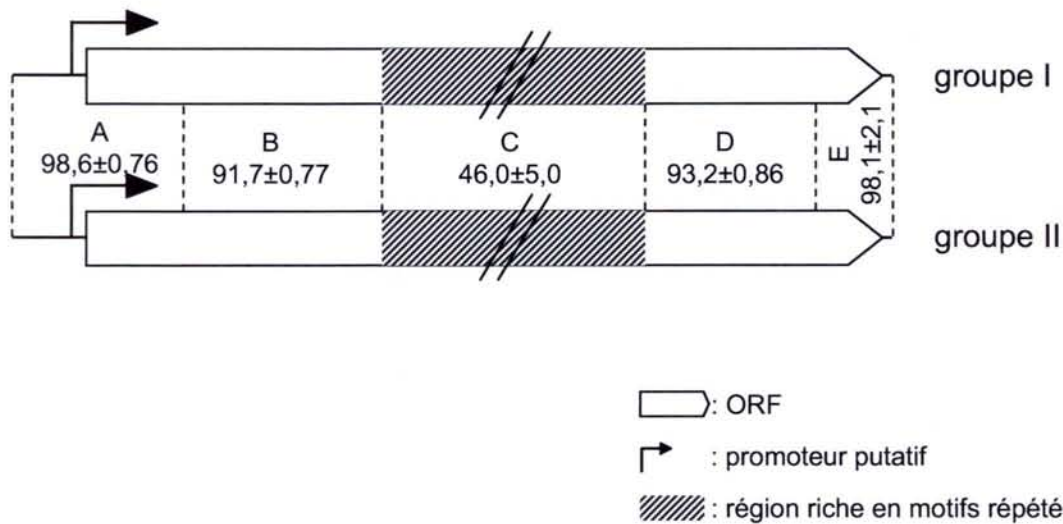


Figure 49 : Caractère mosaïque de l'ORF *cse*

Schéma représentant le locus *cse* des groupes I et II de souches de *S. thermophilus*. Le pourcentage d'identité entre les subdivisions (de A à E) du locus *cse* est indiqué entre les représentations schématiques des loci *cse*. La subdivision C, correspondant à *var-cse*, est partiellement représentée comme le symbolisent les deux traits parallèles obliques. Les groupes I et II sont définis figure 42

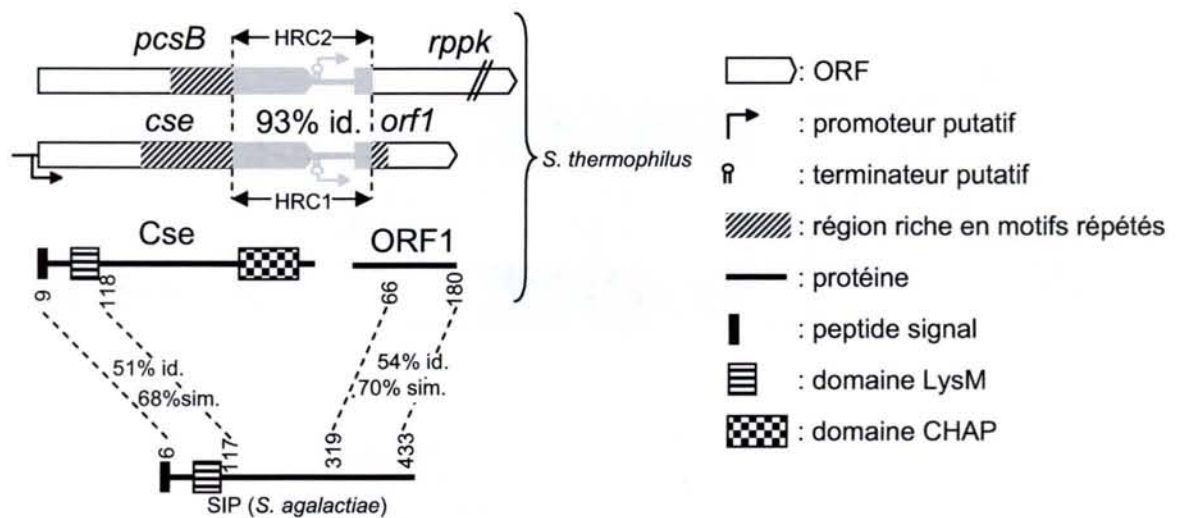


Figure 50 : Les gènes *cse* et *orf1* sont chimériques

Les régions d'ADN présentant de fortes identités sont représentées en gris. Les régions présentant des identités significatives sont reliées entre elles par des traits interrompus, la position des acides aminés délimitant ces régions est également indiquée. Les valeurs d'identité (id.) et de similarité (sim.) sont mentionnées. Les deux traits obliques indiquent que l'ORF est partiellement représentée.

cse (Figure 50). Ces régions protéiques ne présentent pas d'identité significative avec les domaines conservés répertoriés dans la banque de données CDD.

En conclusion, l'ORF cse est chimérique. La région HRC1 du locus cse est homologue de la région HRC2 du locus pcsB. Chacune des régions HRC code un domaine CHAP. De plus la région HRC1 est flanquée par des régions codant des fragments de protéines homologues d'une même protéine, la protéine SIP, présente chez d'autres streptocoques mais pas chez S. thermophilus.

2.9 Les ORF *pcsB* et *orfI* présentent des régions riches en répétitions

Les séquences environnant les frontières de HRC1 et HRC2 (Figure 50) ont été analysées et comparées chez la souche LMG18311. A l'instar de *cse*, la région située en amont de HRC2 est riche en motifs répétés (Figure 51). Ces motifs répétés sont similaires à ceux identifiés dans *var-cse*. Ce sont des motifs de type A, C, E, I et H (Figure 51). Aucun motif répété n'a été identifié dans la région située en aval de HRC2 (Figure 51).

De façon analogue, dans la région située en aval de HRC1, des motifs répétés similaires à ceux identifiés dans *var-cse*, les motifs E, C et F, ont été identifiés (Figure 51). Tous les motifs identifiés dans *orfI* sauf un, divergent de 17 % par rapport à la séquence consensus définie Figure 42 A. De plus, cette région riche en répétitions présente deux fois la succession ECCF, succession également présente dans la région *var-cse*.

Ainsi, les régions situées en amont d'HRC2 et en aval d'HRC1 présentent des motifs répétés communs à ceux identifiés dans la région var-cse, elle-même située en amont de la région HRC1.

2.10 L'ORF *pcsB* est variable au niveau intraspécifique

Compte tenu de la ressemblance de la région centrale de l'ORF *pcsB* avec la région variable *var-cse*, nous avons fait l'hypothèse qu'elle est également variable au niveau intraspécifique. Afin de tester cette hypothèse, la région centrale de l'ORF *pcsB* a été amplifiée par PCR à partir de l'ADN génomique de 7 souches de *S. thermophilus*. La comparaison des produits de PCR par électrophorèse a permis de révéler l'existence d'au moins trois allèles de taille différente (Figure 52 A). Les produits de PCR obtenus avec l'ADN génomique des souches CNRZ368, CNRZ302 et IP6757, chacun de taille différente, ont été séquencés. L'analyse des séquences a révélé que la variabilité de la taille des produits de PCR est liée à une modification du nombre de la succession des motifs C, H, E et I (Figure 52 B). En effet, l'allèle *pcsB* des souches LMG18311/IP6757, CNRZ302 et CNRZ368 présente respectivement 2, 3 et 1 fois la succession CHEI.

En conclusion la région centrale de l'ORF pcsB est variable au niveau intraspécifique. A l'instar de cse, cette variabilité se manifeste par une modification du nombre de motifs répétés.

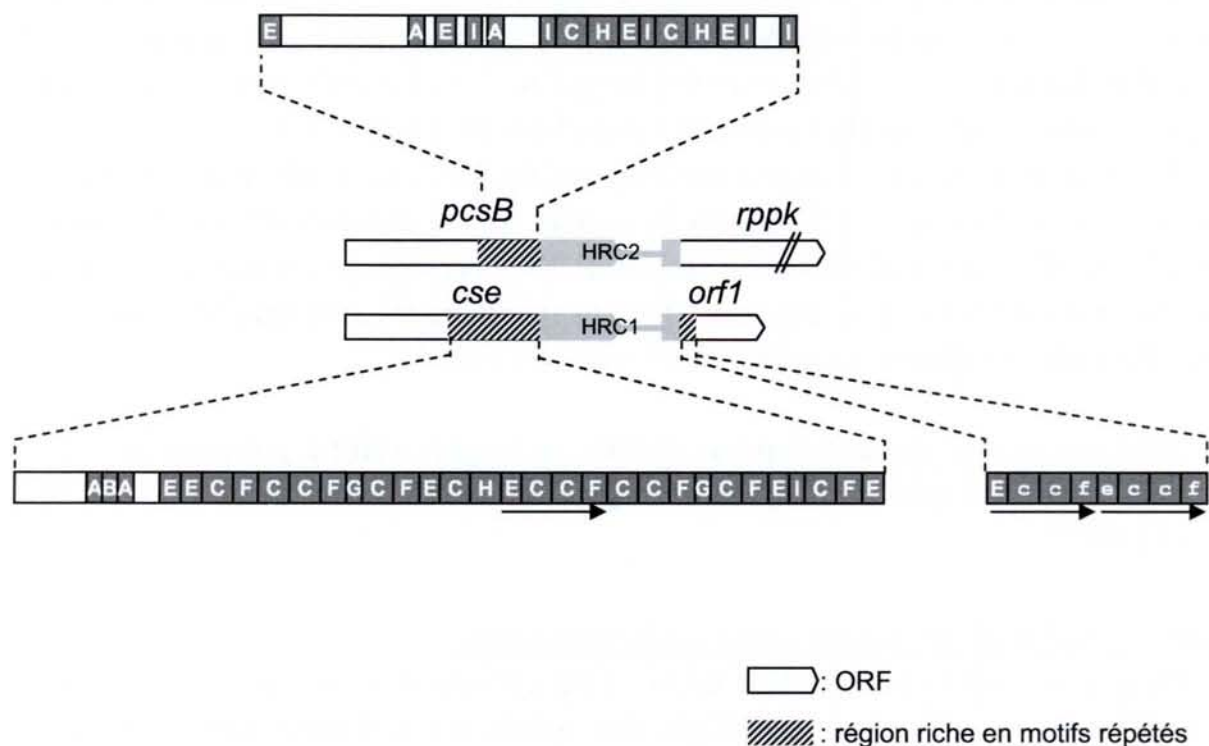
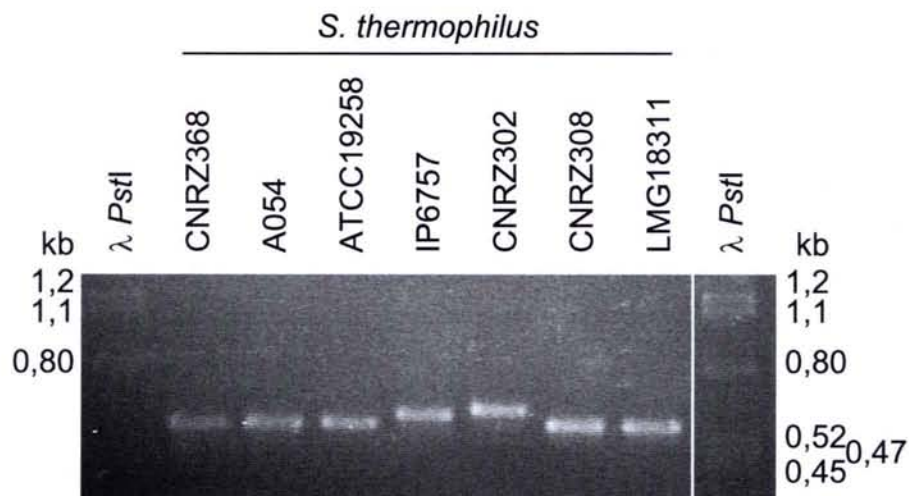


Figure 51 : Représentation schématique des régions riches en motifs répétés de *cse*, *pcsB* et *orf1*

Représentation schématique des loci *cse* et *pcsB* de la souche *S. thermophilus* LMG18311. Les régions d'ADN HRC sont représentées en gris. Les régions riches en répétitions sont agrandies au-dessus et en dessous de la représentation schématique des loci *cse* et *pcsB*. Les motifs répétés sont représentés sous la forme de rectangles contenant le nom de l'unité répétée. Les unités répétées de séquence conformes à la séquence consensus (séquences consensus définies figure 42) sont en lettres majuscules, celles qui ne le sont pas sont en lettres minuscules. Les flèches noires désignent la succession de motifs répétés ECCF.

A.



B.

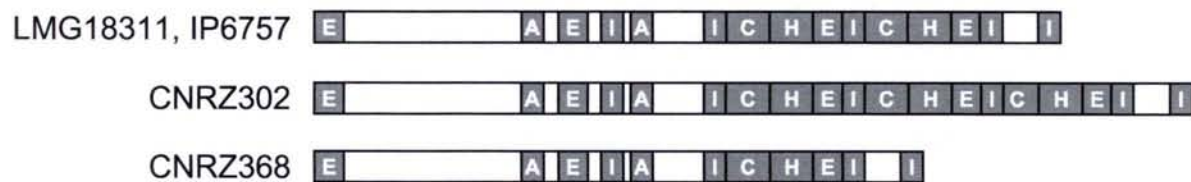


Figure 52 : Variabilité de la région riche en motifs répétés de *pcsB*

A. Migration électrophorétique des produits de PCR obtenus à partir de l'ADN génomique de huit souches de *S. thermophilus* et grâce aux oligonucléotides Forvar.pcsB et Revvar.pcsB. L'ADN du bactériophage λ digéré par l'enzyme de restriction *PstI* est utilisé comme marqueur de taille.

B. Représentation schématique de la région riche en motifs répétés de *pcsB*. Les motifs répétés sont représentés sous la forme de rectangles contenant le nom de l'unité répétée (séquences consensus définies figure 42).

2.11 Impact de la variabilité de *var-cse* sur l'activité de ségrégation cellulaire

2.11.1 Variabilité intraspécifique de la longueur des chaînes de cellules

L'hypothèse de travail était que la variabilité de la région *var-cse* pouvait avoir des conséquences sur l'activité de ségrégation cellulaire. Si tel était le cas, une variabilité intraspécifique de la longueur des chaînes de cellules devait être observable.

Le nombre de cellules par chaîne de sept souches de *S. thermophilus* a été comptabilisé dans les mêmes conditions que décrites précédemment (voir § 2.1.3). Les résultats montrent que la répartition du nombre de cellules par chaîne est variable d'une souche à l'autre (Figure 53). Certaines souches se développent sous la forme de très courtes chaînes, comme c'est le cas de la souche CNRZ445, dont 95 % des cellules constituent des chaînes de 1 à 20 cellules. D'autres souches peuvent se développer sous la forme de longues chaînes comme c'est le cas de la souche LMG18311, dont 23 % des cellules forment des chaînes présentant plus de 200 cellules.

Ainsi, la longueur des chaînes de cellules est variable au sein de l'espèce S. thermophilus.

2.11.2 La variabilité intraspécifique de la longueur des chaînes de cellules est corrélée à un facteur de variabilité du locus *cse*

2.11.2.1 Choix des souches

Dans le but de tester l'hypothèse selon laquelle la variabilité intraspécifique de la longueur des chaînes de cellules est liée à une variabilité allélique de *cse*, des souches mérodiplôides pour le gène *cse* ont été construites. Les souches CNRZ368 et LMG18311 ont été choisies d'une part parce que leur répartition du nombre de cellules par chaîne sont différentes (Figure 53), et d'autre part parce que leur aptitude à être transformée était connue.

2.11.2.2 Influence de l'allèle *cse*_{CNRZ368} sur la longueur des chaînes de la souche LMG18311

Cette expérience a été réalisée avec le plasmide initialement construit pour compléter le phénotype du mutant Δcse de la souche CNRZ368. Ce plasmide porte l'ORF *cse* ainsi que son propre promoteur et son propre terminateur putatif. Ainsi, la souche LMG18311 a été transformée avec le plasmide pNST260+::*cse*_{CNRZ368}. L'intégration du plasmide dans le chromosome a été sélectionnée par culture à 42°C, température à laquelle l'origine de réplication du plasmide n'est plus fonctionnelle, et en présence d'érythromycine. La présence des deux types d'allèles dans le génome de cette souche a été vérifiée grâce à une amplification par PCR de la région *var-cse*, les deux allèles *var-cse* étant discernables par électrophorèse (Figure 54). L'identité de la souche réceptrice a été vérifiée par hybridation d'une sonde révélant spécifiquement un fragment du locus *eps* des souches LMG18311 et IP6757 (Pluvinet *et al.*, 2004) (résultat non montré). Le nombre de copies intégrées du plasmide a également été vérifié : celui-ci est intégré une seule fois dans le génome de la souche LMG18311 (résultat non montré).

La souche LMG18311-pNST260+::*cse*_{CNRZ368} ainsi obtenue, de génotype (*cse*_{LMG18311}/*cse*_{CNRZ368}), a été cultivée jusqu'en phase stationnaire (20 h de culture), puis le nombre de cellules par chaîne a été compté. Les cellules de la souche

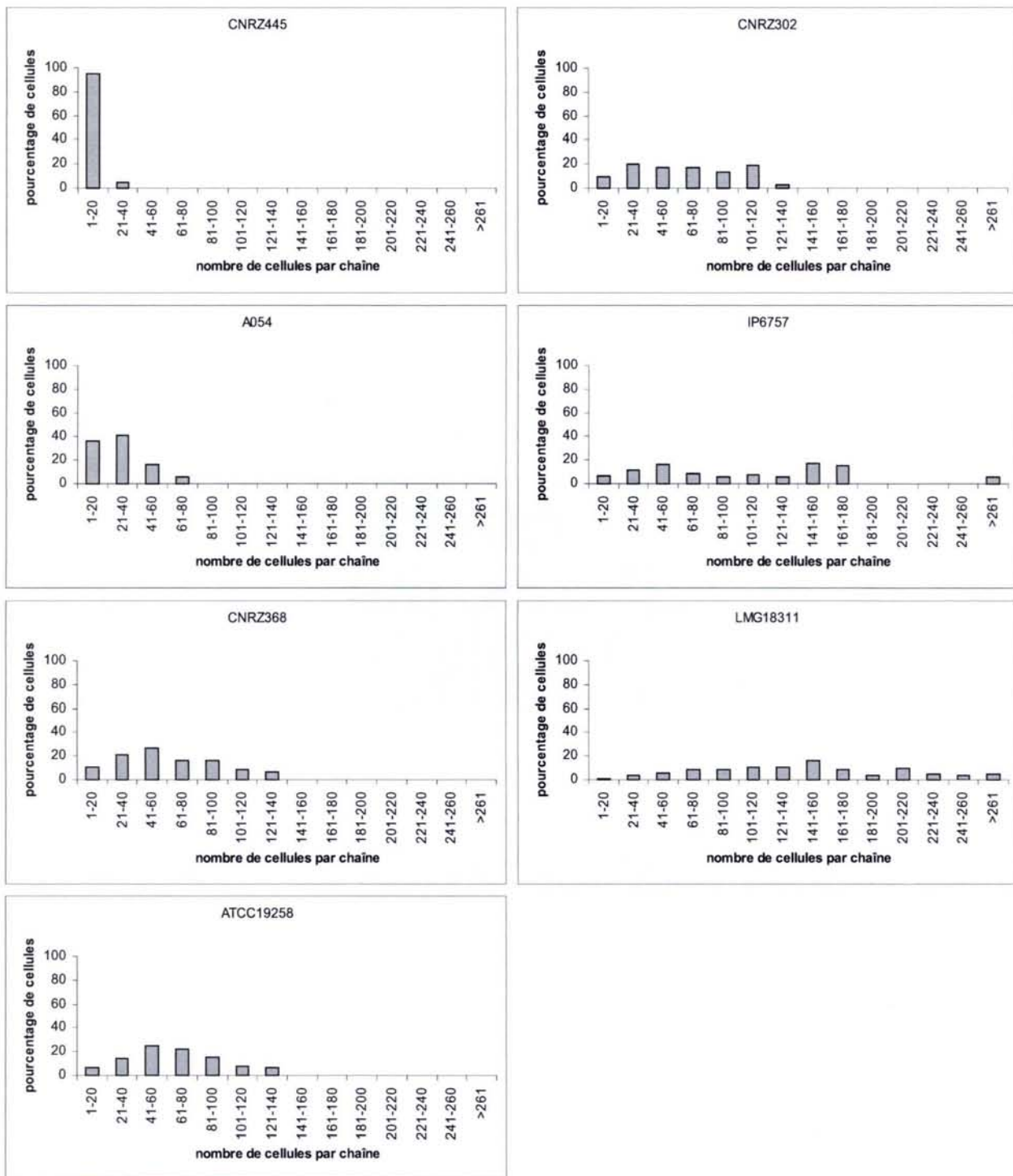


Figure 53 : Polymorphisme intraspécifique du nombre de cellules par chaîne

Différentes souches de *S. thermophilus*, dont le nom figure en titre de chaque graphique, ont été cultivées en milieu TPPY liquide pendant 20h. Les cellules sont alors en phase stationnaire. Ensuite, pour chaque souche au moins 1000 cellules ont été comptées par observation au microscope photonique à contraste de phase à un grossissement de 1000.

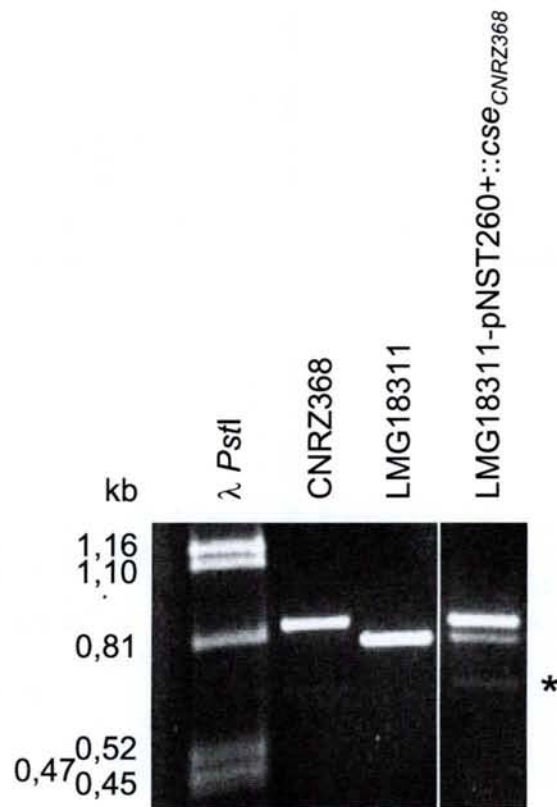


Figure 54 : Profils électrophorétiques des produits d'amplification de la région centrale du gène *cse* des souches sauvages CNRZ368 et LMG18311, et de la souche mérodiplöide hétérozygote pour le gène *cse*.

Les produits de PCR ont été obtenus avec le couple d'oligonucléotides Forvar/Revvar2 et les ADN génomiques des souches CNRZ368, LMG18311 et LMG18311-pNST260+::*cse*_{CNRZ368}. L'ADN du bactériophage λ digéré par l'enzyme de restriction *Pst*I est utilisé comme marqueur de taille.

* : produit secondaire.

LMG18311-pNST260+::*cse*_{CNRZ368} se répartissent dans les chaînes de façon similaire à celles de la souche CNRZ368 (Figure 55 A). En guise de témoin, la même expérience a été réalisée avec le plasmide pNST260+. Dans ce cas, les cellules se répartissent dans les chaînes de façon analogue à celles de la souche sauvage LMG18311 (Figure 55 B), montrant que le plasmide n'a pas d'influence sur la longueur des chaînes de cette souche.

2.11.2.3 Influence d'une copie surnuméraire de *cse*_{CNRZ368} sur la longueur des chaînes de la souche CNRZ368

Deux hypothèses simples peuvent expliquer le phénotype de chaînes courtes de la souche mérodiplôïde LMG18311-*cse*_{CNRZ368} : (1) l'augmentation du nombre de copies du gène *cse* induit une augmentation de l'activité de ségrégation cellulaire par augmentation de la quantité de protéine Cse synthétisée, (2) et l'allèle *cse*_{CNRZ368} confère une activité de ségrégation cellulaire plus importante que l'allèle *cse*_{LMG18311}.

Afin de trancher entre ces deux hypothèses, l'effet d'une augmentation du nombre de copies de *cse* dans la souche CNRZ368 a été estimée par construction d'une souche CNRZ368-pNST260+::*cse*, dans laquelle le plasmide pNST260+::*cse*_{CNRZ368} est intégré dans le chromosome de la souche CNRZ368. Les cellules de cette souche de génotype (*cse*_{CNRZ368}/*cse*_{CNRZ368}) se répartissent dans les chaînes de façon similaire à celles de la souche sauvage (Figure 55 C). La même expérience a été réalisée avec le plasmide pNST260+ en guise de témoin et donne le même résultat (Figure 55 D). Ces résultats indiquent que l'augmentation du nombre de copies de *cse* n'induit pas d'augmentation de l'activité de ségrégation cellulaire, suggérant que le phénotype de chaînes courtes de la souche LMG18311 mérodiplôïde, de génotype (*cse*_{LMG18311}/*cse*_{CNRZ368}), n'est pas dû à l'augmentation du nombre de copies du gène *cse*.

Ce résultat suggère fortement que la région du génome de la souche CNRZ368 comprenant l'allèle cse, son promoteur putatif ainsi que son terminateur putatif, confère une activité de ségrégation cellulaire plus importante que celle de la souche LMG18311.

2.11.3 Remplacement de la région var-cse de la souche LMG18311 par celle de la souche CNRZ368 : absence d'effet sur la ségrégation cellulaire

Nous avons fait l'hypothèse que le facteur de divergence responsable de la variabilité de la longueur des chaînes est la région *var-cse*. Dans le but de tester cette hypothèse, une souche de LMG18311 délétée du gène *cse* a été construite dans un premier temps (voir § 2.1.2). Cette délétion entraîne une augmentation de la longueur des chaînes (Figure 56), montrant que l'allèle *cse* de la souche LMG18311 est fonctionnel.

Ensuite la région *var-cse* de la souche LMG18311 sauvage a été remplacée par celle de la souche CNRZ368. Le remplacement allélique a été sélectionné avec un plasmide portant une chimère du gène *cse*. En effet, les régions flanquant la région *var-cse* ne sont pas à 100% identiques entre les souches CNRZ368 et LMG18311. C'est pourquoi un plasmide portant d'une part les régions flanquant la région *var-cse* de la souche LMG18311, et d'autre part la région *var-cse* de la souche CNRZ368, a été construit (Figure 57). Le plasmide a été vérifié par séquençage, puis une souche de LMG18311 pour laquelle la région *var-cse* a été

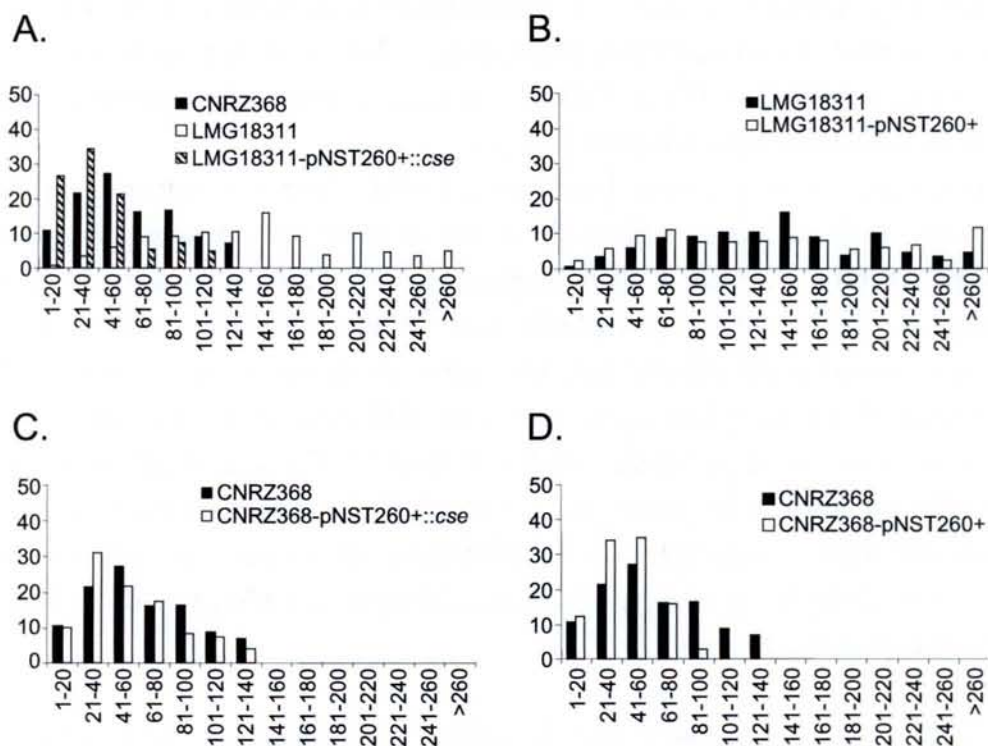


Figure 55 : Influence de l'addition d'une copie de $cse_{CNRZ368}$ dans le génome des souches LMG18311 et CNRZ368 sur la longueur des chaînes de cellules

Les souches bactériennes ont été cultivées en milieu TPTY liquide non agité pendant 20 h. Les cellules sont alors en phase stationnaire. Au total, au moins 1000 cellules ont été comptées pour chacune des souches, à partir de trois cultures indépendantes. Les cellules ont été comptées par observation au microscope photonique à contraste de phase à un grossissement de 1000.

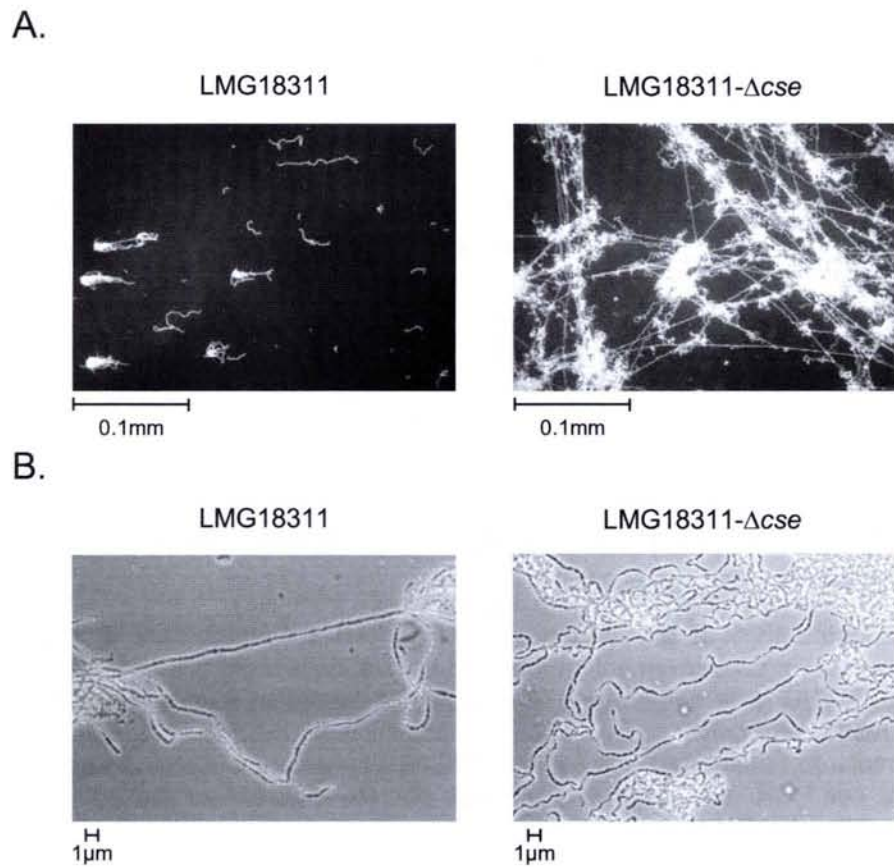


Figure 56 : Effet de la délétion de l'allèle cse de la souche LMG18311 sur la longueur des chaînes

Les souches bactériennes ont été cultivées en milieu TPPY liquide non agité pendant 20h. Les cellules sont alors en phase stationnaire. Les chaînes ont été photographiées à des grossissements de 100 (A.) et de 1000 (B.).

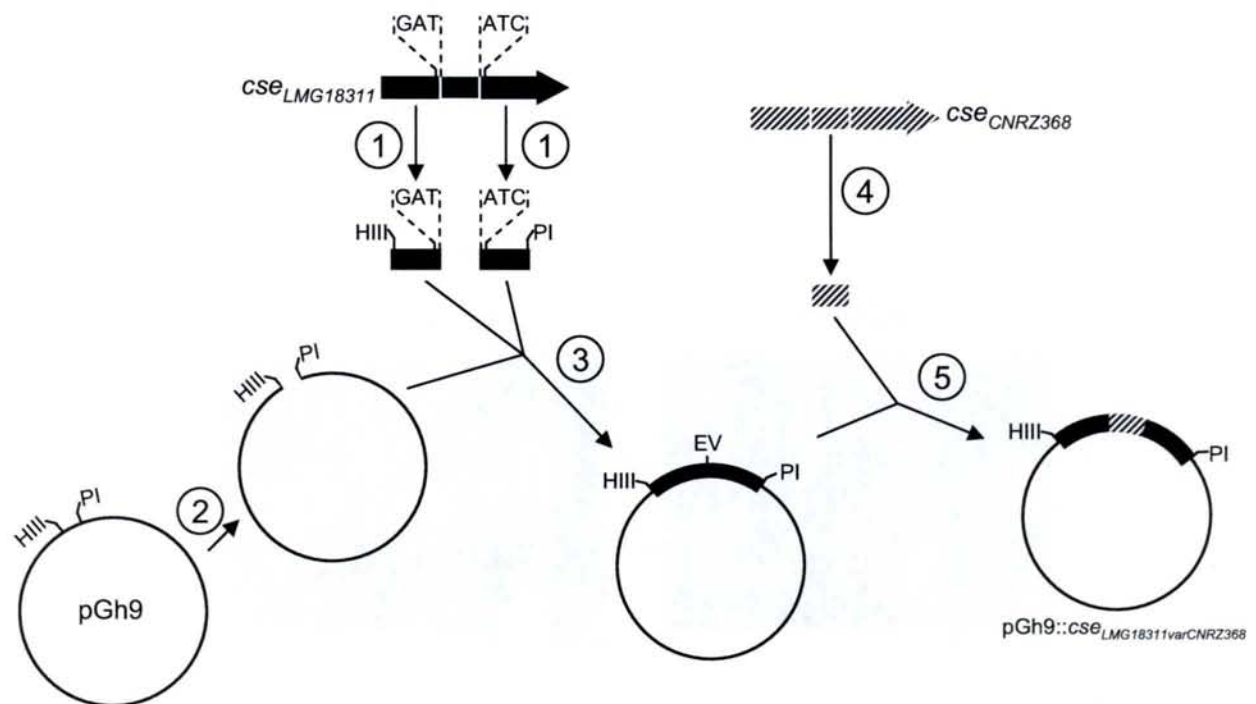


Figure 57 : Représentation schématique des différentes étapes de construction du plasmide pGh9::cse_{LMG18311varCNRZ368}

HIII : *HindIII*, PI : *PstI*, EV : *EcoRV* (génère des extrémités franches).

L'approche adoptée implique les séquences GAT et ATC, présentes dans la séquence sauvage et localisées respectivement en amont et en aval de la région *var-cse*. La réunion de la séquence GAT à la séquence ATC entraîne la formation d'un site reconnu par l'enzyme de restriction *EcoRV*.

1: Les fragments de l'allèle *cse* de la souche LMG18311, flanquant la région *var-cse*, ont été amplifiés par PCR. Les fragments obtenus ont été digérés par les enzymes de restriction *HindIII* et *EcoRV* ou *PstI* et *EcoRV*.

2 : L'ADN du plasmide pGh9 a été digéré par les enzymes de restriction *HindIII* et *PstI*.

3 : Les deux produits de PCR ainsi que l'ADN plasmidique ont été mélangés et ligés ensemble. Puis, des cellules d'*E. coli* ont été transformées avec l'ADN du mélange de ligation afin de cloner un plasmide dans lequel les deux produits de PCR sont reliés ensemble.

4 : La région *var-cse* de la souche CNRZ368 a été amplifiée par PCR. La polymérase ajoutant des nucléotides A simple brin aux extrémités 3' des brins néosynthétisés, les fragments d'ADN ont été incubés en présence de Pfu polymérase qui, par son activité 3'-5' exonucléase, élimine ces A simple brin. Des extrémités franches ont ainsi été générées.

5 : Le plasmide a été digéré par l'enzyme de restriction *EcoRV*, puis mélangé avec le produit de PCR correspondant à la région *var-cse* de la souche CNRZ368. Des cellules d'*E. coli* ont été transformées avec le mélange de ligation, afin de cloner le fragment *var-cse* dans le plasmide. L'orientation de la région *var-cse* a été déterminée par PCR puis par séquençage afin de sélectionner le plasmide pGh9::cse_{LMG18311varCNRZ368}.

remplacée par celle de la souche CNRZ368 a été sélectionnée. Le remplacement effectif de cette région a été vérifié par PCR puis séquençage de la région remplacée.

Après avoir été cultivées jusqu'en phase stationnaire (20 h de culture), les cellules de la souche LMG18311-*cse*_{varCNRZ368} ont été comptées. Les cellules de cette souche mutante se répartissent dans les chaînes de façon similaire à celle de la souche LMG18311 (Figure 58). Le remplacement de la région *var-cse* a ainsi peu ou pas d'effet sur l'activité de ségrégation cellulaire de Cse, sur le critère du nombre de cellules par chaîne.

Ce résultat montre que la variabilité de la région var-cse a peu ou pas d'influence sur l'activité de ségrégation cellulaire de la protéine Cse.

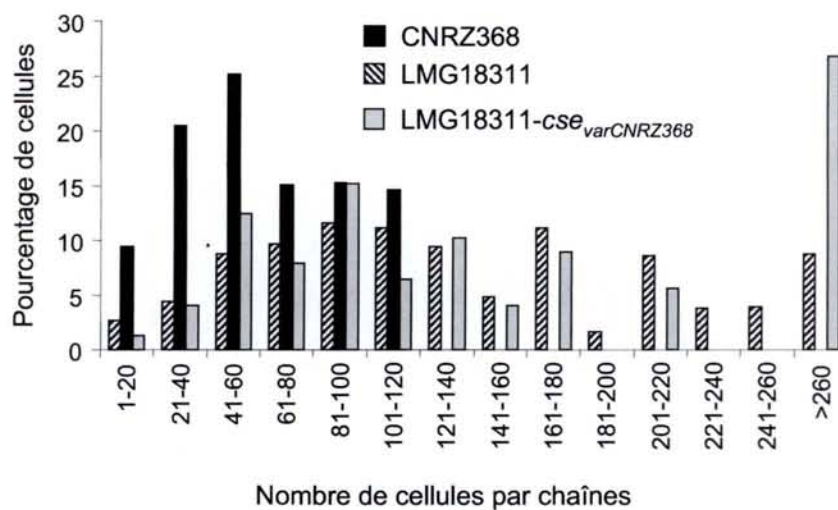


Figure 58 : Impact du remplacement de la région *var-cse* de la souche LMG18311 par celle de la souche CNRZ368, sur la longueur des chaînes de cellules

Les souches bactériennes ont été cultivées en milieu TPTY liquide non agité pendant 20 h. Les cellules sont alors en phase stationnaire. Au total, au moins 1000 cellules ont été comptées pour chacune des souches, à partir de trois cultures indépendantes. Les cellules ont été comptées par observation au microscope photonique à contraste de phase à un grossissement de 1000.

3 Sélection de variants [EPS-, adhésion+] par culture en milieu semi-liquide

L'objectif initial du travail présenté dans cette partie était de déterminer si la méthode de criblage de souches bactériennes, sur le critère de la longueur des chaînes, méthode décrite pour *L. lactis* par Mercier *et al.* (2002), était applicable à *S. thermophilus*. Les expériences ont été réalisées en collaboration avec S. Kulakauskas à l'URLGA du centre INRA de Jouy-en-Josas.

3.1 Corrélation entre vitesse de sédimentation en milieu semi-liquide et longueur des chaînes de cellules

Le milieu semi-liquide est un milieu contenant une faible concentration en agar (de l'ordre de 0,05 % d'agar). A une température de 30-42°C, l'agar augmente la viscosité du milieu de culture, entraînant une diminution de la vitesse de sédimentation des cellules en croissance. Cette diminution de la vitesse de sédimentation est proportionnelle à l'augmentation de la longueur des chaînes de cellules (Figure 59, site <http://www.jouy.inra.fr/unites/urlga/biofilms.shtml>). Cette propriété a été utilisée afin de trier des chaînes de cellules, en fonction de leur longueur, dans une population hétérogène (Mercier *et al.*, 2002). Les auteurs ont ainsi pu isoler des mutants de chaînes courtes à partir d'une souche se développant sous la forme de chaînes longues.

3.2 Sélection de variants de croissance en milieu semi-liquide chez *S. thermophilus*

Environ 50 cellules d'une préculture de 12 h des souches *S. thermophilus* CNRZ368 et CNRZ368- Δcse ont été cultivées indépendamment, à 42°C sans agitation, dans 30 ml de milieu M17 lactose semi-liquide (agar : 0,05 %).

A 14 h d'incubation, l'observation des flacons révèle que la souche sauvage CNRZ368 ainsi que la souche CNRZ368- Δcse (caractérisée par de longues chaînes), croissent de façon similaire. En effet, dans les deux cas les colonies ont une forme globulaire (Figure 60). Ce phénotype est similaire à celui observé pour la souche $\Delta acmA$, souche se développant sous la forme de longues chaînes. En conclusion, la méthode de criblage ne serait pas applicable pour différencier les chaînes de la souche *S. thermophilus* CNRZ368 sauvage d'une souche présentant des chaînes plus longues.

A 28 h d'incubation, une minorité de colonies de la souche CNRZ368 présente une morphologie coloniale variante. En effet, celles-ci présentent un filament de cellules solidaire de la masse globulaire dont il semble être issu (Figure 61). Des cellules de ce filament ont été prélevées avec une seringue, puis striées sur boîte. Des colonies ont ensuite été cultivées en milieu liquide, puis en milieu semi-liquide. La forme des colonies de la souche variante, nommée CNRZ368-6.1, est différente de celle de la souche sauvage (Figure 61). En effet, les colonies du variant ont une forme plus allongée que celles de la souche sauvage, suggérant que les cellules de la souche variante sédimentent plus vite.

L. Lactis $\Delta acmA$

L. lactis

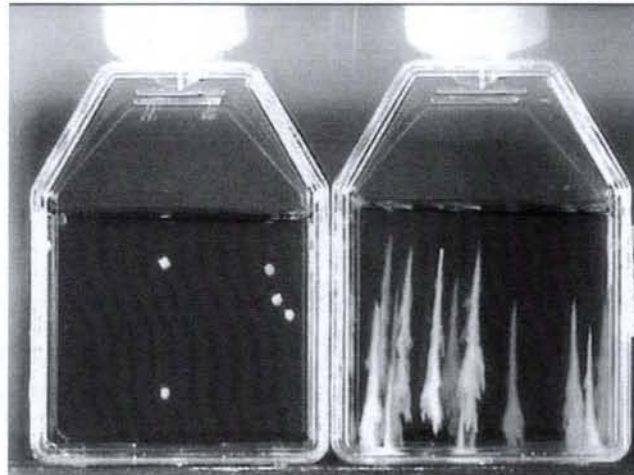
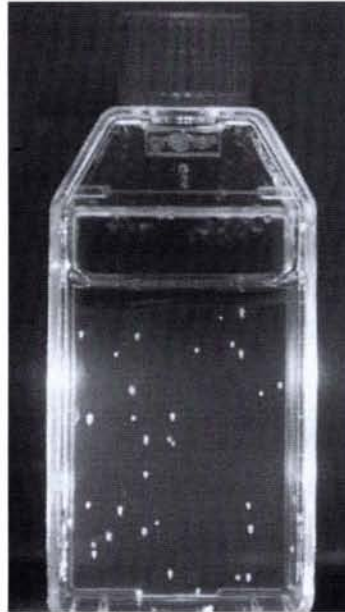


Figure 59 : Influence de la viscosité du milieu de culture sur la sédimentation de chaînes de différentes longueurs (d'après Mercier et al., 2002)

Ce cliché photographique a été pris après culture des souches de *L. lactis* en milieu M17 semi-liquide (agar : 0,05 %) pendant 28 h à 30°C. La souche sauvage *L. lactis* MG1363 forme de courtes chaînes de cellules, alors que la souche $\Delta acmA$ en forme de longues. Photos provenant du site <http://www.jouy.inra.fr/unites/urlga/biofilms.shtml>

A.

CNRZ368



B.

CNRZ368

CNRZ368- Δ cse

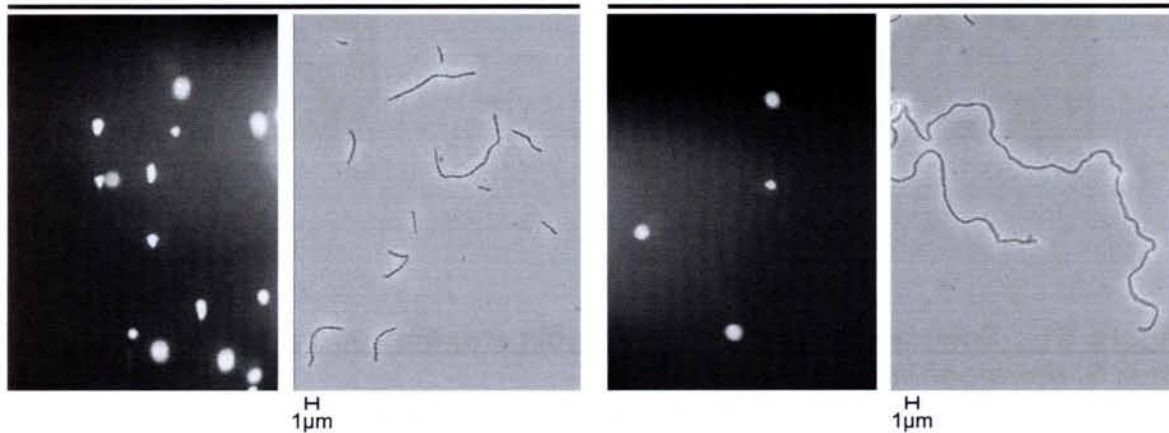


Figure 60 : Culture en milieu semi-liquide de différentes souches de *S. thermophilus* CNRZ368

Les souches ont été cultivées en milieu M17 semi-liquide pendant 14 h à 42°C.

A. Cliché photographique d'un flacon de culture de la souche *S. thermophilus* CNRZ368.

B. Clichés photographiques de colonies des souches *S. thermophilus* CNRZ368 et CNRZ368- Δ cse. Des cellules ont été prélevées de ces flacons, puis photographiées au microscope photonique à contraste de phase, à un grossissement de 1000.

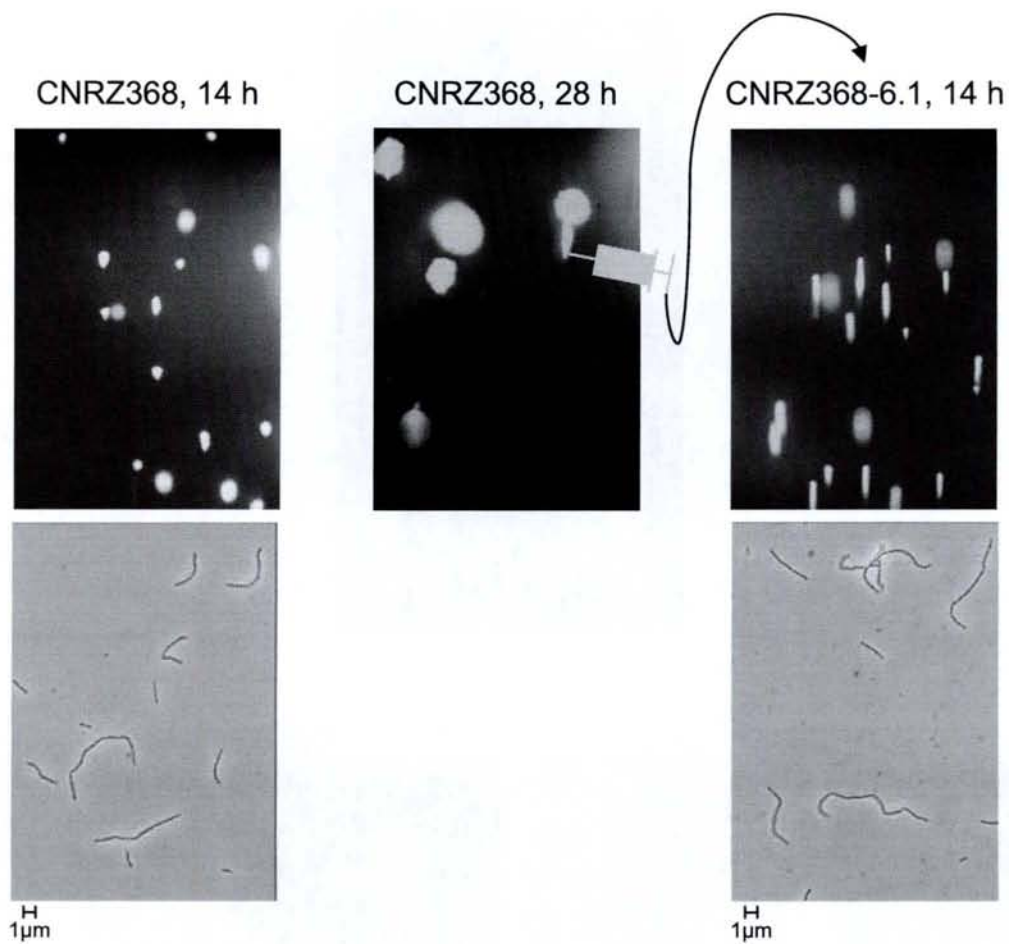


Figure 61 : Sélection de variants de sédimentation en milieu semi-liquide chez *S. thermophilus* CNRZ368

Les souches ont été cultivées en milieu M17 semi-liquide à 42°C. Le flacon de culture de la souche CNRZ368 a été photographié à 14 h, puis à 28 h d'incubation. A 28 h d'incubation, une colonie présente un filament de cellules solidaire de la masse globulaire dont il semble être issu. Des cellules de ce filament ont été prélevées avec une seringue, puis striées sur boîte. Plusieurs colonies ont été cultivées 12 h en milieu liquide puis cultivées 14 h en milieu semi-liquide. Les colonies du variant nommé CNRZ368-6.1 ont été photographiées. Des cellules d'une colonie du variant et de la souche sauvage ont été prélevées, puis photographiées en microscopie photonique à contraste de phase, à un grossissement de 1000.

L'observation des chaînes au microscope photonique n'a pas révélé de changement majeur de leur longueur chez le variant CNRZ368-6.1 par rapport à la souche sauvage (Figure 61).

Un autre variant a été obtenu indépendamment, le variant CNRZ368-2.1, et donne les mêmes résultats.

En conclusion, l'augmentation de la vitesse de sédimentation des variants CNRZ368-6.1 et CNRZ368-2.1 ne serait pas due à une réduction de la longueur des chaînes.

3.3 Les variants de sédimentation sont affectés dans la production d'EPS

Une hypothèse simple permettant d'expliquer le fait que les variants CNRZ368-6.1 et CNRZ368-2.1 sédimentent plus vite que la souche sauvage, est qu'ils sont affectés dans la production d'exopolysaccharides (EPS). Il est vraisemblable que les EPS forment une matrice extracellulaire dans laquelle les cellules pourraient être emprisonnées en milieu semi-liquide. En effet, l'ensemble cellules/EPS pourrait avoir une masse moléculaire suffisante pour ralentir la sédimentation au point de ne plus être sensible à la modification de la longueur des chaînes. La souche *L. lactis* MG1363 n'étant pas productrice d'EPS, il aurait été possible dans ce cas de trier les chaînes de cette souche en fonction de leur longueur (Mercier *et al.*, 2002).

Les EPS produits par la souche sauvage CNRZ368, et les deux variants CNRZ368-2.1 et CNRZ368-6.1, ont été extraits puis dosés (Dubois *et al.*, 1951; Robson *et al.*, 1974). Les résultats montrent que les deux variants CNRZ368-2.1 et CNRZ368-6.1 produisent moins d'EPS que la souche sauvage (Figure 62).

3.4 Le variant de sédimentation CNRZ368-6.1 a un pouvoir d'adhésion à une surface en verre augmenté

L'équipe de S. Kulakauskas a récemment montré que les EPS diminuent la capacité de la bactérie les produisant à adhérer à une surface en verre (communication personnelle, <http://www.jouy.inra.fr/unites/urlga/biofilms.shtml>). Ainsi, la capacité à adhérer à une surface en verre de l'un des deux variants, le variant CNRZ368-6.1, a été mesurée. Les résultats montrent que le variant CNRZ368-6.1 adhère significativement mieux à la surface en verre que la souche sauvage (Figure 63).

En conclusion, la culture de la souche S. thermophilus CNRZ368 en milieu semi-liquide permettrait de sélectionner des variants [EPS-] au pouvoir d'adhésion augmenté à une surface en verre. Mercier et al. (2002) ont utilisé la croissance en milieu semi-liquide pour sélectionner des mutants présentant des chaînes courtes chez une souche de L. lactis ne produisant pas d'EPS (Kulakauskas, communication personnelle). Il est ainsi probable que cette méthode de criblage soit utilisable sur des souches de S. thermophilus ne produisant pas d'EPS.

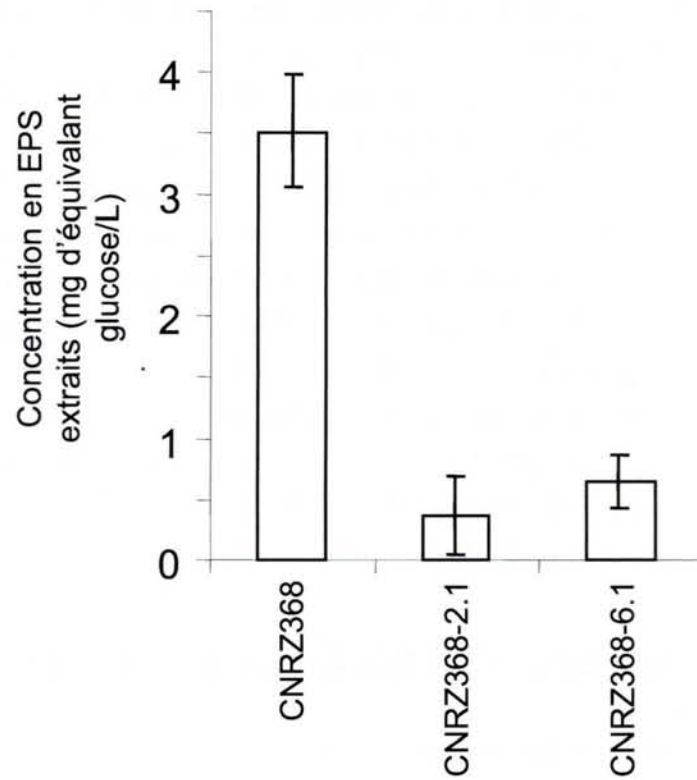


Figure 62 : Production d'EPS des variants de sédimentation

Les souches bactériennes ont été cultivées en milieu lait pendant 12 h, puis les EPS ont été extraits selon la méthode de Dubois *et al.* (1951). L'expérience a été réalisée trois fois de façon indépendante.

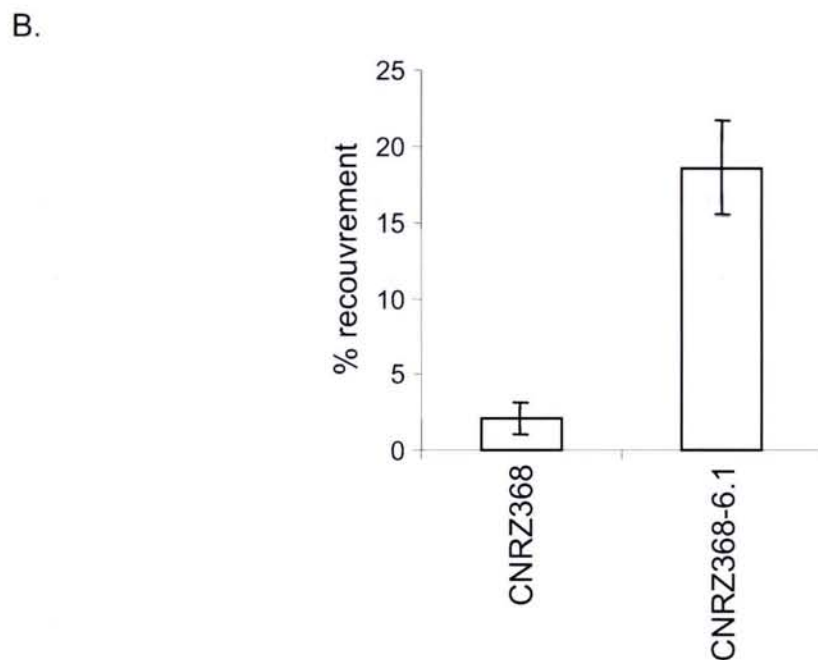
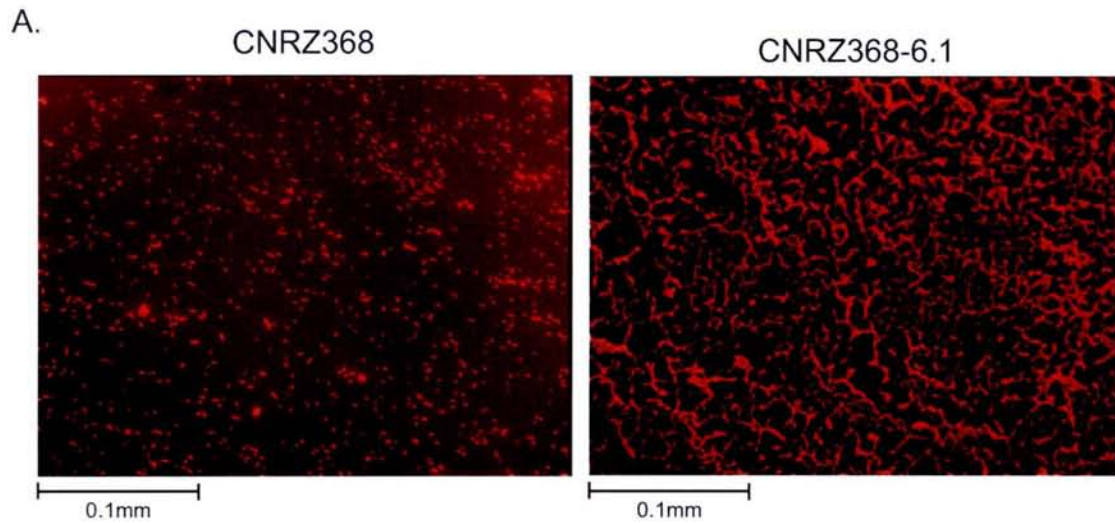


Figure 63 : Pouvoir d'adhésion à une surface en verre du variant de sédimentation CNRZ368-6.1

Les lames de verre ont été photographiées en microscopie à fluorescence (A.) afin de calculer le pourcentage de la surface recouverte par les cellules (B.). Trois expériences indépendantes ont été réalisées pour chacune des souches bactériennes. Pour chaque expérience d'adhésion, le pourcentage de recouvrement a été calculé à partir de 10 clichés photographiques.

Discussion

Ce travail a permis d'identifier un gène nommé *cse* (pour cell segregation), impliqué dans la ségrégation cellulaire chez *S. thermophilus*. Ce gène présente trois particularités génétiques : il est chimérique, variable et mosaïque. Il a été montré que *cse* influence les propriétés adhésives des cellules de *S. thermophilus*. Par ailleurs, un autre facteur influençant potentiellement le pouvoir adhésif de façon négative a été identifié : la production d'EPS. L'ensemble de ces résultats est discuté en trois parties distinctes.

1 Le gène *cse* code une protéine extracellulaire impliquée dans la ségrégation cellulaire

1.1 Architecture modulaire des protéines extracellulaires impliquées dans la ségrégation cellulaire

Streptococcus thermophilus, tout comme les autres streptocoques, se développe sous la forme de chaînes de cellules. Un mutant de *S. thermophilus* dans lequel l'ORF *cse* est déléetée présente de longues chaînes de cellules et ne présente pas de modifications majeures de la morphologie cellulaire. Un test de complémentation, réalisé par ajout d'un plasmide intégratif portant une copie de l'ORF *cse* précédée de son propre promoteur, et suivi de son terminateur putatif, permet de restaurer le phénotype sauvage. Ces résultats montrent que *cse* est impliqué dans le raccourcissement des chaînes de cellules, et par conséquent dans la ségrégation des cellules.

Des gènes impliqués dans la ségrégation cellulaire ont déjà été identifiés chez d'autres coques se développant sous la forme de chaînes tels que *acmA* chez *L. lactis* (Buist *et al.*, 1995), *lytA* et *lytB* chez *S. pneumoniae* (Sanchez-Puelles *et al.*, 1986; Garcia *et al.*, 1999). Chacun de ces gènes code des protéines extracellulaires présentant une activité hydrolase du peptidoglycane. Ces protéines ont une structure modulaire : elles présentent un ou plusieurs domaines d'attachement à la paroi (Sanchez-Puelles *et al.*, 1990; Steen *et al.*, 2003), et un domaine enzymatique catalysant l'hydrolyse d'une liaison du peptidoglycane (Sanchez-Puelles *et al.*, 1986; Buist *et al.*, 1995; De Las Rivas *et al.*, 2002).

La structure de la protéine Cse semble également être modulaire : Cse présenterait un domaine LysM d'attachement à la paroi, à l'instar d'AcmA, ainsi qu'un domaine catalytique CHAP (pour cysteine, histidine-dependent amidohydrolase/peptidase). Les domaines CHAP sont souvent présents dans des protéines de surface (Bateman et Rawlings, 2003; Rigden *et al.*, 2003). Celui de l'autolysine Aae de *Staphylococcus epidermidis* présente une activité hydrolase du peptidoglycane (Heilmann *et al.*, 2003). Une hypothèse simple est donc que Cse corresponde à une protéine extracellulaire, attachée à la surface des cellules, et catalysant l'hydrolyse de liaisons du peptidoglycane.

1.2 Est-ce que la protéine extracellulaire Cse est une hydrolase du peptidoglycane ?

La construction de fusions traductionnelles de *cse* avec un gène rapporteur a montré que la protéine Cse est extracellulaire. Ensuite, une éventuelle activité hydrolase du peptidoglycane

a été recherchée. Pour cela, des extraits protéiques des souches *S. thermophilus* CNRZ368 sauvage et Δcse ont été testés en zymographie, afin de révéler une éventuelle activité hydrolase du peptidoglycane. Deux substrats différents ont été testés : des cellules *S. thermophilus* et des cellules de *Micrococcus lysodeikticus*. Ces cellules ont, au préalable, subi un traitement par chauffage à 120°C. Ces expériences n'ont pas révélé d'activité lytique associée à Cse.

Deux fragments de la protéine Cse ont également été surproduit chez *E. coli* : un fragment allant des résidus d'acides aminés 1 à 183, contenant notamment le domaine putatif LysM, et un fragment allant des résidus d'acides aminés 184 à 461, contenant la région riche en répétitions ainsi que le domaine putatif CHAP. Des extraits protéiques des souches d'*E. coli* surproduisant les fragments de Cse ont été testés par zymographie, dans les mêmes conditions que précédemment. Aucune activité lytique associée aux fragments de Cse n'a été détectée. Les zymogrammes ayant été réalisés par dénaturation des protéines, migration électrophorétique en SDS-PAGE, puis renaturation, l'absence de détection d'une activité hydrolase du peptidoglycane pouvait s'expliquer par une renaturation incorrecte de Cse.

D'autres tests ont été réalisés (thèse de Séverine Layec). L'activité lytique d'extraits de protéines natives des souches recombinantes d'*E. coli* surproduisant des fragments de la protéine Cse, a été testée en milieu liquide, avec comme substrat des cellules autoclavées de *S. thermophilus* ou de *M. lysodeikticus*. Cette expérience n'a pas révélé d'activité lytique associée à un des fragments de la protéine Cse (résultats non montrés). Enfin, une souche mutante de *L. lactis* se développant sous la forme de chaînes longues a été transformée avec le plasmide autorépliquatif pNST260+::cse. Cette expérience avait pour but de déterminer si cse était capable de compléter un phénotype de chaînes longues résultant d'une déficience en une hydrolase du peptidoglycane. La souche mutante utilisée est délétée du gène *acmA* (fourni par le Pr J. Kok) et la croissance de ce mutant sous la forme de chaînes longues résulte de l'absence de l'hydrolyse des liaisons du peptidoglycane normalement hydrolysées par AcmA. L'ajout du plasmide pNST260+::cse dans le génome du mutant $\Delta acmA$ ne se traduit pas par une diminution de la longueur des chaînes de cellules (résultats non montrés).

L'hypothèse d'une activité hydrolase du peptidoglycane portée par PcsB, une autre protéine présentant un domaine CHAP, avait déjà été testée chez *S. agalactiae*. Les auteurs ont effectué des tests de zymographie, mais ont également recherché à révéler une activité hydrolase du peptidoglycane par cartographie du peptidoglycane en HPLC. Aucune de ces approches n'a permis de détecter une éventuelle activité hydrolase du peptidoglycane de PcsB (Reinscheid *et al.*, 2001; Reinscheid *et al.*, 2003).

L'ensemble de ces résultats suggère soit que la protéine Cse n'a pas d'activité hydrolase du peptidoglycane, soit que les conditions testées ne permettent pas de révéler cette activité.

1.3 Fonction de la région Var-Cse

1.3.1 Une charnière interdomaine

La région Var-Cse est localisée entre le domaine LysM et le domaine CHAP. Par conséquent, une hypothèse simple serait que Var-Cse soit une charnière joignant ces deux domaines. Une telle fonction est supposée pour les séquences interdomaines appelées

charnières de type Q (ou Q linker) qui joignent les domaines de protéines régulant l'expression de gènes du métabolisme de l'azote (Wootton et Drummond, 1989). La séquence de ces charnières est riche en résidus d'acides aminés glutamine (ou Q), d'où leur nom. Des mutants ont été construits dans le but d'augmenter la taille de cette région dans les protéines NifA et NtrC. Ces variations de la longueur de la charnière n'ont pas de conséquence sur l'activité régulatrice de ces protéines (Wootton et Drummond, 1989). De façon analogue, l'influence de modifications de la séquence de la région Var-Cse sur l'activité de ségrégation cellulaire de la protéine Cse a été étudiée. Pour cela, la région *var-cse* de la souche LMG18311 a été remplacée au locus chromosomique par celle de la souche CNRZ368. La région Var-Cse de la souche CNRZ368 est 60 pb plus courte que celle de la souche LMG18311, et leurs séquences présentent 40% de divergence. De façon similaire aux protéines NifA et NtrC, les modifications apportées à la région Var-Cse ont peu ou pas d'influence sur l'activité de ségrégation cellulaire de la protéine Cse. Ainsi, un rôle de charnière interdomaine pour cette région Var-Cse serait en accord avec sa capacité à tolérer la variabilité.

La région Var-Cse mesure entre 136 et 180 résidus d'acides aminés selon les allèles séquencés. Or les régions charnières mesurent généralement entre 4 et 20 résidus d'acides aminés (Tanaka *et al.*, 2003). Par exemple, les charnières de type Q mesurent entre 15 et 25 résidus d'acides aminés (Wootton et Drummond, 1989). Compte tenu de la grande taille de la région Var-Cse, il semble alors probable qu'elle ait une autre fonction que celle de charnière interdomaine.

1.3.2 Un espaceur

La région Var-Cse pourrait constituer un espaceur entre les domaines LysM et CHAP nécessaires à l'activité de ségrégation cellulaire. Un rôle similaire a été suggéré pour la région riche en répétitions de la protéine de liaison au fibrinogène Clf de *S. aureus* (Hartford *et al.*, 1997). Le fibrinogène est un composant des matrices extracellulaires des mammifères (pour revue Hay, 1991). La protéine Clf est ancrée à la paroi *via* son extrémité C-terminale et interagit avec le fibrinogène grâce à un domaine localisé à son extrémité N-terminale. Le site de liaison covalent à la paroi et le domaine d'interaction au fibrinogène sont espacés par une région riche en motifs répétés DS. Cette région riche en répétitions permettrait au domaine de liaison au fibrinogène d'être exposé à la surface. En l'absence de cette région riche en répétitions, le domaine d'interaction avec le fibrinogène resterait enfoui au coeur de la paroi cellulaire et ne pourrait ainsi pas assurer son rôle (Hartford *et al.*, 1997). De façon analogue, la région Var-Cse pourrait permettre d'espacer physiquement le domaine catalytique CHAP du domaine d'attachement LysM, afin que le domaine CHAP puisse être à proximité de son substrat.

1.3.3 Un domaine de liaison à la paroi

Une hypothèse alternative serait que la région Var-Cse soit un domaine de liaison à la paroi. En effet, cette région de la protéine Cse est notamment riche en résidus d'acides aminés proline, sérine et thréonine. Or, le domaine de liaison à la paroi de la protéine Ftf de *Streptococcus salivarius* est également riche en ces résidus d'acides aminés (Rathsam et

Jacques, 1998; Sara *et al.*, 1998). De plus, les propriétés adhésives de ce domaine pourraient résulter de l'enrichissement en ces résidus d'acides aminés (Rathsam et Jacques, 1998). Les fonctions hydroxyles des sérines et thréonines pourraient interagir avec des composants de la paroi en formant des liaisons hydrogène. Dans ce contexte, la tolérance de la région Var-Cse envers la variabilité pourrait signifier que les interactions entre la région Var-Cse et la paroi ne requièrent pas de conservation de sa structure mais plutôt une conservation de sa composition en acides aminés. Or, malgré sa variabilité, la région Var-Cse reste enrichie en les mêmes acides aminés.

Un rôle d'attachement à la paroi a également été proposé pour un domaine d'une autre protéine de *S. thermophilus* : l'hydrolase du peptidoglycane Mur1. Cette protéine, pourtant associée à la paroi, ne présente pas de similarité avec les domaines d'attachement connus. Elle présente en revanche, une région particulièrement riche en résidus glycine, sérine et thréonine. Les auteurs ont par conséquent fait l'hypothèse que cette région permet à la protéine d'être liée à la paroi (Husson-Kao *et al.*, 2000).

1.3.4 Avantage adaptatif conféré par la variabilité de la région Var-Cse

La variabilité de la région centrale du gène *cse* a peu ou pas d'influence sur l'activité de ségrégation cellulaire de la protéine Cse, comme le montre les résultats issus du remplacement de la région *var-cse* d'une souche de *S. thermophilus* par celle d'une autre. Par conséquent, si la variabilité de cette région du gène confère un avantage adaptatif, il est probable qu'il soit indépendant de l'activité de ségrégation cellulaire de Cse. Chez les streptocoques pathogènes, la variabilité de protéines de surface leur confère la capacité à échapper au système immunitaire de leur hôte (Madoff *et al.*, 1996). Cette variabilité leur permet ainsi de s'adapter à un milieu pouvant être hostile à leur survie. Une hypothèse simple, serait que la variabilité de la protéine Cse confère à *S. thermophilus* la capacité à échapper à un facteur de son environnement affectant sa survie ou sa capacité à se diviser. Ce facteur reconnaîtrait les cellules de *S. thermophilus* via la protéine Cse. La seule niche écologique connue de *S. thermophilus* est le lait et des produits alimentaires issus de sa fermentation. Dans ces environnements, il existe des bactériophages pouvant entraîner la lyse des cellules de *S. thermophilus*. Ces bactériophages ont un spectre d'hôte très étroit, suggérant qu'un facteur de variabilité entre les souches de *S. thermophilus* en est responsable. Une hypothèse serait que la protéine Cse soit un récepteur de phages et que la variabilité de sa région centrale permette de générer des variants de la protéine qui ne seraient plus reconnus par ces phages.

Cette hypothèse a été testée en collaboration avec Sylvain Moineau (Université Laval, Canada)(résultats non montrés). Deux phages infectant la souche *S. thermophilus* CNRZ368 ont été identifiés. Toutefois, ces deux phages restent capable d'infecter la souche CNRZ368 lorsque le gène *cse* est délété de son génome, suggérant fortement que la région Var-Cse de la protéine Cse n'est pas le récepteur de ces phages. La question d'un avantage adaptatif conféré par la variabilité de la région Var-Cse reste ouverte.

2 Origine et évolution du gène *cse*

2.1 Origine et conséquences fonctionnelles du caractère chimérique de *cse*

2.1.1 Le gène *cse* aurait été généré par réassortiment de modules

La région HRC1 (homologous region, copy 1), localisée dans le locus *cse*, présente un fort pourcentage d'identité (93%) avec une autre région du génome nommée HRC2 qui est localisée dans le locus *pcsB* (Figure 64). Le fort pourcentage d'identité entre les régions HRC suggère que l'une des deux provient d'un événement de duplication de l'autre région.

La région HRC1 est flanquée par des fragments d'ORF de gènes distincts : *cse* et *orf1*. Ces régions codent toutes deux des fragments de protéines homologues de la même protéine : la protéine SIP de *S. agalactiae* (Rioux *et al.*, 2001). Plus précisément, la partie N-terminale de *Cse* est homologue de la partie N-terminale de SIP, alors que la partie C-terminale de *ORF1* est homologue de la partie C-terminale de SIP. Par ailleurs, l'analyse du génome complet n'a pas révélé l'existence de protéine de *S. thermophilus* homologue de SIP sur toute sa longueur. En revanche, de tels homologues ont été trouvés chez tous les autres streptocoques dont le génome est entièrement séquencé, excepté chez *S. mutans*. Le gène *pcsB* quant à lui semble ubiquitaire chez les streptocoques, y compris *S. thermophilus*. Ces résultats suggèrent que c'est la copie HRC2 qui aurait subi un événement de duplication. Cette duplication aurait eu lieu, soit en intragénomique, soit par transfert horizontal.

Les résultats suggèrent également que l'insertion de la copie dupliquée aurait eu lieu à l'intérieur d'une ORF ancestrale. Cette ORF aurait codé une protéine SIP (Figure 64). Les conséquences de cet événement d'insertion auraient été la génération de deux gènes chimériques : *cse* et *orf1*, et la disparition du gène ancestral *sip*. En conclusion, le gène *cse* est chimérique, il aurait été généré par réassortiment de fragments de gènes non homologues.

2.1.2 La protéine *Cse* résulterait d'un réassortiment de domaines

Les régions HRC englobent chacune une séquence codant l'intégralité d'un domaine CHAP. Ainsi, le réarrangement qui aurait été à l'origine de la génération du gène *cse* aurait engendré la fusion d'un fragment de gène, codant entre autre un domaine LysM putatif, à un autre fragment de gène, codant un domaine CHAP putatif. Ainsi, ce réarrangement aurait eu pour conséquence l'association d'un domaine LysM à un domaine CHAP. Par conséquent, *Cse* aurait été généré par réassortiment de domaines et le gène *cse* par réassortiment de modules. En effet, un module a été défini précédemment comme un fragment de gène codant un domaine protéique.

2.1.3 Conséquence fonctionnelle du réassortiment de domaines à l'origine de *Cse*

2.1.3.1 *Introduction*

PcsB et *Cse* ont toutes deux un domaine CHAP du côté C-terminal et ne présentent pas d'identité significative du côté N-terminal. Ces deux protéines seraient donc constituées de combinaisons différentes de domaines. Par conséquent il est probable que *PcsB* et *Cse* aient des fonctions différentes.

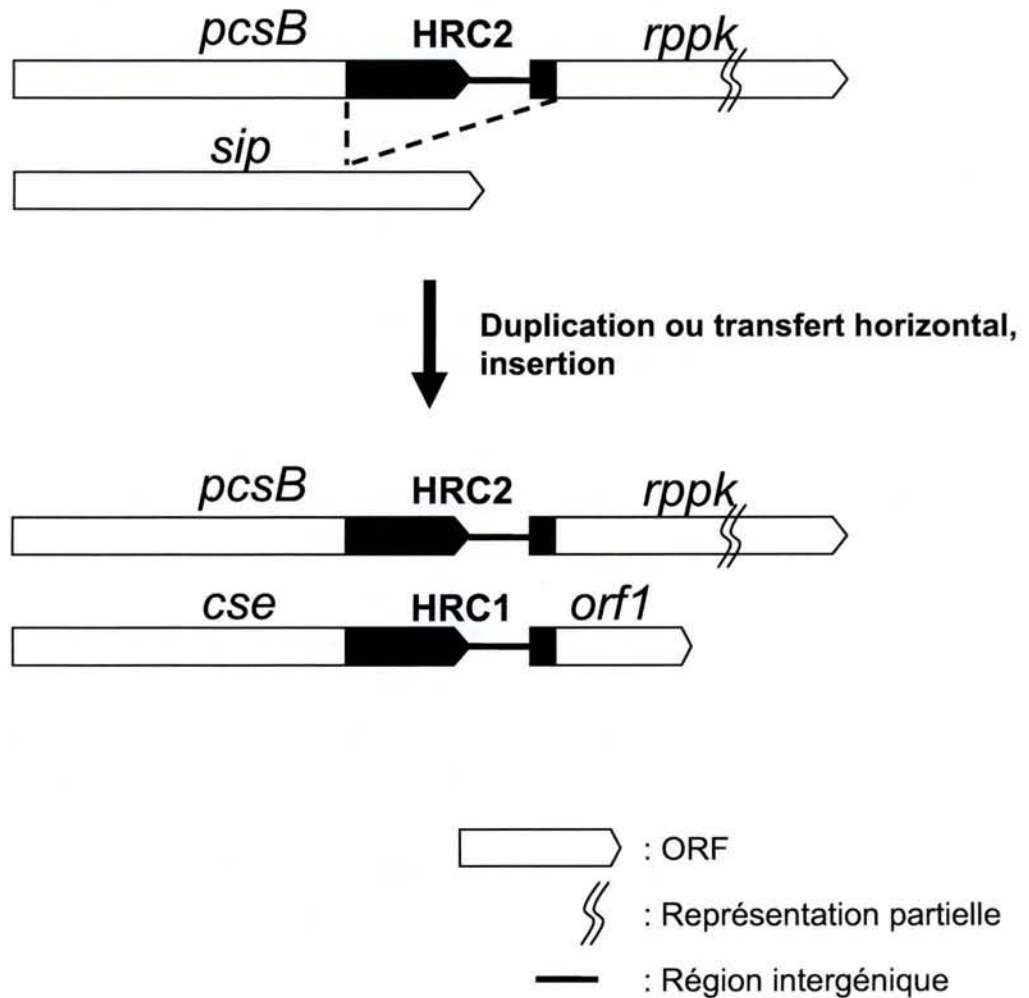


Figure 64 : Génération des gènes *cse* et *orf1* par réassortiment de modules

Les régions HRC1 et HRC2 (pour homologous region, copy 1 et 2) sont représentées en noir. Les pointillés délimitent les bornes de la région dupliquée et son site d'insertion.

2.1.3.2 *Fonction de PcsB*

L'inactivation du gène codant PcsB (pour protein for cell separation) a été entreprise chez trois streptocoques : *S. agalactiae* (Reinscheid *et al.*, 2001), *S. mutans* (Chia *et al.*, 2001a), et *S. pneumoniae* (Ng *et al.*, 2004). Chez *S. mutans*, un mutant présentant une inactivation du gène *sagA*, codant PcsB (appelé GSP-781 ou SMU.22) a été obtenu. Chez *S. agalactiae*, l'obtention d'un mutant présentant une inactivation du gène *pcsB* a nécessité dans le milieu de culture la présence de sorbitol, une molécule osmoprotectante. En revanche, chez *S. pneumoniae*, il n'a pas été possible d'obtenir de mutant inactivé dans le gène *pcsB*, y compris lorsqu'un osmoprotectant tel que le sorbitol est ajouté dans le milieu de culture. Par conséquent, les auteurs ont opté pour la construction d'un mutant conditionnel dans lequel le promoteur sauvage de *pcsB* a été remplacé par un promoteur dont l'activité est modulable en fonction de la concentration en fucose. En conclusion, le gène *pcsB* est essentiel à la croissance de *S. pneumoniae* alors qu'il ne l'est pas chez *S. mutans* et *S. agalactiae*.

Les mutants *pcsB* de *S. mutans* et de *S. agalactiae*, et le mutant de *S. pneumoniae* pour lequel l'expression de *pcsB* est drastiquement diminuée, présentent des caractéristiques communes. Les trois souches croissent lentement en milieu liquide, et sont toutes affectées dans la morphologie cellulaire (Figure 65). Plus précisément, la forme des cellules est très hétérogène et il semble que les cellules se séparent selon de multiples plans. Il a également été observé, chez les mutants *pcsB* de *S. mutans* et *S. agalactiae*, que plusieurs plans de division se forment au sein d'une même cellule (Figure 66) (Chia *et al.*, 2001a; Reinscheid *et al.*, 2001). Une hypothèse simple expliquant ces modifications morphologiques est que ces mutants sont affectés dans la structure du peptidoglycane. Cette hypothèse est confortée par des expériences montrant que le mutant *pcsB* de *S. pneumoniae* est affecté dans le contrôle de la biosynthèse de peptidoglycane (Ng *et al.*, 2004). Cela pourrait expliquer pourquoi le mutant *pcsB* de *S. agalactiae* nécessiterait la présence d'un osmoprotectant pour croître : sa paroi ne serait plus assez rigide pour supporter la pression osmotique imposée par le milieu intracellulaire.

2.1.3.3 *Divergence fonctionnelle entre Cse et PcsB*

Les mutants Δcse de *S. thermophilus* ne semblent pas présenter les mêmes caractéristiques phénotypiques que les mutants *pcsB* des autres streptocoques. En effet, les mutants Δcse ne sont pas affectés drastiquement dans leur vitesse de doublement, et n'arborent pas de modification majeure de la morphologie cellulaire. Ces divergences phénotypiques suggèrent que PcsB et Cse ont des fonctions différentes. Les deux régions C-terminales, incluant le domaine CHAP, étant homologues, il est probable que cette divergence fonctionnelle entre PcsB et Cse résulte d'une spécialisation de l'activité des domaines CHAP, spécialisation qui serait assurée par la région N-terminale qui leurs sont associées.

2.1.3.4 *Recouvrement fonctionnel partiel entre PcsB et Cse*

Lorsque l'expression du gène *pcsB* de *S. pneumoniae* est faiblement abaissé, et non pas drastiquement abaissé comme présenté précédemment, *S. pneumoniae* se développe, tout comme la souche *S. thermophilus* Δcse , sous la forme de longues chaînes de cellules dont la

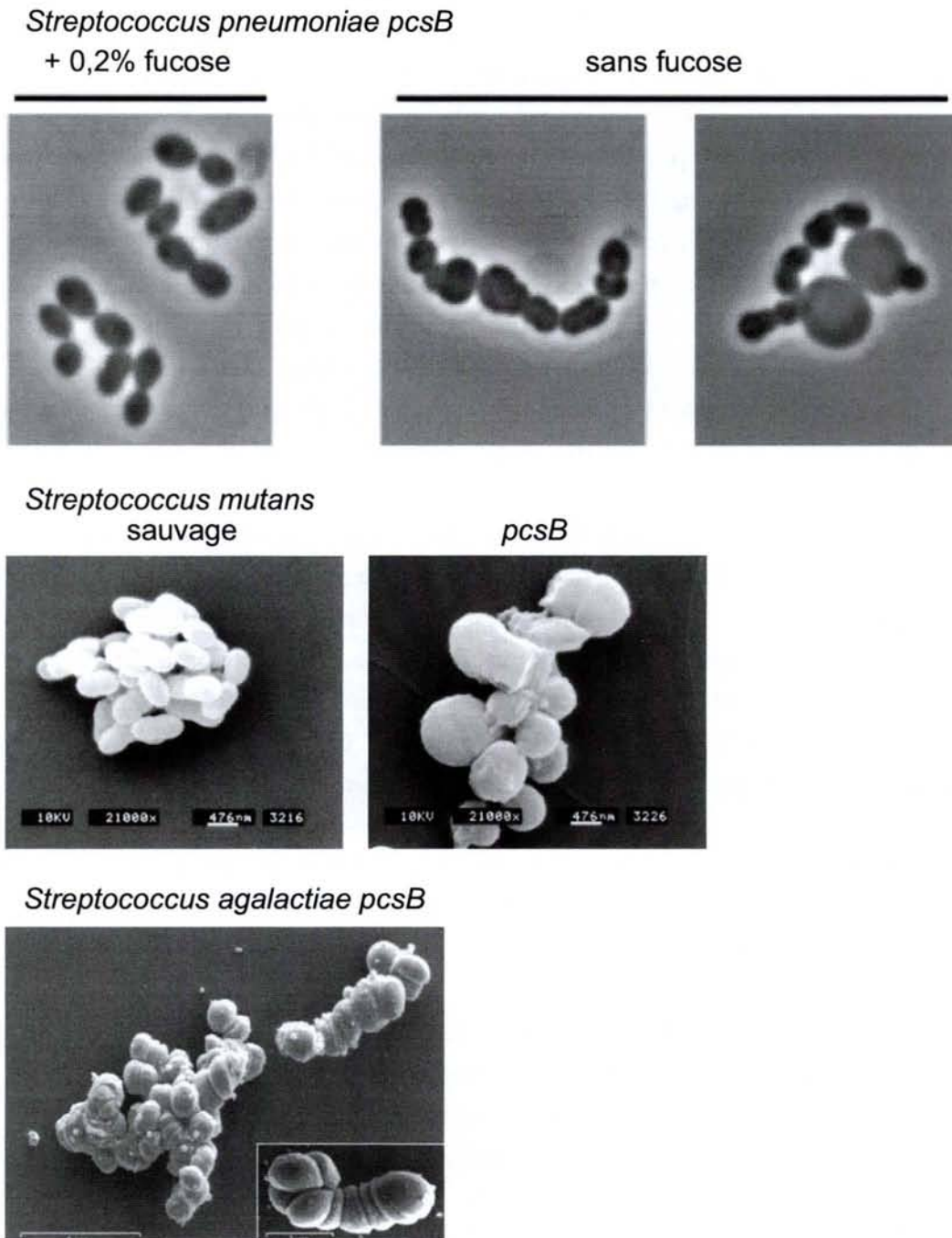


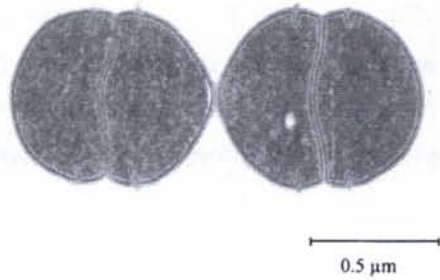
Figure 65 : Morphologie cellulaire de mutants *pcsB* chez trois streptocoques (d'après Ng et al., 2004 ; Chia et al., 2001 ; Reinscheid et al., 2001)

Les cellules de *S. pneumoniae* ont été photographiées en microscopie optique à contraste de phase à un grossissement de 1000 (les auteurs n'ont pas représenté l'échelle). La souche de *S. pneumoniae* photographiée est un mutant pour lequel le niveau d'expression du gène *pcsB* est sous le contrôle de la concentration en fucose du milieu de culture. Deux conditions sont présentées : en présence de 0,2% de fucose, le gène est exprimé, en absence de fucose, le gène est sous exprimé.

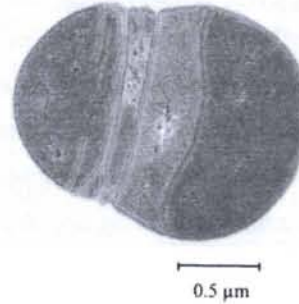
Les cellules de *S. mutans* et de *S. agalactiae* ont été photographiées en microscopie électronique à balayage. Chez ces deux streptocoques, les mutants *pcsB* ont été obtenus par inactivation de l'ORF *pcsB*.

Streptococcus agalactiae

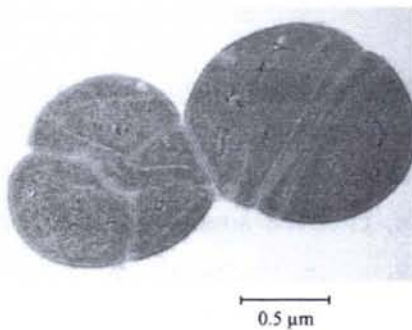
A.



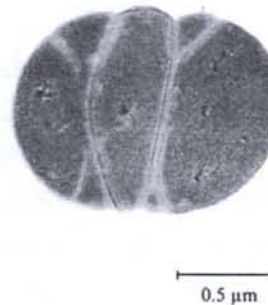
C.



B.



D.

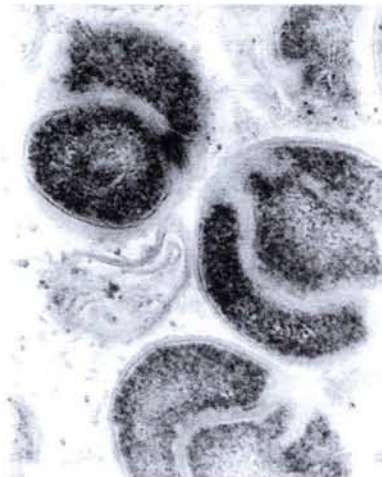


Streptococcus mutans

E.



F.



G.



Figure 66 : Observation des cellules de deux mutants *pcsB* en microscopie électronique à transmission (d'après Reinscheid *et al.*, 2001 ; Chia *et al.*, 2001)

Les clichés photographiques des souches sauvages de *S. agalactiae* et *S. mutans* sont présentés respectivement en A et E, ceux des mutants *pcsB* sont présentés en B, C, et D pour *S. agalactiae*, et en F et G pour *S. mutans*. Les clichés photographiques des souches de *S. mutans* ont été pris à un grossissement de 50000, les auteurs n'ont pas représenté l'échelle.

morphologie ne semble pas affectée (Ng *et al.*, 2004). Il est donc probable que les protéines Cse et PcsB présentent un recouvrement fonctionnel partiel.

Un recouvrement fonctionnel des protéines présentant des domaines CHAP avait déjà été suggéré par Ng *et al.* (2004) pour expliquer le caractère essentiel à la croissance du gène *pcsB* chez *S. pneumoniae*, et non essentiel chez *S. mutans* et *S. agalactiae*. Les auteurs ont constaté que le génome de *S. pneumoniae*, dont le gène *pcsB* est indispensable à la croissance, coderait en plus de PcsB une autre protéine avec un domaine CHAP. En revanche, le génome de *S. mutans* et celui de *S. agalactiae*, dont le gène *pcsB* est non indispensable à la croissance, coderaient 6 protéines avec un domaine CHAP. Ces données suggèrent un recouvrement fonctionnel partiel entre PcsB et une ou plusieurs des autres protéines présentant un domaine CHAP chez *S. mutans* et *S. agalactiae*.

Dans le cadre de cette hypothèse, Cse et PcsB seraient toutes deux impliquées dans la ségrégation cellulaire. En plus d'être impliqué dans la ségrégation cellulaire, la protéine PcsB serait impliquée dans la morphogénèse cellulaire, contrairement à Cse.

2.2 Mécanismes d'évolution de la région centrale du gène *cse*

2.2.1 Variabilité de la région *var-cse*

La région centrale de *cse* est quasi exclusivement constituée de motifs répétés. L'analyse de la séquence de cette région, nommée *var-cse*, chez 18 souches de *S. thermophilus*, a montré que elle est variable en séquence et en taille.

L'analyse détaillée des séquences *var-cse* a révélé au total 11 motifs répétés différents positionnés en tandem ou de façon dispersée. Cette caractéristique est peu communément décrite dans la littérature. En effet, les régions riches en motifs répétés sont généralement constituées de motifs répétés en tandem (Fischetti *et al.*, 1991; Dramsi *et al.*, 1993; van Belkum *et al.*, 1998). Certains gènes présentent cependant des régions riches en répétitions constituées de plusieurs motifs répétés différents positionnés de façon dispersée (Fahnestock *et al.*, 1986; Filpula *et al.*, 1987; Kehoe, 1994; Giraud et Cuny, 1997). Toutefois, à notre connaissance, la variété des motifs répétés décrits n'atteint jamais celle observée pour *cse*.

2.2.2 La variabilité : une conséquence de l'instabilité génétique résultant de la richesse en répétitions

L'alignement de la séquence de certains allèles révèle que la variabilité de la taille peut résulter de la variabilité du nombre de motifs répétés (Figure 67). Il est probable que les événements d'insertions et de délétions de motifs aient eu lieu par glissements de fourche de réplication ou par « single-strand annealing » (SSA), deux mécanismes de recombinaison illégitime (pour revue Michel, 1999). Ces deux mécanismes font intervenir un appariement de séquences complémentaires ectopiques, et induisent la délétion ou l'insertion de la région située entre les unités répétées ainsi que de l'une des deux unités répétées (Figure 14 et 15 pages 12 bis et 13 bis). Les alignements de séquences des régions *var-cse* sont compatibles avec ces deux mécanismes de recombinaison illégitime (Figure 67, les unités répétées ayant potentiellement été impliquées dans un appariement ectopique ont été surlignées en gris).

1)

		D		D		D		D
368.1.385	TACTATCTGAA	SAAGT	ACCAGTTGAA	SAAGCACCAGTTGAA	SAAGCACTAGTTGAA	SAAGCACCAGTTGAA		
A054	V L S E	E V P V E	E A P V E	E A L V E	E A P V E			

		D		D	
368.1.385	AGAAGCACCAGTTGAA	SAAGCACCAGCAGAA	CTAGCAGCTACATACGAATATGCATCAACTCCTAATACA		
A054	E A P V E	E A P A E	L A A T Y E Y A S T P N T		

2)

		A		B		A						E
308	ACT	SAACATCT	SCACAAGCT	SAACAACCT	SAAGTAGCAGCATACCAAGAAGCAGCAACA	TCAGAAGCTG						
6757.7.1066.Sfi6	T	E T S	A Q A	E T T	E V A A Y Q E A A T	S E A						

		E		C		C						H
308	CA	TCAGAAGCTGCA	TCAGAA	GCACCAGCTGCA	TCAGAAGCTCCTGCTAGC	GTAGTACCTGTCCGCAACA	TC					
6757.7.1066.Sfi6	A	S E A A A	S E A P A A	S E A P A S	V V P V A T S							

		E		C		C						C
308	AGAAGCTGTA	TCAGAAGCACCAGCTGCA						TCAGAAGT	GCCAGCAGAA	CAACTA		
6757.7.1066.Sfi6	E A V	S E A P A A						S E V	P A E Q L			

3)

		C		F		E		C
LMD9.1402	AGC	TCAGAAGCGCCAGCAGAA	CAACCAGCAGCAACAT	TCAGAAGCTGCA	TCAGAAGCTCCTGCTAGC	GTAG		
302.307	S	S E A P A E	Q P A A T	S E A A	S E A P A S V			

		H		E		C		C		F
LMD9.1402	TACCTGTCCGCAACA	TCAGAAGCT	GTA	TCAGAAGCACCAGCTGTA	TCAGAAGTGCCAGCAGAA	CAACTAGC				
302.307	V P V A T	S E A V		S E A P A V	S E V P A E Q L A					

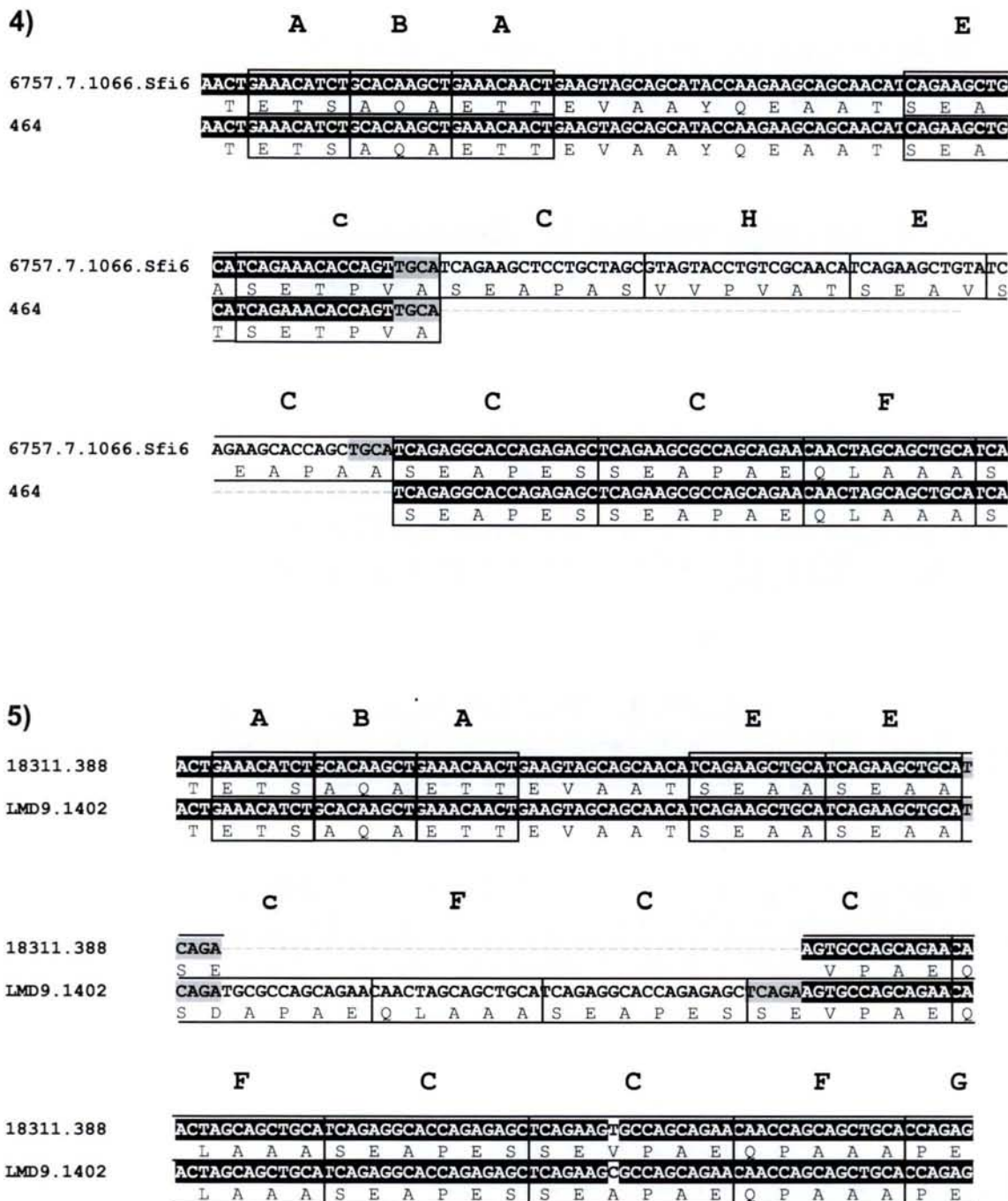


Figure 67 : Mise en évidence d'insertions et de délétions de motifs répétés

Les séquences nucléotidiques des régions *var-cse* des différentes souches de *S. thermophilus* ont été alignées grâce au logiciel CLUSTALW. Seuls les fragments d'alignements fiables et permettant de visualiser une divergence caractéristique d'événements d'insertion et/ou de délétion d'ADN sont présentés. Dans chacun des cas, deux alignements sont possibles concernant le positionnement des trous, un seul des deux alignements possibles est ici représenté. Les unités répétées ayant potentiellement été impliquées dans un appariement ectopique ont été surlignées en gris. Les résultats obtenus avec les séquences *var-cse* des souches CNRZ368 (368), NST1 (1), CNRZ385 (385), A054, CNRZ308 (308), IP6757 (6757), CNRZ7 (7), CNRZ1066 (1066), Sfi6, LMD-9, CNRZ1402 (1402), CNRZ302 (302), CNRZ307 (307), CNRZ464 (464), LMG18311 (18311), CNRZ388 (388) sont ainsi présentés. Les séquences nucléotidiques identiques sont surlignées en noir. En-dessous de chaque séquence nucléotidique est présentée la séquence protéique correspondante. Les motifs répétés sont encadrés et leur nom est indiqué au-dessus de l'encadré.

Les unités répétées, supposées avoir été sujettes à un appariement ectopique, peuvent constituer tout ou partie d'un même motif répété. Par exemple, l'événement d'insertion ou de délétion d'un motif E, révélé par l'alignement des régions *var-cse* présenté Figure 67 2), aurait pu faire intervenir un appariement ectopique de deux séquences, constituant chacune l'intégralité d'un motif E. En revanche, les appariements de séquences ectopiques pourraient avoir lieu entre des séquences constituant seulement une partie de motifs répétés (Figure 67 2),4),5)). Compte tenu des similarités constatées entre certains motifs (Figure 44 page 42 bis), il est par conséquent probable que des remaniements aient impliqué l'appariement de séquences ectopiques de motifs répétés différents. Cette hypothèse est soutenue par l'événement d'insertion ou de délétion révélé par l'alignement présenté Figure 67 3). En effet, ce remaniement aurait pu avoir lieu par appariement d'une partie d'un motif C avec une partie d'un motif E.

L'appariement partiel de motifs répétés pourrait non seulement être responsable d'insertions et de délétions d'un nombre entier de motifs répétés successifs, mais également de leur délétion partielle. En effet, certains motifs sont similaires entre eux. Par exemple, le motif G est entièrement compris dans un exemplaire du motif C (Figure 44 page 42 bis). Cette similarité suggère que ces deux motifs sont apparentés. Une hypothèse serait que le motif G aurait été généré par délétion partielle d'un motif C. Un scénario possible ferait intervenir un appariement ectopique des trois nucléotides GCA de l'extrémité 3' de l'unité répétée F qui précède le motif G (Figure 68). Cet appariement ectopique aurait engendré la délétion de la région codant la séquence protéique SEA du motif ancestral C, pour générer le motif G. Il doit être noté que ce scénario explique comment a pu être généré le motif G, mais n'explique pas comment celui-ci, qui n'est toujours pas un motif répété au terme de ce scénario, a été dupliqué.

Les allèles *var-cse* de 11 souches de *S. thermophilus* présentent une répétition d'une succession d'unités répétées. Dans le cas de la souche LMG18311, CNRZ388, LMD-9, CNRZ1402, CNRZ302 et CNRZ307 c'est la succession de motifs CFCCFGCFE, de 140 pb, qui est dupliquée de façon dispersée. Deux hypothèses peuvent expliquer l'existence de cette grande duplication. La première est que la même succession est apparue deux fois de façon indépendante suite à une succession de plusieurs événements d'insertion et de duplication de motifs répétés. La seconde est que l'une des deux unités a été entièrement dupliquée. Il semble que la seconde hypothèse soit plus favorable. En effet, les deux unités répétées qui mesurent chacune 140 pb, ne divergent que d'un nucléotide. Cette faible divergence contraste avec la haute dégénérescence observée pour certains motifs (Figure 42 A). Notamment, le motif C est le plus dégénéré de tous les motifs identifiés : 8 positions sur 18 sont dégénérées. Ainsi, il est vraisemblable que la répétition de deux exemplaires de la succession CFCCFGCFE soit apparue suite à un événement de duplication d'une des deux copies.

Il est à noter que cette grande duplication aurait généré une duplication du motif G, motif qui aurait été généré par délétion partielle d'un motif ancestral C selon le scénario présenté au paragraphe précédent. L'hypothèse de la duplication de la succession CFCCFGCFE, contenant un motif G, expliquerait comment le motif G, initialement en copie unique, aurait été dupliqué. Cette hypothèse est favorisée par la présence exclusive du motif G dans la succession CFCCFGCFE. Un mécanisme équivalent de génération de motifs répétés avait été

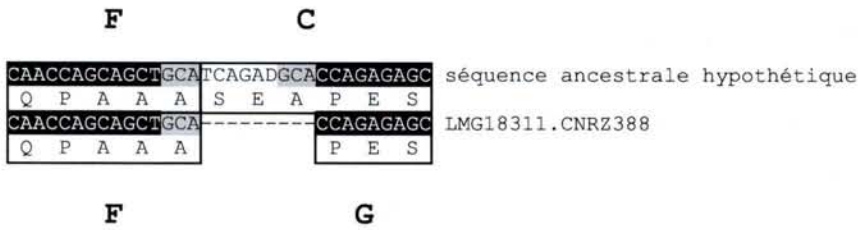


Figure 68 : Comparaison de séquences permettant d'illustrer un mécanisme hypothétique de génération du motif G

Alignement d'un fragment de la séquence de la région *var-cse* des souches *S. thermophilus* LMG18311 et CNRZ388 avec une séquence ancestrale hypothétique. Les nucléotides identiques entre les deux séquences sont surlignés en noir. La séquence de trois nucléotides dupliqués dans la séquence ancestrale et qui ne l'est pas dans la séquence des souches de LMG18311 et CNRZ388 est surlignée en gris. Dans cette hypothèse, le motif G aurait été généré par délétion d'un fragment du motif ancestral C, suite à un événement de recombinaison illégitime faisant intervenir l'appariement ectopique de deux copies d'une séquence répétée de trois nucléotides GCA.

suggéré suite à l'analyse de la séquence de la protéine G de deux souches de streptocoques de groupe G (Figure 13 page 11 bis) (Fahnestock *et al.*, 1986; Filpula *et al.*, 1987).

2.2.3 Implication du transfert horizontal

Les analyses de séquences montrent que *cse* est un gène mosaïque : il est en effet divisible en 5 parties nommées de A à E sur le critère d'identités de séquences entre les allèles *cse* de différentes souches de *S. thermophilus* (Figure 49 page 44 bis). Sur ce critère de comparaison, les allèles *cse* de *S. thermophilus* peuvent être classés en deux groupes : les groupes I et II (pour la définition des groupes, voir la légende de la Figure 42 B page 41 quinto). En effet, les séquences au sein d'un même groupe présentent entre elles environ 98% d'identité sur les régions A, B, D et E, et au moins 70% d'identité au niveau de la région C (Figure 48 page 43 quarto). En revanche, lorsqu'on compare les séquences du groupe I avec celles du groupe II, le pourcentage d'identité est également de 98% environ dans les régions A et E mais il chute à 93% environ dans les régions B et D. De plus, entre les deux groupes les séquences présentent non pas 70% d'identité au moins, mais moins de 52% d'identité dans la région C. Ainsi, les séquences des groupes I et II divergent plus des séquences de l'autre groupe que des séquences du même groupe sur les régions BCD. Cette distinction en deux groupes n'est en revanche plus possible sur les régions A et E.

De nombreux gènes mosaïques ont été décrits dans la littérature, notamment chez les streptocoques (Whatmore et Kehoe, 1994; Kapur *et al.*, 1995; Whatmore *et al.*, 1999; Filipe *et al.*, 2000; Hakenbeck, 2000), et en particulier chez *S. thermophilus* (P. Renault, communication personnelle). Ces gènes sont formés par recombinaison allélique entre un locus du génome et un allèle de ce même locus acquis par transfert horizontal. Il est probable que le caractère mosaïque de *cse* résulte de ce mécanisme. En effet, un événement de recombinaison aurait pu engendrer le remplacement des régions BCD chez l'ancêtre des souches du groupe I ou II, par celle d'un gène *cse* acquis par transfert horizontal et provenant d'une autre espèce de streptocoque (Figure 69). En effet, il a été montré que les séquences des souches de *S. thermophilus* divergent entre elles de 0,2% (P. Renault, communication personnelle) (Bolotin *et al.*, 2004). Or, les séquences du groupe I et du groupe II divergent de 7% environ au niveau des régions B et D, régions ne présentant *a priori* pas de structure favorisant l'instabilité génétique (telle que des répétitions). En conclusion, il est probable que la région BCD du gène *cse* des souches du groupe I ou du groupe II provienne d'un autre streptocoque et qu'elle ait été acquise par transfert horizontal. Une hypothèse, concernant le streptocoque donneur, serait qu'il appartienne aux espèces *Streptococcus vestibularis* ou *Streptococcus salivarius*. En effet, le degré de divergence entre les gènes des souches de *S. salivarius*, de *S. vestibularis* et de *S. thermophilus* est de l'ordre de 7% (P. Renault, communication personnelle).

Ce remplacement implique que le streptocoque donneur présente le gène chimérique *cse* au sein de son génome, et donc par conséquent que celui-ci soit proche de *S. thermophilus*. Cette hypothèse est corroborée par le faible taux de divergence (7%) entre les séquences du groupe I et celles du groupe II, dans les régions B et D.

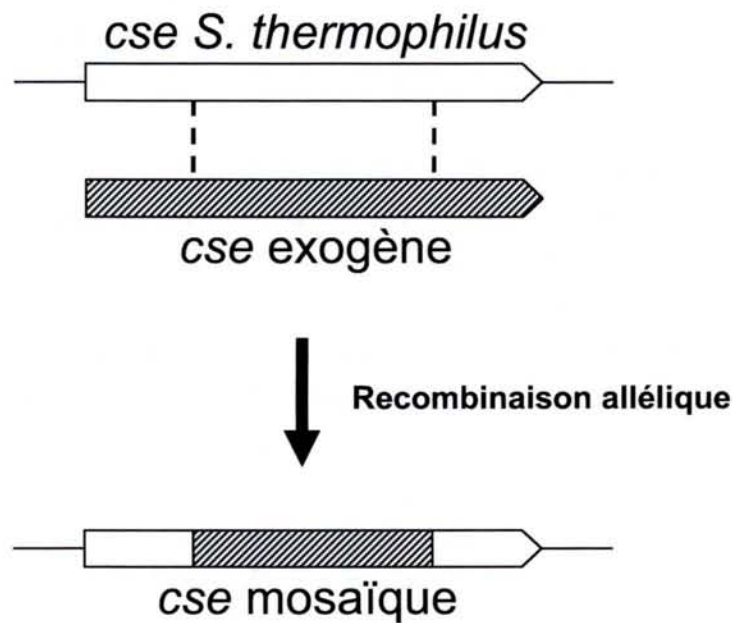


Figure 69 : Génération du gène mosaïque *cse*

Une souche de *S. thermophilus* aurait acquis par transfert horizontal un gène *cse* provenant d'une autre espèce du genre *Streptococcus* (appelé *cse* exogène). La recombinaison allélique entre l'allèle chromosomique résident et l'allèle exogène entraîne la formation d'un allèle *cse* mosaïque.

Les ORF *cse* sont représentées par une flèche, les régions chromosomiques flanquant l'ORF *cse* de la souche de *S. thermophilus* sont représentées par des traits fins. Les bornes de recombinaison sont indiquées par des traits interrompus.

2.3 Histoire évolutive de *cse*

A la lumière de l'ensemble de ces résultats il est possible d'imaginer un scénario décrivant l'histoire évolutive de *cse*. L'ancêtre des streptocoques aurait présenté un gène ancestral *sip* toujours présent dans le génome de *S. pneumoniae*, *S. agalactiae*, *S. pyogenes* et *S. suis*. En revanche, un remaniement chromosomique aurait engendré l'interruption du gène *sip* et la formation du gène chimérique *cse* chez l'ancêtre commun à *S. thermophilus* et d'autres streptocoques phylogénétiquement proches (Figure 64). Ces autres espèces sont encore non identifiées, mais il est probable que *S. salivarius* et *S. vestibularis*, deux espèces très proches de *S. thermophilus*, en face partie. L'instabilité de la région riche en répétitions aurait entraîné une évolution rapide de la région centrale de *cse* depuis sa création.

Un fragment d'ADN contenant le gène *cse*, provenant d'un streptocoque proche de *S. thermophilus*, aurait été acquis par une souche de *S. thermophilus* (Figure 69). Ce transfert horizontal aurait impliqué le remplacement d'un fragment du gène *cse* résidant, par le fragment équivalant exogène. La conséquence de ce remplacement aurait été la formation d'un gène *cse* mosaïque.

2.4 Importance des régions de faible complexité et tolérant la variabilité dans les réassortiments de domaines

2.4.1 Exemples de protéines de surface ayant été potentiellement générées par réassortiment de domaines

2.4.1.1 *La protéine Cse*

La région HRC1 est flanquée par deux régions riches en répétitions : la région *var-cse* et une autre région localisée dans le gène *orf1*. Tous les motifs identifiés dans la région riche en répétitions de *orf1* sont communs à ceux identifiés dans la région *var-cse* des souches de *S. thermophilus* du groupe II (Figure 51 page 45 bis). De plus, la région riche en répétitions de *orf1* est constituée de la répétition en tandem de la succession de motifs ECCF, succession identifiée à plusieurs reprises dans la région *var-cse* de plusieurs souches du groupe II. Cette région présente également une séquence de faible complexité.

La région HRC2 quant à elle jouxte une région de faible complexité, également riche en répétitions, qui est localisée en amont d'HRC2, dans l'ORF *pcsB* (Figure 51 page 45 bis). Cette région, également variable au niveau intraspécifique, est constituée de motifs répétés communs à ceux identifiés dans la région *var-cse*.

Dans le cadre de l'hypothèse de la génération de *cse* par réassortiment de modules, ces résultats suggèrent qu'une des bornes de la duplication était localisée dans une région ancestrale variable et riche en répétitions du gène *pcsB* (Figure 70). Ils suggèrent également que l'insertion de la copie dupliquée a eu lieu au sein d'une région variable et riche en répétitions du gène ancestral *sip* (Figure 70).

Si les régions centrales des gènes *pcsB* et *sip* contiennent une région de faible complexité, variable et riche en répétitions, alors il est possible que leur gène ancêtre présentait également ces mêmes caractéristiques.

L'analyse par DOT PLOT de l'ORF *sip* de *S. agalactiae* ainsi que celles qui lui sont homologues chez *S. pyogenes*, *S. pneumoniae* et *S. suis* n'a pas révélé l'existence de région

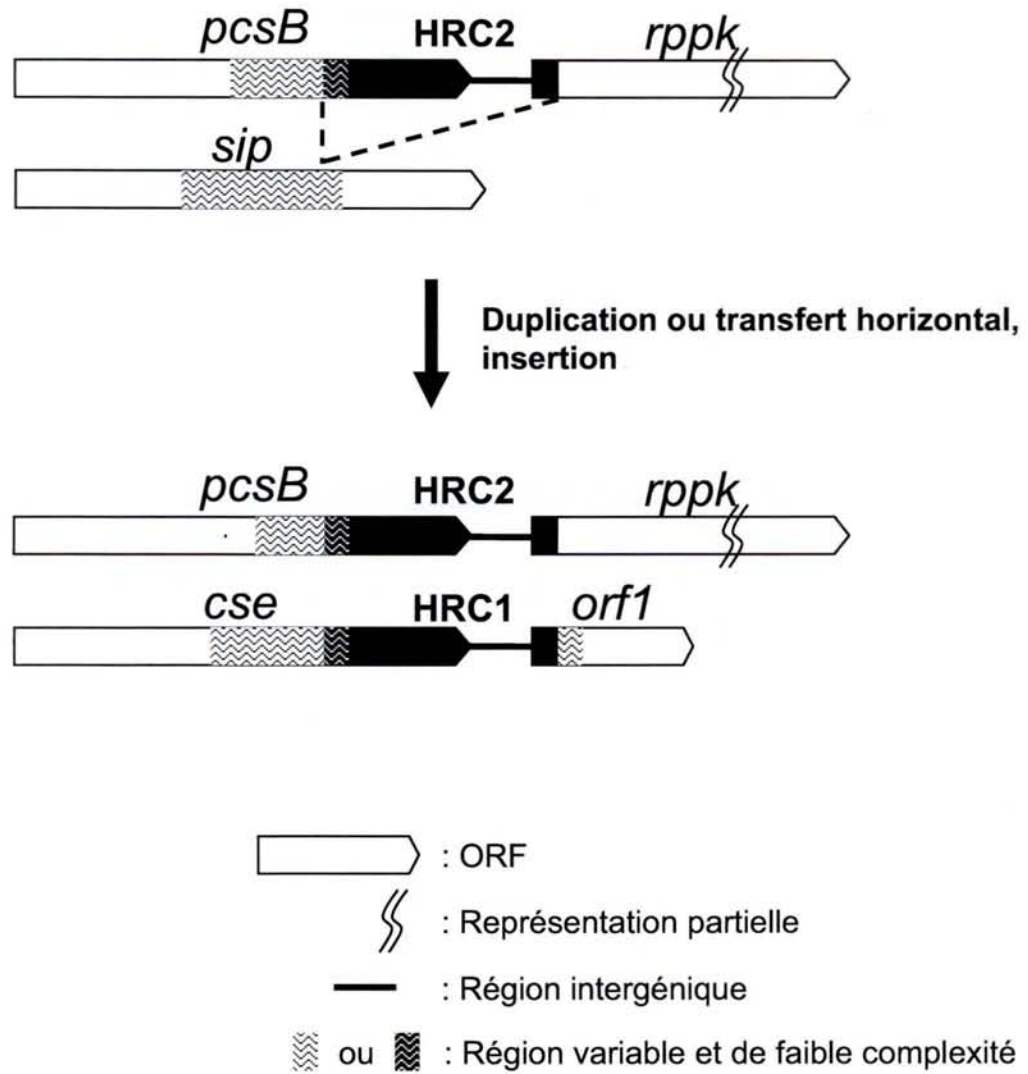


Figure 70 : Localisation des bornes du réarrangement à l'origine de la génération des gènes *cse* et *orf1*

Les régions HRC1 et HRC2 (pour homologous region, copy 1 and 2) sont représentées en noir. Les pointillés délimitent les bornes de la région dupliquée et son site d'insertion.

riche en répétitions. En revanche, l'alignement de la séquence de ces ORF, ou de leur produit, montre qu'à défaut d'être hautement divergentes au niveau intraspécifique, les régions centrales sont très divergentes au niveau interspécifique (Figure 71). De plus, l'analyse de ces séquences avec logiciel SEG montre que ces régions centrales présentent une séquence de faible complexité (résultat non montré). Les mêmes résultats sont obtenus lorsque les séquences de *pcsB* et de leur produit sont analysées (Figure 72). Compte tenu de la variabilité intraspécifique de la région centrale des gènes *sip* et *pcsB* actuel, il est probable que cette même région était également variable et de faible complexité au moment où le réarrangement à l'origine de *cse* a eu lieu (Figure 70).

L'ensemble de ces données suggèrent que certaines bornes du réarrangement d'ADN à l'origine du gène chimérique *cse*, étaient localisées à l'intérieur de régions variables et présentant une séquence de faible complexité.

2.4.1.2 La protéine PcsB ou SagA

Comme mentionné précédemment, tous les génomes entièrement séquencés de streptocoques, ainsi que celui de *L. lactis*, codent une protéine homologue de PcsB. Le génome d'*E. faecium* code également une protéine homologue de PcsB, nommée SagA (pour secreted antigen A) (Teng *et al.*, 2003). Toutefois, seule la région N-terminale de la protéine présente des similarités significatives (42% d'identité, 64% de similarité) avec les protéines PcsB (Figure 73). Pourtant, à l'instar des gènes *pcsB* de *S. mutans*, *S. thermophilus*, *S. pneumoniae* et *L. lactis*, le gène *sagA* est précédé de l'opéron putatif *mreCD*. Cette organisation commune, en plus des similarités de séquence, suggère que les régions 5' des gènes *sagA* et *pcsB* sont orthologues. Quant à l'extrémité C-terminale de SagA, celle-ci présente des similarités significatives (63% d'identité, 76% de similarité) avec un fragment d'une autre protéine : une muramidase codée par le génome du même organisme, *E. faecium*. Les régions homologues entre ces protéines contiennent l'intégralité d'un domaine NPLC/P60. Ce domaine, souvent présent dans des lipoprotéines, est de fonction inconnue. Ainsi, SagA ne présente pas de domaine CHAP comme PcsB, mais présente à la place un autre domaine, de fonction inconnue, ne présentant pas d'identité significative avec les domaines CHAP. Il est donc probable que PcsB ou SagA soit une protéine chimérique ayant été générée par réassortiment de domaines.

L'analyse de la séquence de la protéine SagA avec le logiciel SEG montre que la séquence localisée entre le domaine N-terminal homologue de PcsB, et le domaine C-terminal homologue de la muramidase, est de faible complexité. De plus, la séquence est riche en motifs répétés (Teng *et al.*, 2003). Elle contient deux motifs répétés différents de douze résidus d'acides aminés. L'un d'eux est présent en 4 exemplaires et l'autre en 3. Une recherche effectuée avec le logiciel BLAST a permis d'identifier des protéines homologues de SagA présentant les deux mêmes types de domaines N- et C-terminaux. L'alignement montre que les régions centrales de ces protéines sont très divergentes contrairement aux régions N- et C-terminales, traduisant la variabilité de la région centrale (Figure 74).

Ainsi, les protéines PcsB et SagA présentent chacune deux domaines reliés l'un à l'autre par une séquence variable de faible complexité. Or, un réarrangement génomique, ayant pour conséquence la génération d'une protéine par réassortiment de domaine, nécessite qu'une

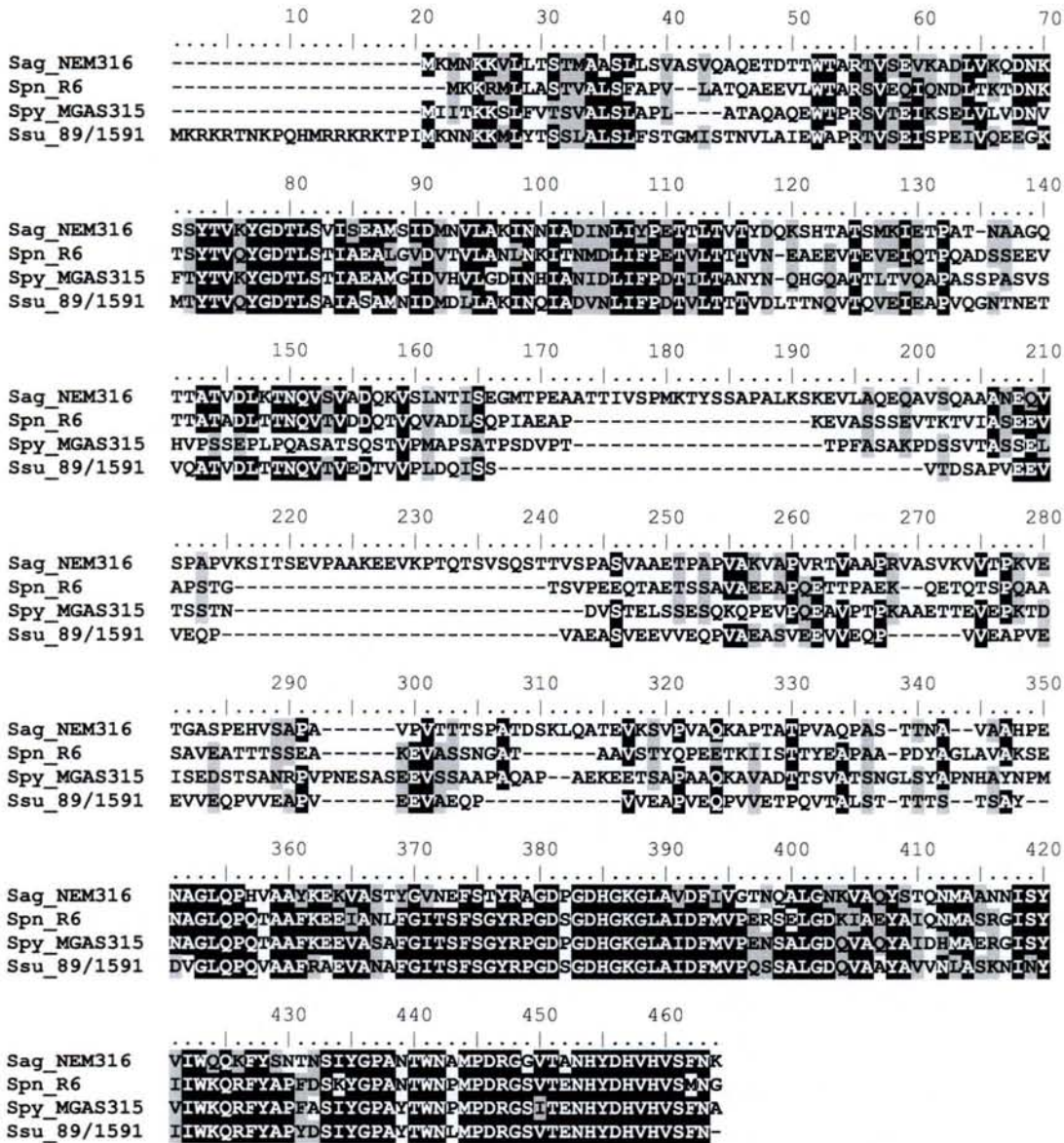


Figure 71 : Alignement de la protéine SIP de *S. agalactiae* avec ses homologues les plus proches chez trois autres streptocoques

L'alignement a été construit grâce au logiciel CLUSTALW. Les résidus sont surlignés en noir lorsque 75% des séquences présentent le même résidu d'acides aminés à la même position, et en gris lorsque 75% des résidus d'acides aminés sont chimiquement équivalents. Les séquences sont accessibles grâce aux numéros d'accès suivants : NP_734501 pour celle de *S. agalactiae* NEM316 (Sag_NEM316), NP_359466 pour celle de *S. pneumoniae* R6 (Spn_R6), NC_004070 pour celle de *S. pyogenes* MGAS315 (Spy_MGAS315) et ZP_00333146 pour celle de *S. suis* 89/1591 (Ssu_89/1591).

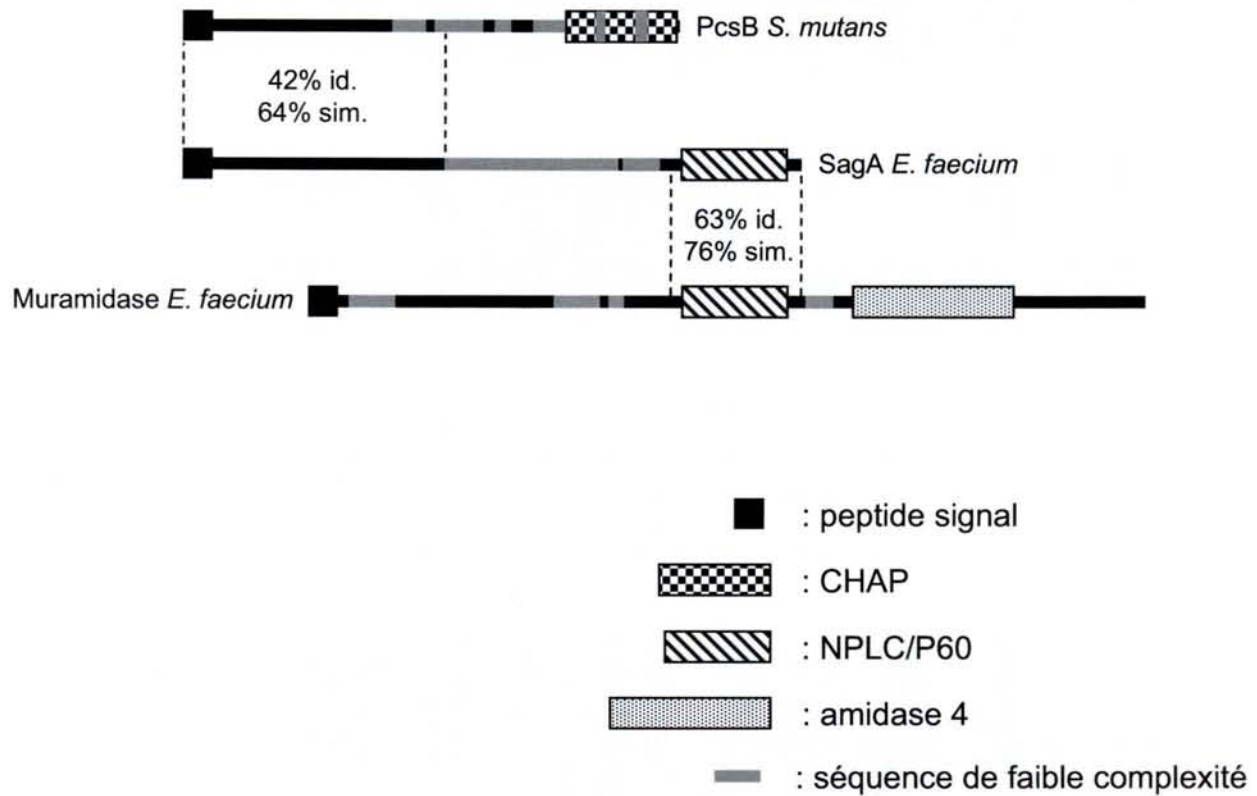


Figure 73 : SagA ou PcsB est une protéine chimérique

Les régions présentant des identités significatives sont reliées entre elles par des traits interrompus. Les valeurs d'identité (id.) et de similarité (sim.) sont mentionnées. Les domaines ont été prédits grâce à CDD et les régions de faible complexité ont été détectées grâce au logiciel SEG.

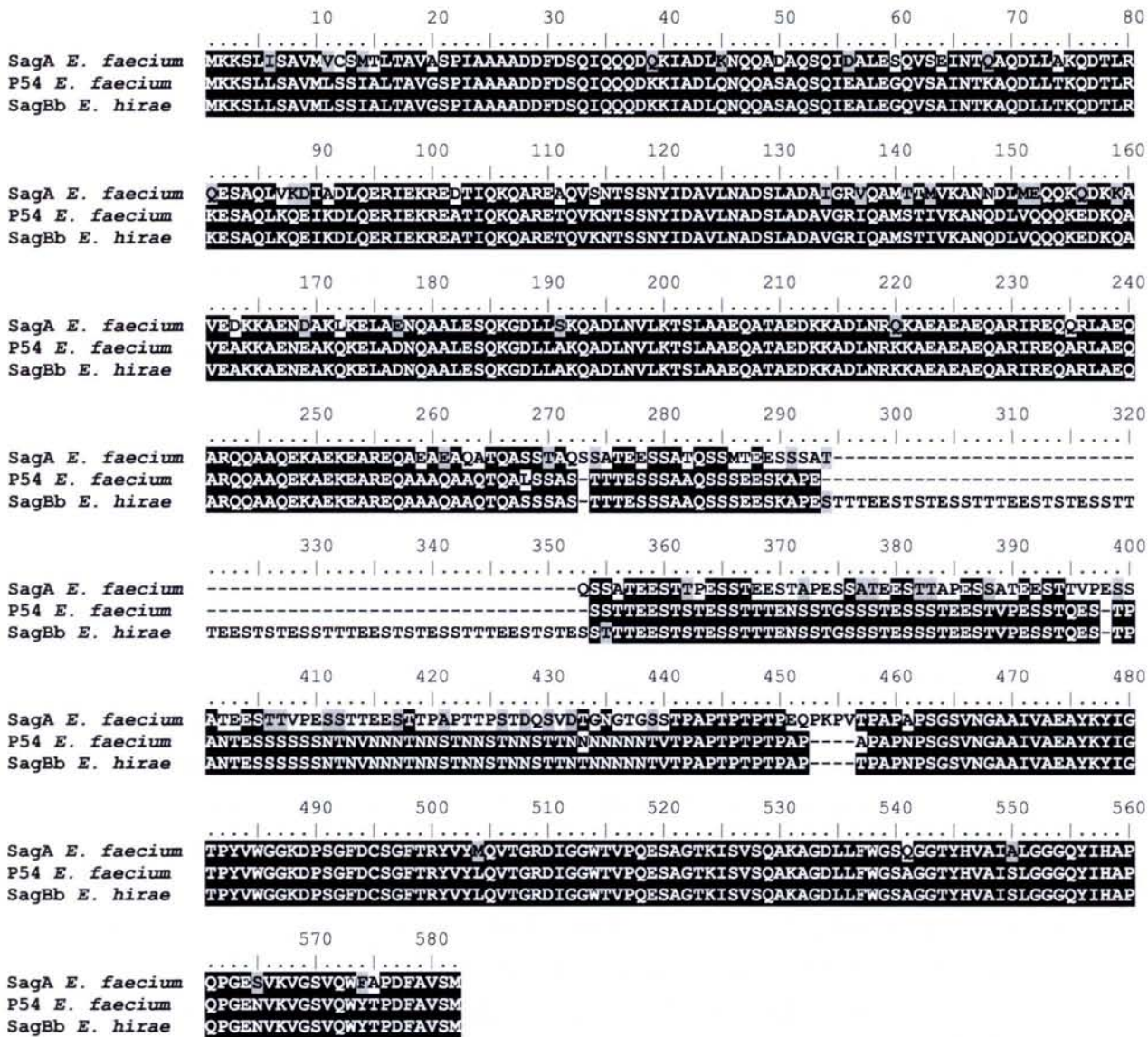


Figure 74 : Alignement de SagA avec les proteines qui lui sont homologues sur les regions N- et C-terminales

L'alignement a été construit grâce au logiciel CLUSTALW. Les résidus sont surlignés en noir lorsque 2/3 des séquences présentent le même résidu d'acides aminés à la même position, et en gris lorsque 2/3 des résidus d'acides aminés sont chimiquement équivalents. Les séquences sont accessibles grâce aux numéros d'accession suivants : AAF86217 pour SagA, P13692 pour P54 et AAF87093 pour SagBb.

borne soit localisée entre les séquences codant les domaines réassortis. Par conséquent, il est probable qu'une des bornes du réarrangement supposé à l'origine de la protéine PcsB ou SagA était localisée à l'intérieur d'une région variable de faible complexité.

2.4.1.3 La protéine APF1

Une protéine de surface chimérique a été identifiée chez *Lactobacillus johnsonii* (Turner *et al.*, 2004). La région C-terminale de la protéine APF1 présente 86% d'identité avec la région C-terminale de la protéine Sep de *Lactobacillus fermentum* (Figure 75). En revanche, les autres régions de ces deux protéines ne présentent pas d'identité significative entre elles. De plus, la protéine Sep présenterait dans sa région N-terminale un domaine LysM que ne présenterait pas la protéine APF1. Ces données suggèrent que la protéine Sep, ou que la protéine APF1, a été générée par réassortiment de domaines.

Les génomes d'*E. faecium* et de *S. pneumoniae* présentent chacun un gène codant une protéine homologue des régions N- et C-terminales de la protéine Sep de *L. fermentum*. En revanche, ces mêmes génomes ne présenteraient pas d'homologue du gène codant APF1. Ainsi, il est probable que l'ancêtre commun aux bactéries du genre *Lactobacillus*, *Enterococcus* et *Streptococcus*, ait présenté un gène codant une protéine Sep. La protéine APF1, serait ainsi une protéine chimérique, codée par un gène chimérique issu de la fusion de la région 3' d'un gène ayant codé une protéine Sep ancestrale, à un fragment d'un autre gène.

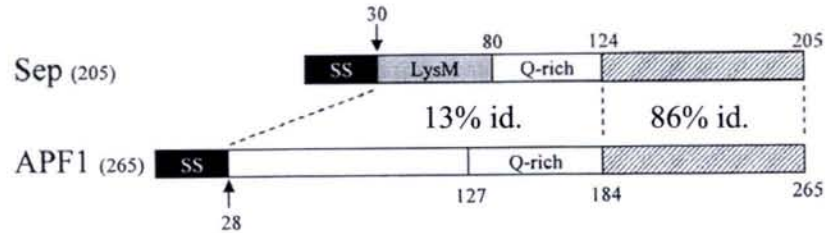
La région centrale des gènes codant des protéines Sep et APF1 ne présente pas de motifs répétés. Toutefois, ces deux protéines présentent dans leur région centrale, une séquence de faible complexité riche en résidus glutamine. De plus, l'alignement de deux protéines d'*E. faecium* et d'une protéine d'*E. faecalis*, toutes homologues de Sep, avec la protéine Sep de *L. fermentum*, révèle que la région centrale est très divergente comparativement aux régions N- et C-terminales (Figure 76). La région centrale de la protéine Sep serait par conséquent variable.

L'analyse de l'alignement des protéines Sep et APF1 montre que leurs régions C-terminales présentent entre elles des similarités significatives jusqu'à la limite C-terminale de leur région centrale de faible complexité respective (Figure 75 B). Par conséquent, il semble alors possible que les fragments de gènes aient été fusionnés au niveau de régions variables et de faible complexité pour former le gène chimérique codant la protéine APF1. La région centrale du gène codant APF1 est une région de faible complexité. Or, les régions de faible complexité ont tendance à évoluer rapidement (Wootton et Federhen, 1996). Ainsi, il est probable que la région centrale du gène codant APF1 soit variable. La région centrale de ce gène étant probablement variable, la borne de la fusion aurait pu être effacée suite à de multiples mutations.

2.4.2 Les régions variables et de faible complexité pourraient favoriser la génération de nouvelles protéines par réassortiments de domaines

La création de nouvelles protéines par réassortiments de domaines peut théoriquement permettre de générer une nouvelle fonction, grâce à un unique événement mutationnel, en recombinant des fragments de gènes entre eux. Toutefois, pour être efficace, la localisation des bornes des réarrangements ne doit pas induire d'inactivation de la fonction propre à

A.



B.

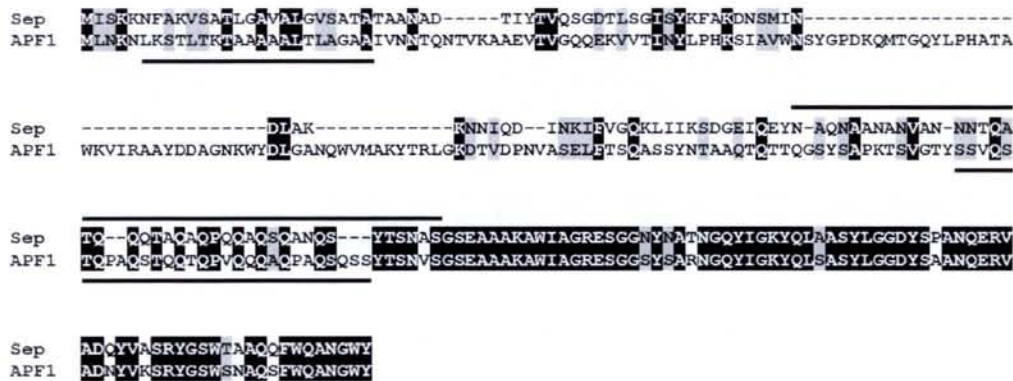


Figure 75 : Identité partielle des deux protéines Sep et APF1 (d'après Turner, 2004)

A. Représentation schématique des protéines Sep de *Lactobacillus fermentum* et APF1 de *Lactobacillus johnsonii*. Le pourcentage d'identité (id.) entre les régions délimitées par des pointillés est indiqué. Les positions sont indiquées en acides aminés.

B. Alignement des séquences Sep et APF1 obtenu grâce au logiciel CLUSTALW. Les positions identiques sont surlignées en noir, celles chimiquement équivalentes sont surlignées en gris. Les régions de faible complexité détectées par le logiciel SEG sont indiquées par des traits noirs épais.

```

      10      20      30      40      50      60      70      80
L. fermentum MISKRNFAKVSATLGAVALGVSATATAANADTIYTVQSGDTLSGISYKFAKDNSMINDLAKKNNIQDINKLIVGQKLIHK
E. faecium   MTSIKRILLFGTTLAAGAAF---FMCTTAFADEAYTVQSGDTLSTISQKYVGDNSLINAIAESNSISDINLIYSGOQLTIP
E. faecalis  MKSIKRTILLGTTLAAGLGL---FLGIDANAESLYTVKAGDTLSTISHQFAGDNSLQKIASDNKLPNLDLIFECEQLVIR

      90      100     110     120     130     140     150     160
L. fermentum SD--GEIQEYNAQNAANANVANN---NTQATQOQTAAQAPQ-----QAOSQANQSYTSNAGSEAAAKAWIAGRESGQNY
E. faecium   TEGSAQAAAEQAAVQEAPEVQAEQVQAEQPVVQETVQETQAAAPVAETQPAPAVTETAATPASTSSAKEWIAQKESGCSY
E. faecalis  SE--KEVANTPAPAVEVAPVQVQV---VEQVPAQP-VQEEVQ-----QPVAQEVAPAPAPAASSDAKEWIAQRESGCSY

      170     180     190     200     210
L. fermentum NATNGQYIGRYQLAASYLNGDYSANQERVAQYVASRYGSWTAAQQFWQANGWY
E. faecium   TATNGRYIGRYQLDSSYLNQDYSANQERVAQYVTSRYGSWEAAKAFWEANGWY
E. faecalis  DATNGQYIGRYQLSASYLNGDYSANQERVAQYVAGRYGSWDAAKSFWLANGWY

```

Figure 76 : Alignement de Sep avec les proteines qui lui sont homologues sur les regions N- et C-terminales

L'alignement a été construit grâce au logiciel CLUSTALW. Les résidus sont surlignés en noir lorsque 100% des séquences présentent le même résidu d'acides aminés à la même position, et en gris lorsque 100% des résidus d'acides aminés sont chimiquement équivalents. Les séquences sont accessibles grâce aux numéros d'accension suivants : AAS55430, ZP_00285760 et NP_814228 respectivement pour les protéines de *L. fermentum*, *E. faecium* et *E. faecalis*.

chaque domaine. La fonctionnalité d'un domaine aura notamment plus de chance de rester intacte si le site de fusion n'est pas localisé à l'intérieur des domaines. En effet, les contraintes de structure pouvant être élevées, un domaine tronqué ou chimérique aura de fortes chances d'être non fonctionnel. Par conséquent, il est plus favorable que les sites de fusion soient localisés à l'intérieur des charnières interdomaines ou alors à l'intérieur de domaines protéiques dont la fonction impose de faibles contraintes sur la conservation de la séquence.

Peu de protéines ou gènes chimériques bactériens sont décrits de façon détaillée du point de vue des mécanismes génétiques à l'origine de leur formation. Nous décrivons ici trois protéines chimériques supposées avoir été générées par réassortiment de domaines : les protéines Cse, APF1 et PcsB (ou SagA). Ces protéines ont plusieurs caractéristiques communes. Elles présenteraient toutes une structure quadripartite : un peptide signal suivi de deux domaines espacés par une région variable et de faible complexité. De plus, dans le cas des protéines Cse, PcsB et SagA, la région variable et de faible complexité est également riche en répétitions. Une autre caractéristique commune à ces 4 protéines est qu'elles seraient extracellulaires et toutes, sauf la protéine APF1, liées ou potentiellement liées à la surface des cellules. La protéine APF1 ne présente pas de domaine connu d'attachement à la paroi ce qui n'exclut pas qu'elle puisse en présenter un encore non identifié. Lopez *et al.* (2000) ont proposé que les protéines de surface des bactéries à Gram positif évoluent fréquemment par réassortiment de domaines. Deux hypothèses non exclusives permettraient d'expliquer le succès évolutif de ce mécanisme de génération de nouvelles fonctions à la surface des cellules des bactéries à Gram positif.

La première serait que l'interdépendance fonctionnelle de domaines présents au sein d'une même protéine serait faible. Par exemple, certaines protéines de surface sont constituées d'un domaine de liaison à la paroi et d'un domaine catalytique. Dans de nombreux cas, il a été montré que chacun des domaines pouvait assurer son rôle propre indépendamment du domaine qui lui est associé (Diaz *et al.*, 1991; Buist *et al.*, 1995; Rashid *et al.*, 1995; Mesnage et Fouet, 2002; Huard *et al.*, 2003; Steen *et al.*, 2003). Ainsi, une nouvelle combinaison de domaines peut potentiellement être fonctionnelle immédiatement après sa génération, contrairement à certains domaines pouvant nécessiter une co-adaptation, comme les domaines de certaines protéines régulatrices (Juarez *et al.*, 2000).

La deuxième hypothèse émane de l'étude des protéines chimériques Cse, APF1 et PcsB (ou SagA). Nous proposons que les événements de réassortiments de domaines sont facilités par la présence de séquences variables et de faible complexité. En effet, l'existence de ces séquences pourrait augmenter la probabilité pour que les bornes des réarrangements de gènes ne soient pas localisées à l'intérieur de modules codant un domaine. Si la probabilité pour que ces bornes aient une localisation respectant l'intégrité des domaines est augmentée, alors la probabilité pour que des réassortiments de domaines efficaces aient lieu sera également plus importante. Dans le cadre de cette hypothèse, plus la région variable et de faible complexité sera grande et plus la probabilité pour qu'ait lieu un réassortiment de domaines efficace augmentera.

Or les protéines de surface des bactéries à Gram positif présenteraient fréquemment des régions variables (Kehoe, 1994; van Belkum *et al.*, 1998). La présence fréquente de ces régions répétées variables dans les protéines de surface pourrait favoriser la génération de

nouvelles protéines par réassortiment de domaines. Ces régions variables et riches en répétitions pourraient ainsi être un facteur à l'origine du succès de ce mécanisme dans l'évolution des protéines de surface des bactéries à Gram positif.

Chez les eucaryotes, les réassortiments de domaines seraient favorisés par la structuration des gènes en introns et en exons ainsi que par la coïncidence des bornes des exons avec celles des domaines protéiques (van Belkum *et al.*, 1998). Les introns constitueraient des régions favorables à l'accueil de bornes de réarrangements responsables de réassortiments d'exons, car la fonction d'intron impose peu de contraintes de conservation de la séquence (Gilbert, 1978; Patthy, 1991; 1994). Les régions intragéniques variables et de faible complexité présentent des similitudes avec les introns eucaryotes. En effet, les introns et les régions variables de faible complexité ont des fonctions tolérant d'importantes modifications. Elles peuvent tolérer en particulier des modifications telles que des délétions, des insertions ou alors la formation de chimères, qui sont des mutations inhérentes aux réarrangements de fragments de gènes.

On peut alors se demander si les régions intragéniques variable et de faible complexité ne joueraient pas le même rôle que les introns dans la création de gènes par réassortiment de modules.

3 Facteurs influençant l'adhésion de *S. thermophilus* à une surface non biotique

3.1 Introduction

Les bactéries peuvent vivre sous la forme de biofilms, un biofilm étant défini comme une communauté bactérienne, généralement incluse dans une matrice de polymères organiques et adhérant à une surface (Costerton *et al.*, 1995). La capacité à former un biofilm n'est pas seulement un facteur de virulence des bactéries pathogènes. En effet, la capacité de bactéries non pathogènes à former des biofilms est utilisée dans l'industrie pour la formation de « biofilms positifs ». Le terme de « biofilms positifs » désigne des biofilms protecteurs pouvant contrecarrer l'implantation de micro-organismes, dont ceux indésirables.

La première étape de formation d'un biofilm est l'adhésion des bactéries à une surface. Cette capacité a auparavant été étudiée notamment chez des bactéries lactiques telles que *L. lactis* et *S. thermophilus* (Mercier *et al.*, 2002). Ces études montrent que le pouvoir adhésif de *S. thermophilus* est faible comparativement à celui de *L. lactis*. Ce résultat suggère que *S. thermophilus* est déficient en facteur d'adhésion, ou alors qu'elle présente des facteurs antiadhésifs que *L. lactis* ne présente pas.

3.2 Influence du gène *cse* sur le pouvoir adhésif de *S. thermophilus*

3.2.1 L'augmentation de la longueur des chaînes résultant de la délétion de *cse* induirait une diminution du pouvoir adhésif de *S. thermophilus*

Il a été montré que la délétion du gène *cse* conduit à une réduction du pouvoir adhésif de *S. thermophilus* à une surface en verre. Or, le gène *cse* est impliqué dans la ségrégation cellulaire, son absence entraîne en effet la formation de longues chaînes de cellules. Mercier *et al.* (2002) ont montré que la longueur des chaînes est un facteur jouant un rôle important dans l'aptitude de bactéries telles que *L. lactis* et *S. thermophilus*, à adhérer à des surfaces non biotiques. Plus précisément, l'allongement des chaînes de cellules entraîne une diminution de la capacité des cellules à adhérer à une surface. Il est donc probable que le faible pouvoir adhésif du mutant Δcse , par rapport à la souche sauvage, résulte de sa croissance sous la forme de longues chaînes.

3.2.2 Variabilité de la longueur des chaînes et variabilité du pouvoir adhésif

L'analyse des chaînes de cellules de différentes souches de *S. thermophilus* montre que la longueur des chaînes peut être très variable d'une souche à l'autre. Certaines souches se développent sous la forme de chaînes courtes, comme la souche CNRZ445 ou CNRZ368, et d'autres sous la forme de chaînes longues, comme la souche LMG18311. Les chaînes d'un mutant Δcse de la souche LMG18311 étant plus longues que celles de la souche sauvage, il semble qu'il y ait une activité de ségrégation cellulaire conférée par le gène *cse* de cette souche, bien qu'elle se développe sous la forme de chaînes longues. L'analyse des chaînes de cellules de souches mérodiplôïdes pour le gène *cse* suggère fortement qu'une origine génétique de la variabilité de la longueur des chaînes est localisée dans le locus *cse*. Il est ainsi probable qu'un ou plusieurs facteurs de divergence génétique entre les allèles *cse* soit responsable de la variabilité intraspécifique de la longueur des chaînes. Ces facteurs de divergences sont probablement localisés en dehors de la région *var-cse* car il a été montré que le remplacement de la région *var-cse* de la souche LMG18311 par celle de la souche CNRZ368 a peu ou pas d'influence sur l'activité de ségrégation cellulaire de la protéine Cse.

La longueur des chaînes de cellules est un paramètre influençant la capacité des cellules à adhérer à des surfaces (Mercier *et al.*, 2002). De plus, il a été montré que *cse* joue un rôle dans le pouvoir adhésif des cellules de *S. thermophilus*. Etant donné que la longueur des chaînes de *S. thermophilus* est variable au niveau intraspécifique et que cette variabilité résulte d'un facteur de variabilité du locus *cse*, il est plausible que le pouvoir adhésif des cellules de *S. thermophilus* soit variable au niveau intraspécifique.

3.3 Influence des exopolysaccharides sur le pouvoir adhésif de *S. thermophilus*

La culture de la souche de *S. thermophilus* sauvage en milieu semi-liquide a permis de sélectionner des souches variantes produisant moins d'exopolysaccharides (ou EPS). Il a été montré que ces souches variantes ont un pouvoir d'adhésion aux surfaces en verre qui est environ dix fois supérieur à celui de la souche sauvage. Il est probable que cette augmentation du pouvoir adhésif résulte de la diminution de la production d'EPS, même si on ne peut pas totalement exclure qu'elle résulte d'une modification qualitative de la structure des EPS

produits. Chez *E. coli*, la présence de la capsule diminue la capacité des cellules à s'agréger. En effet, la présence de la capsule empêcherait l'agrégation cellulaire en masquant les adhésines, telle que la protéine Ag43, qui ne pourraient plus interagir ensemble d'une cellule à l'autre (Schembri *et al.*, 2004). Il a également été montré que les cellules de *S. pneumoniae* adhèrent plus efficacement aux cellules hôtes lorsque la production de capsule est diminuée (Talbot *et al.*, 1996). Un phénomène similaire pourrait expliquer l'augmentation du pouvoir adhésif du variant de *S. thermophilus* produisant moins d'EPS. Des facteurs adhésifs présents à la surface des cellules de *S. thermophilus* pourraient être dans l'impossibilité d'entrer en contact avec le verre lorsque les cellules produisent des EPS.

Perspectives

Les données accumulées jusque là sur la fonction, la structure et l'évolution du gène *cse*, conduisent à poser plusieurs questions. Notamment, est-ce que le réassortiment de modules à l'origine du gène *cse* a été générateur d'une nouvelle fonction ? Est-ce que la région Var-Cse joue un rôle dans la fonction de ségrégation cellulaire de la protéine Cse ? Est-elle non indispensable ?

De façon plus générale, Cse constitue un exemple supplémentaire de protéines de surface ayant évoluées par réassortiment de domaines. Par conséquent, les protéines de surface évolueraient-elles plus fréquemment par réassortiment de domaines que les protéines intracellulaires ?

L'étude de *cse* permet d'échafauder un modèle moléculaire pour sa création, faisant intervenir une recombinaison entre régions variables et de faible complexité. Ce modèle peut-il être appliqué à d'autres protéines elles aussi créées par réassortiment de domaines ?

La formation de longues chaînes de cellules ainsi que la production d'EPS, entravant l'adhésion cellulaire, sont-ils des facteurs responsables du faible pouvoir probiotique de *S. thermophilus* ?

Les stratégies proposées dans le but de répondre à ces questions sont présentées ici.

1 Conséquences fonctionnelles du réassortiment de domaines à l'origine de Cse

Il est probable que la fonction de ségrégation cellulaire de Cse résulte de l'association des domaines putatifs LysM et CHAP. Toutefois, on ne peut pas exclure que l'un des deux domaines ne soit pas impliqué dans cette fonction. C'est pourquoi il sera nécessaire d'analyser les conséquences fonctionnelles de la délétion des domaines LysM et CHAP.

L'hypothèse la plus simple à élaborer concernant la fonction de la protéine Cse est que celle-ci est associée à la paroi *via* son domaine LysM et que son domaine CHAP catalyse l'hydrolyse de liaisons du peptidoglycane. La fonctionnalité du domaine LysM pourrait être testée grâce à des expériences de détection à la surface des cellules d'une protéine issue de la fusion traductionnelle entre le domaine LysM et une protéine rapporteuse telle que la GFP, protéine fluorescente.

Les tests d'activité de lyse de la paroi qui ont été utilisés ici permettent classiquement de détecter les hydrolases de la paroi dont l'activité entraîne la lyse cellulaire. Or, Cse étant une protéine impliquée dans la séparation des cellules, il est possible que son activité n'entraîne pas la lyse cellulaire, même lorsque la protéine est en excès. Par conséquent, un test permettant de détecter directement l'activité d'hydrolyse de la paroi pourrait révéler que Cse présente bien une activité hydrolase du peptidoglycane. Pour cela, les tests d'activité pourront être réalisés avec comme substrat des extraits de paroi de *S. thermophilus* préalablement marqués radioactivement. Une approche similaire a été utilisée avec succès pour révéler l'activité lytique de la LytB de *S. pneumoniae*, protéine également impliquée dans la ségrégation cellulaire (De Las Rivas *et al.*, 2002).

Les protéines Cse et PcsB présentent toutes deux dans leur région C-terminale un domaine CHAP. Celui de la protéine Cse serait issu de la duplication d'un module du gène *pcsB* codant ce domaine CHAP. Les régions N-terminales de ces deux protéines ne présentent pas d'identité de séquence significative, suggérant que ces deux protéines pourraient avoir une fonction différente. Dans le but de tester cette hypothèse, il serait nécessaire d'analyser le phénotype d'un mutant *pcsB* de *S. thermophilus* et d'un double mutant *cse-pcsB*. Il serait ensuite nécessaire de déterminer si les domaines CHAP assurent la même fonction au sein de ces protéines. Pour cela, un gène chimérique présentant la partie 5' du gène *cse*, et la partie 3' du gène *pcsB* codant le domaine CHAP, pourra être construit. L'analyse de la fonctionnalité de ce gène pour être effectuée grâce à un test de complémentation d'un mutant Δcse avec ce gène chimérique. La restauration du phénotype sauvage signifiera que le domaine CHAP de la protéine PcsB assure la même fonction que celui de la protéine Cse. Il serait également intéressant de connaître en quoi les domaines N-terminaux des protéines Cse et PcsB spécialiseraient la fonction des domaines CHAP. Une hypothèse qui pourra être testée est que ces domaines N-terminaux confèrent une localisation géographique différente à la surface des cellules de *S. thermophilus*.

Il serait également important de connaître la fonction de la protéine SIP pour déterminer si la création du gène *cse* a effectivement engendré une nouvelle fonction. La fonction de cette protéine pourrait être étudiée par l'analyse des conséquences phénotypiques de la délétion du gène *sip* chez un autre streptocoque.

L'ensemble de ces expériences devrait permettre de savoir si la création du gène *cse* par réassortiment de modules a permis de générer une nouvelle fonction.

2 Fonction de la région Var-Cse dans la ségrégation cellulaire

L'importance de la région Var-Cse dans la fonction de ségrégation cellulaire de la protéine Cse pourrait être évaluée par l'analyse de la longueur des chaînes d'un mutant, dans lequel la totalité de la région *var-cse* aurait été déléetée. Dans le cas où cette délétion entraînerait l'abolition de l'activité de ségrégation cellulaire, d'autres mutants correspondant à la délétion d'une partie plus ou moins importante de cette région pourraient être construits. L'analyse de la longueur des chaînes de ces mutants permettrait de déterminer quelle doit être la longueur minimale de la région Var-Cse pour que la protéine Cse présente sa pleine activité de ségrégation cellulaire.

3 Importance des régions variables et de faible complexité dans la génération de nouvelles fonctions

Nous proposons l'hypothèse que les régions variables et de faible complexité favorisent les réassortiments de domaines, générateurs de nouvelles fonctions. Cette hypothèse pourrait être testée tout d'abord par l'établissement d'un répertoire de gènes chimériques à partir des séquences des génomes de micro-organismes entièrement séquencés. Ce répertoire pourrait

contenir les informations issues des prédictions de domaines protéiques. L'ensemble pourrait notamment permettre de connaître le pourcentage de sites de fusion localisés à l'intérieur de module et ceux situés à l'extérieur de modules. Par ailleurs, l'analyse de ce répertoire pourrait permettre de déterminer si les protéines de surface évoluent plus fréquemment que les protéines intracellulaires par réassortiments de domaines, comme l'ont suggéré Lopez *et al* (2000).

En parallèle, les séquences intragéniques variables et de faible complexité pourront être recherchées dans les génomes de micro-organismes entièrement séquencés. Cette analyse permettra de déterminer le pourcentage de gènes contenant ce type de région. Le pourcentage de gènes chimériques pour lesquels le site de fusion est localisé à l'intérieur d'une région variable et de faible complexité pourra également être déterminé. Il sera ainsi possible de savoir si les gènes chimériques présentent favorablement des séquences variables et de faible complexité. Si tel est le cas, cela pourrait signifier que ces régions variables et de faible complexité favorisent les réarrangements entraînant des réassortiments de domaines.

Une approche moléculaire pourra également être envisagée. Un système expérimental pourra être élaboré dans le but de déterminer la fréquence relative de réassortiment de domaines en fonction de la longueur de la région variable et de faible complexité. Par exemple, deux régions variables et de faible complexité pourront être clonées séparément : l'une dans un plasmide, l'autre dans le chromosome. Chacune de ces régions pourra être fusionnée à un module de telle sorte que si *in vivo*, un événement de recombinaison entre les deux régions variables et de faible complexité se produit, il puisse entraîner la formation d'un gène conférant un phénotype sélectionnable positivement, comme par exemple un gène de résistance à un antibiotique.

4 Adhésion de *S. thermophilus* à la surface de la paroi intestinale

Bien que *S. thermophilus* soit utilisé comme probiotique, il semble que cette bactérie ne soit pas un probiotique très efficace. Une hypothèse pouvant expliquer son faible pouvoir probiotique serait que les souches de *S. thermophilus* testées adhèrent peu efficacement aux parois intestinales. En effet, cette capacité adhésive peut être requise pour une action probiotique (pour revue Saarela *et al.*, 2000). Des résultats préliminaires obtenus au cours de cette étude suggèrent que la longueur des chaînes de cellules et la production d'exopolysaccharides seraient des facteurs influençant négativement le pouvoir d'adhésion de *S. thermophilus* à une surface en verre. Il serait ainsi intéressant de tester si la longueur des chaînes et la production d'EPS influencent l'aptitude de *S. thermophilus* à adhérer à la surface de la paroi intestinale.

Références

- Apic, G., J. Gough and S. A. Teichmann** (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* **310**(2): 311-25.
- Areschoug, T., S. Linse, M. Stalhammar-Carlemalm, L. O. Heden and G. Lindahl** (2002). A proline-rich region with a highly periodic sequence in Streptococcal beta protein adopts the polyproline II structure and is exposed on the bacterial surface. *J Bacteriol* **184**(22): 6376-83.
- Bateman, A. and N. D. Rawlings** (2003). The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem Sci* **28**(5): 234-7.
- Beveridge, T. J.** (2000). Ultrastructure of Gram-Positive Cell Walls. Gam-Positive Pathogens. F. V. A., N. R. P., F. J.J., P. D.A. et R. J.I. Washington D.C., ASM press: 3-10.
- Bi, X. and L. F. Liu** (1994). *recA*-independent and *recA*-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. *J Mol Biol* **235**(2): 414-23.
- Bianchi, M. E. and C. M. Radding** (1983). Insertions, deletions and mismatches in heteroduplex DNA made by *recA* protein. *Cell* **35**(2 Pt 1): 511-20.
- Bierne, H. and B. Michel** (1994). When replication forks stop. *Mol Microbiol* **13**(1): 17-23.
- Bollinger, J. M., Jr., D. S. Kwon, G. W. Huisman, R. Kolter and C. T. Walsh** (1995). Glutathionylspermidine metabolism in *Escherichia coli*. Purification, cloning, overproduction, and characterization of a bifunctional glutathionylspermidine synthetase/amidase. *J Biol Chem* **270**(23): 14031-41.
- Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau and P. Hols** (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**(12): 1554-8.
- Brady, L. J., D. G. Cvitkovitch, C. M. Geric, M. N. Addison, J. C. Joyce, P. J. Crowley and A. S. Bleiweis** (1998). Deletion of the central proline-rich repeat domain results in altered antigenicity and lack of surface expression of the *Streptococcus mutans* P1 adhesin molecule. *Infect Immun* **66**(9): 4274-82.
- Broadbent, J. R., D. J. McMahon, D. L. Welker, C. J. Oberg and S. Moineau** (2003). Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J Dairy Sci* **86**(2): 407-23.
- Brunings, A. M. and D. W. Gabriel** (2003). *Xanthomonas citri* : breaking the surface. *Mol Plant Pathol* **4**(3): 141-57.
- Buist, G., J. Kok, K. J. Leenhouts, M. Dabrowska, G. Venema and A. J. Haandrikman** (1995). Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation. *J Bacteriol* **177**(6): 1554-63.
- Burrus, V., Y. Roussel, B. Decaris and G. Guedon** (2000). Characterization of a novel integrative element, ICES*StI*, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* **66**(4): 1749-53.
- Cabanes, D., P. Dehoux, O. Dussurget, L. Frangeul and P. Cossart** (2002). Surface proteins and the pathogenic potential of *Listeria monocytogenes*. *Trends Microbiol* **10**(5): 238-45.
- Carroll, S. A., T. Hain, U. Technow, A. Darji, P. Pashalidis, S. W. Joseph and T. Chakraborty** (2003). Identification and characterization of a peptidoglycan hydrolase, MurA, of *Listeria monocytogenes*, a muramidase needed for cell separation. *J Bacteriol* **185**(23): 6801-8.
- Charlesworth, B., P. Sniegowski and W. Stephan** (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**(6494): 215-20.

- Chedin, F., E. Dervyn, R. Dervyn, S. D. Ehrlich and P. Noirot** (1994). Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol Microbiol* **12**(4): 561-9.
- Chia, J. S., L. Y. Chang, C. T. Shun, Y. Y. Chang, Y. G. Tsay and J. Y. Chen** (2001a). A 60-kilodalton immunodominant glycoprotein is essential for cell wall integrity and the maintenance of cell shape in *Streptococcus mutans*. *Infect Immun* **69**(11): 6987-98.
- Chia, J. S., Y. Y. Lee, P. T. Huang and J. Y. Chen** (2001b). Identification of stress-responsive genes in *Streptococcus mutans* by differential display reverse transcription-PCR. *Infect Immun* **69**(4): 2493-501.
- Clark, A. J.** (1971). Toward a metabolic interpretation of genetic recombination of *E. coli* and its phages. *Annu Rev Microbiol* **25**: 437-64.
- Costerton, J. W., Z. Lewandowski, D. E. Caldwell, D. R. Korber and H. M. Lappin-Scott** (1995). Microbial biofilms. *Annu Rev Microbiol* **49**: 711-45.
- Croux, C., C. Ronda, R. Lopez and J. L. Garcia** (1993). Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Mol Microbiol* **9**(5): 1019-25.
- de Benito, I., M. E. Cano, J. Agüero and J. M. Garcia Lobo** (2004). A polymorphic tandem repeat potentially useful for typing in the chromosome of *Yersinia enterocolitica*. *Microbiology* **150**(Pt 1): 199-204.
- De Las Rivas, B., J. L. Garcia, R. Lopez and P. Garcia** (2002). Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* **184**(18): 4988-5000.
- Delcour, J., T. Ferain, M. Deghorain, E. Palumbo and P. Hols** (1999). The biosynthesis and functionality of the cell-wall of lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**(1-4): 159-84.
- Deretic, V. and W. M. Konyecsni** (1990). A prokaryotic regulatory factor with a histone H1-like carboxy-terminal domain: clonal variation of repeats within *algP*, a gene involved in regulation of mucoidy in *Pseudomonas aeruginosa*. *J Bacteriol* **172**(10): 5544-54.
- Diaz, E., R. Lopez and J. L. Garcia** (1990). Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. *Proc Natl Acad Sci U S A* **87**(20): 8125-9.
- Diaz, E., R. Lopez and J. L. Garcia** (1991). Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J Biol Chem* **266**(9): 5464-71.
- Doolittle, R. F.** (1995). The multiplicity of domains in proteins. *Annu Rev Biochem* **64**: 287-314.
- Dramsi, S., P. Dehoux and P. Cossart** (1993). Common features of gram-positive bacterial proteins involved in cell recognition. *Mol Microbiol* **9**(5): 1119-21.
- Dubois, M., K. Gilles, J. K. Hamilton, P. A. Rebers and F. Smith** (1951). A colorimetric method for the determination of sugars. *Nature* **168**(4265): 167.
- Dybvig, K., R. Sitaraman and C. T. French** (1998). A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc Natl Acad Sci U S A* **95**(23): 13923-8.
- Eriksen, T. A., A. Kadziola and S. Larsen** (2002). Binding of cations in *Bacillus subtilis* phosphoribosyl diphosphate synthetase and their role in catalysis. *Protein Sci* **11**(2): 271-9.
- Fahnestock, S. R., P. Alexander, J. Nagle and D. Filpula** (1986). Gene for an immunoglobulin-binding protein from a group G streptococcus. *J Bacteriol* **167**(3): 870-80.
- Filipe, S. R., E. Severina and A. Tomasz** (2000). Distribution of the mosaic structured *murM* genes among natural populations of *Streptococcus pneumoniae*. *J Bacteriol* **182**(23): 6798-805.

- Filpula, D., P. Alexander and S. R. Fahnestock** (1987). Nucleotide sequence of the protein G gene from *Streptococcus* GX7805, and comparison to previously reported sequences. *Nucleic Acids Res* **15**(17): 7210.
- Fischetti, V. A., R. P. Novick, J. J. Ferreti, P. D.A. and J. L. Rood** (2000). Gram-Positive Pathogens, ASM press.
- Fischetti, V. A., V. Pancholi and O. Schneewind** (1991). Common characteristics of the surface proteins from gram-positive cocci. Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci. G. M. Dunny, P. P. Cleary et L. L. Mc Kay, ASM press: 290-4.
- Flor, H. H.** (1971). Current status of the gene-for-gene concept. *Annu Rev Phytopathol* **9**: 275-96.
- Foucaud-Scheunemann, C. and I. Poquet** (2003). HtrA is a key factor in the response to specific stress conditions in *Lactococcus lactis*. *FEMS Microbiol Lett* **224**(1): 53-9.
- Fridovich, I.** (1983). Superoxide radical: an endogenous toxicant. *Annu Rev Pharmacol Toxicol* **23**: 239-57.
- Fuss, C., J. C. Palmaz and E. A. Sprague** (2001). Fibrinogen: structure, function, and surface interactions. *J Vasc Interv Radiol* **12**(6): 677-82.
- Garcia, P., J. L. Garcia, E. Garcia, J. M. Sanchez-Puelles and R. Lopez** (1990). Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* **86**(1): 81-8.
- Garcia, P., M. P. Gonzalez, E. Garcia, R. Lopez and J. L. Garcia** (1999). LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol* **31**(4): 1275-7.
- Giard, J. C., A. Rince, H. Capioux, Y. Auffray and A. Hartke** (2000). Inactivation of the stress- and starvation-inducible *gls24* operon has a pleiotrophic effect on cell morphology, stress sensitivity, and gene expression in *Enterococcus faecalis*. *J Bacteriol* **182**(16): 4512-20.
- Gilbert, W.** (1978). Why genes in pieces? *Nature* **271**: 501.
- Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, Jr. and R. A. Warren** (1991). Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev* **55**(2): 303-15.
- Giraud, E. and G. Cuny** (1997). Molecular characterization of the alpha-amylase genes of *Lactobacillus plantarum* A6 and *Lactobacillus amylovorus* reveals an unusual 3' end structure with direct tandem repeats and suggests a common evolutionary origin. *Gene* **198**(1-2): 149-57.
- Glickman, B. W. and L. S. Ripley** (1984). Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc Natl Acad Sci U S A* **81**(2): 512-6.
- Goffin, C. and J. M. Ghuyssen** (1998). Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. *Microbiol Mol Biol Rev* **62**(4): 1079-93.
- Hakenbeck, R.** (2000). Transformation in *Streptococcus pneumoniae*: mosaic genes and the regulation of competence. *Res Microbiol* **151**(6): 453-6.
- Hallet, B.** (2001). Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiol* **4**(5): 570-81.
- Hartford, O., P. Francois, P. Vaudaux and T. J. Foster** (1997). The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. *Mol Microbiol* **25**(6): 1065-76.
- Hay, E. D.** (1991). Cell biology of extracellular matrix, Plenum Press, New York, NY.
- Heddle, C., A. H. Nobbs, N. S. Jakubovics, M. Gal, J. P. Mansell, D. Dymock and H. F. Jenkinson** (2003). Host collagen signal induces antigen I/II adhesin and invasin gene expression in oral *Streptococcus gordonii*. *Mol Microbiol* **50**(2): 597-607.

- Heidrich, C., M. F. Templin, A. Ursinus, M. Merdanovic, J. Berger, H. Schwarz, M. A. de Pedro and J. V. Holtje** (2001). Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**(1): 167-78.
- Heilmann, C., G. Thumm, G. S. Chhatwal, J. Hartleib, A. Uekotter and G. Peters** (2003). Identification and characterization of a novel autolysin (Aae) with adhesive properties from *Staphylococcus epidermidis*. *Microbiology* **149**(Pt 10): 2769-78.
- Henderson, I. R., P. Owen and J. P. Nataro** (1999). Molecular switches--the ON and OFF of bacterial phase variation. *Mol Microbiol* **33**(5): 919-32.
- Holtje, J. V. and A. Tomasz** (1975). Specific recognition of choline residues in the cell wall teichoic acid by the N-acetylmuramyl-L-alanine amidase of *Pneumococcus*. *J Biol Chem* **250**(15): 6072-6.
- Holtje, J. V. and A. Tomasz** (1976). Purification of the pneumococcal N-acetylmuramyl-L-alanine amidase to biochemical homogeneity. *J Biol Chem* **251**(14): 4199-207.
- Houghton, J. E., G. A. O'Donovan and J. R. Wild** (1989). Reconstruction of an enzyme by domain substitution effectively switches substrate specificity. *Nature* **338**(6211): 172-4.
- Huard, C., G. Miranda, F. Wessner, A. Bolotin, J. Hansen, S. J. Foster and M. P. Chapot-Chartier** (2003). Characterization of AcMB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*. *Microbiology* **149**(Pt 3): 695-705.
- Husson-Kao, C., J. Mengaud, L. Benbadis and M. P. Chapot-Chartier** (2000). Mur1, a *Streptococcus thermophilus* peptidoglycan hydrolase devoid of a specific cell wall binding domain. *FEMS Microbiol Lett* **187**(1): 69-76.
- Ishikawa, S., Y. Hara, R. Ohnishi and J. Sekiguchi** (1998). Regulation of a new cell wall hydrolase gene, *cwlF*, which affects cell separation in *Bacillus subtilis*. *J Bacteriol* **180**(9): 2549-55.
- Ishino, F., W. Park, S. Tomioka, S. Tamaki, I. Takase, K. Kunugita, H. Matsuzawa, S. Asoh, T. Ohta, B. G. Spratt and et al.** (1986). Peptidoglycan synthetic activities in membranes of *Escherichia coli* caused by overproduction of penicillin-binding protein 2 and RodA protein. *J Biol Chem* **261**(15): 7024-31.
- Juarez, K., H. Flores, S. Davila, L. Olvera, V. Gonzalez and E. Morett** (2000). Reciprocal domain evolution within a transactivator in a restricted sequence space. *Proc Natl Acad Sci U S A* **97**(7): 3314-8.
- Kadurugamuwa, J. L., A. Mayer, P. Messner, M. Sara, U. B. Sleytr and T. J. Beveridge** (1998). S-layered *Aneurinibacillus* and *Bacillus* spp. are susceptible to the lytic action of *Pseudomonas aeruginosa* membrane vesicles. *J Bacteriol* **180**(9): 2306-11.
- Kapur, V., S. Kanjilal, M. R. Hamrick, L. L. Li, T. S. Whittam, S. A. Sawyer and J. M. Musser** (1995). Molecular population genetic analysis of the streptokinase gene of *Streptococcus pyogenes*: mosaic alleles generated by recombination. *Mol Microbiol* **16**(3): 509-19.
- Karatzas, K. A. and M. H. Bennik** (2002). Characterization of a *Listeria monocytogenes* Scott A isolate with high tolerance towards high hydrostatic pressure. *Appl Environ Microbiol* **68**(7): 3183-9.
- Karatzas, K. A., J. A. Wouters, C. G. Gahan, C. Hill, T. Abee and M. H. Bennik** (2003). The CtsR regulator of *Listeria monocytogenes* contains a variant glycine repeat region that affects piezotolerance, stress resistance, motility and virulence. *Mol Microbiol* **49**(5): 1227-38.
- Kehoe, M. A.** (1994). Cell-wall-associated proteins in Gram-positive bacteria. *Bacterial Cell Wall*. J.-M. G. a. R. Hakenbeck, Elsevier Science B.V: 217-61.
- Kelly, C., P. Evans, L. Bergmeier, S. F. Lee, A. Progulsk-Fox, A. C. Harris, A. Aitken, A. S. Bleiweis and T. Lehner** (1989). Sequence analysis of the cloned streptococcal cell surface antigen I/II. *FEBS Lett* **258**(1): 127-32.
- Koroleva, I. V., A. Efstratiou and A. N. Suvorov** (2002). Structural heterogeneity of the streptococcal C5a peptidase gene in *Streptococcus pyogenes*. *J Bacteriol* **184**(22): 6384-6.

- Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder and W. M. Rehrauer** (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* **58**(3): 401-65.
- Lachica, R. V., C. Genigeorgis and P. D. Hoepflich** (1971). Metachromatic agar-diffusion methods for detecting staphylococcal nuclease activity. *Appl Microbiol* **21**(4): 585-7.
- Le Fleche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoel, V. Ramière, P. Sylvestre, G. Benson, F. Ramière and G. Vergnaud** (2001). A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* **1**(1): 2.
- Lee, S. F., A. Progulsk-Fox and A. S. Bleiweis** (1988). Molecular cloning and expression of a *Streptococcus mutans* major surface protein antigen, P1 (I/II), in *Escherichia coli*. *Infect Immun* **56**(8): 2114-9.
- Levinson, G. and G. A. Gutman** (1987). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* **15**(13): 5323-38.
- Liu, M. and A. Grigoriev** (2004). Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling? *Trends Genet* **20**(9): 399-403.
- Lopez, R., M. P. Gonzalez, E. Garcia, J. L. Garcia and P. Garcia** (2000). Biological roles of two new murein hydrolases of *Streptococcus pneumoniae* representing examples of module shuffling. *Res Microbiol* **151**(6): 437-43.
- Lukowski, S., K. Nakashima, I. Abdi, V. J. Cipriano, R. M. Ireland, S. D. Reid, G. G. Adams and J. M. Musser** (2000). Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. *Infect Immun* **68**(12): 6542-53.
- Madera, M., C. Vogel, S. K. Kummerfeld, C. Chothia and J. Gough** (2004). The SUPERFAMILY database in 2004: additions and improvements. *Nucleic Acids Res* **32 Database issue**: D235-9.
- Madoff, L. C., S. Hori, J. L. Michel, C. J. Baker and D. L. Kasper** (1991). Phenotypic diversity in the alpha C protein of group B streptococci. *Infect Immun* **59**(8): 2638-44.
- Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling and D. L. Kasper** (1996). Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc Natl Acad Sci U S A* **93**(9): 4131-6.
- Maguin, E., H. Prevost, S. D. Ehrlich and A. Gruss** (1996). Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J Bacteriol* **178**(3): 931-5.
- Majewski, J. and F. M. Cohan** (1999). DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* **153**(4): 1525-33.
- Margot, P., M. Pagni and D. Karamata** (1999). *Bacillus subtilis* 168 gene *lytF* encodes a gamma-D-glutamate-meso-diaminopimelate mureopeptidase expressed by the alternative vegetative sigma factor, *sigmaD*. *Microbiology* **145** (Pt 1): 57-65.
- Mazin, A. V., A. V. Kuzminov, G. L. Dianov and R. I. Salganik** (1991). Mechanisms of deletion formation in *Escherichia coli* plasmids. II. Deletions mediated by short direct repeats. *Mol Gen Genet* **228**(1-2): 209-14.
- McDevitt, D., P. Francois, P. Vaudaux and T. J. Foster** (1994). Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*. *Mol Microbiol* **11**(2): 237-48.
- Medvedkin, V. N., E. A. Permyakov, L. V. Klimenko, Y. V. Mitin, N. Matsushima, S. Nakayama and R. H. Kretsinger** (1995). Interactions of (Ala*Ala*Lys*Pro)_n and (Lys*Lys*Ser*Pro)_n with DNA. Proposed coiled-coil structure of AlgR3 and AlgP from *Pseudomonas aeruginosa*. *Protein Eng* **8**(1): 63-70.
- Mercier, C., E. Domakova, J. Tremblay and S. Kulakauskas** (2000). Effects of a muramidase on a mixed bacterial community. *FEMS Microbiol Lett* **187**(1): 47-52.
- Mercier, C., C. Durrieu, R. Briandet, E. Domakova, J. Tremblay, G. Buist and S. Kulakauskas** (2002). Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol Microbiol* **46**(1): 235-43.

- Mesnager, S., T. Fontaine, T. Mignot, M. Delepierre, M. Mock and A. Fouet** (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *Embo J* **19**(17): 4473-84.
- Mesnager, S. and A. Fouet** (2002). Plasmid-encoded autolysin in *Bacillus anthracis*: modular structure and catalytic properties. *J Bacteriol* **184**(1): 331-4.
- Mesnager, S., E. Tosi-Couture, M. Mock, P. Gounon and A. Fouet** (1997). Molecular characterization of the *Bacillus anthracis* main S-layer component: evidence that it is the major cell-associated antigen. *Mol Microbiol* **23**(6): 1147-55.
- Michel, B.** (1999). Illegitimate recombination in bacteria. Organization of the prokaryotic genome. R. L. Charlebois, American Society for Microbiology: 129-50.
- Michel, J. L., L. C. Madoff, K. Olson, D. E. Kling, D. L. Kasper and F. M. Ausubel** (1992). Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci. *Proc Natl Acad Sci U S A* **89**(21): 10060-4.
- Mouw, A. R., E. H. Beachey and V. Burdett** (1988). Molecular evolution of streptococcal M protein: cloning and nucleotide sequence of the type 24 M protein gene and relation to other genes of *Streptococcus pyogenes*. *J Bacteriol* **170**(2): 676-84.
- Mrazek, J., L. H. Gaynon and S. Karlin** (2002). Frequent oligonucleotide motifs in genomes of three streptococci. *Nucleic Acids Res* **30**(19): 4216-21.
- Murray, N. E.** (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**(2): 412-34.
- Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia** (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**(4): 536-40.
- Nagaraja, V., J. C. Shepherd and T. A. Bickle** (1985). A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**(6026): 371-2.
- Navarre, W. W. and O. Schneewind** (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* **63**(1): 174-229.
- Ng, W. L., K. M. Kazmierczak and M. E. Winkler** (2004). Defective cell wall synthesis in *Streptococcus pneumoniae* R6 depleted for the essential PcsB putative murein hydrolase or the VicR (YycF) response regulator. *Mol Microbiol* **53**(4): 1161-75.
- Ng, W. L., G. T. Robertson, K. M. Kazmierczak, J. Zhao, R. Gilmour and M. E. Winkler** (2003). Constitutive expression of PcsB suppresses the requirement for the essential VicR (YycF) response regulator in *Streptococcus pneumoniae* R6. *Mol Microbiol* **50**(5): 1647-63.
- Ni Eidhin, D., S. Perkins, P. Francois, P. Vaudaux, M. Hook and T. J. Foster** (1998). Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. *Mol Microbiol* **30**(2): 245-57.
- Ohnishi, R., S. Ishikawa and J. Sekiguchi** (1999). Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J Bacteriol* **181**(10): 3178-84.
- Okahashi, N., C. Sasakawa, M. Yoshikawa, S. Hamada and T. Koga** (1989). Molecular characterization of a surface protein antigen gene from serotype c *Streptococcus mutans*, implicated in dental caries. *Mol Microbiol* **3**(5): 673-8.
- O'Sullivan, D., D. P. Twomey, A. Coffey, C. Hill, G. F. Fitzgerald and R. P. Ross** (2000). Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. *Mol Microbiol* **36**(4): 866-75.
- Patthy, L.** (1991). Modular exchange principles in proteins. *Curr Opin Struct Biol* **1**: 351-61.

- Patthy, L.** (1994). Introns and exons. *Curr Opin Struct Biol* **4**: 383-92.
- Pavlovic, G.** (2004). Evolution d'une famille d'éléments intégratifs potentiellement conjugatifs et/ou mobilisables de *Streptococcus thermophilus*. Laboratoire de Génétique et Microbiologie, UMR-INRA 1128. Université Henri Poincaré, Nancy 1. Vandoeuvre-lès-Nancy. Pages.
- Peeters, B. P., J. H. de Boer, S. Bron and G. Venema** (1988). Structural plasmid instability in *Bacillus subtilis*: effect of direct and inverted repeats. *Mol Gen Genet* **212**(3): 450-8.
- Petes, T. D. and C. W. Hill** (1988). Recombination between repeated genes in microorganisms. *Annu Rev Genet* **22**: 147-68.
- Pluvinet, A., F. Charron-Bourgoin, C. Morel and B. Decaris** (2004). Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. *International Dairy Journal* **14**: 627-34.
- Popham, D. L. and K. D. Young** (2003). Role of penicillin-binding proteins in bacterial cell morphogenesis. *Curr Opin Microbiol* **6**(6): 594-9.
- Poquet, I., S. D. Ehrlich and A. Gruss** (1998). An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *J Bacteriol* **180**(7): 1904-12.
- Poquet, I., V. Saint, E. Seznec, N. Simoes, A. Bolotin and A. Gruss** (2000). HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Mol Microbiol* **35**(5): 1042-51.
- Price, C., J. Lingner, T. A. Bickle, K. Firman and S. W. Glover** (1989). Basis for changes in DNA recognition by the *EcoR124* and *EcoR124/3* type I DNA restriction and modification enzymes. *J Mol Biol* **205**(1): 115-25.
- Qin, X., K. V. Singh, G. M. Weinstock and B. E. Murray** (2000). Effects of *Enterococcus faecalis* *fsr* genes on production of gelatinase and a serine protease and virulence. *Infect Immun* **68**(5): 2579-86.
- Qin, X., K. V. Singh, G. M. Weinstock and B. E. Murray** (2001). Characterization of *fsr*, a regulator controlling expression of gelatinase and serine protease in *Enterococcus faecalis* OG1RF. *J Bacteriol* **183**(11): 3372-82.
- Qin, X., K. V. Singh, Y. Xu, G. M. Weinstock and B. E. Murray** (1998). Effect of disruption of a gene encoding an autolysin of *Enterococcus faecalis* OG1RF. *Antimicrob Agents Chemother* **42**(11): 2883-8.
- Rajagopal, L., A. Clancy and C. E. Rubens** (2003). A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J Biol Chem* **278**(16): 14429-41.
- Rashid, M. H., M. Mori and J. Sekiguchi** (1995). Glucosaminidase of *Bacillus subtilis*: cloning, regulation, primary structure and biochemical characterization. *Microbiology* **141** (Pt 10): 2391-404.
- Rasmussen, M. and L. Bjorck** (2001). Unique regulation of *SclB* - a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol Microbiol* **40**(6): 1427-38.
- Rathsam, C. and N. A. Jacques** (1998). Role of C-terminal domains in surface attachment of the fructosyltransferase of *Streptococcus salivarius* ATCC 25975. *J Bacteriol* **180**(23): 6400-3.
- Reinscheid, D. J., K. Ehlert, G. S. Chhatwal and B. J. Eikmanns** (2003). Functional analysis of a PcsB-deficient mutant of group B streptococcus. *FEMS Microbiol Lett* **221**(1): 73-9.
- Reinscheid, D. J., B. Gottschalk, A. Schubert, B. J. Eikmanns and G. S. Chhatwal** (2001). Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J Bacteriol* **183**(4): 1175-83.

- Ries, W., C. Hotzy, I. Schocher, U. B. Sleytr and M. Sara** (1997). Evidence that the N-terminal part of the S-layer protein from *Bacillus stearothermophilus* PV72/p2 recognizes a secondary cell wall polymer. *J Bacteriol* **179**(12): 3892-8.
- Rigden, D. J., M. J. Jedrzejewski and M. Y. Galperin** (2003). Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem Sci* **28**(5): 230-4.
- Rioux, S., D. Martin, H. W. Ackermann, J. Dumont, J. Hamel and B. R. Brodeur** (2001). Localization of surface immunogenic protein on group B streptococcus. *Infect Immun* **69**(8): 5162-5.
- Robson, R. L., R. M. Robson and J. G. Morris** (1974). The biosynthesis of granulose by *Clostridium pasteurianum*. *Biochem J* **144**(3): 503-11.
- Romero, D., J. Martinez-Salazar, E. Ortiz, C. Rodriguez and E. Valencia-Morales** (1999). Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. *Res Microbiol* **150**(9-10): 735-43.
- Saarela, M., G. Mogensen, R. Fonden, J. Matto and T. Mattila-Sandholm** (2000). Probiotic bacteria: safety, functional and technological properties. *J Biotechnol* **84**(3): 197-215.
- Sanchez-Puelles, J. M., C. Ronda, J. L. Garcia, P. Garcia, R. Lopez and E. Garcia** (1986). Searching for autolysin functions. Characterization of a pneumococcal mutant deleted in the *lytA* gene. *Eur J Biochem* **158**(2): 289-93.
- Sanchez-Puelles, J. M., J. M. Sanz, J. L. Garcia and E. Garcia** (1990). Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. *Gene* **89**(1): 69-75.
- Sara, M., E. M. Egelseer, C. Dekitsch and U. B. Sleytr** (1998). Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. *J Bacteriol* **180**(24): 6780-3.
- Schembri, M. A., D. Dalsgaard and P. Klemm** (2004). Capsule shields the function of short bacterial adhesins. *J Bacteriol* **186**(5): 1249-57.
- Schleifer, K. H.** (1986). Gram-positive cocci. In P. H. A. Sneath, N. S. Mair, M. E. Sharpe and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 2. William & Wilkins, Baltimore, Md.: 999-1002.
- Schubert, A., K. Zakikhany, M. Schreiner, R. Frank, B. Spellerberg, B. J. Eikmanns and D. J. Reinscheid** (2002). A fibrinogen receptor from group B *Streptococcus* interacts with fibrinogen by repetitive units with novel ligand binding sites. *Mol Microbiol* **46**(2): 557-69.
- Schupp, J. M., A. M. Klevytska, G. Zinser, L. B. Price and P. Keim** (2000). *vrrB*, a hypervariable open reading frame in *Bacillus anthracis*. *J Bacteriol* **182**(14): 3989-97.
- Schwarz-Linek, U., M. Hook and J. R. Potts** (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Mol Microbiol* **52**(3): 631-41.
- Schwecke, T., J. F. Aparicio, I. Molnar, A. Konig, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortes, J. B. Lester and et al.** (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci U S A* **92**(17): 7839-43.
- Smith, H. E., F. H. Reek, U. Vecht, A. L. Gielkens and M. A. Smits** (1993). Repeats in an extracellular protein of weakly pathogenic strains of *Streptococcus suis* type 2 are absent in pathogenic strains. *Infect Immun* **61**(8): 3318-26.
- Snyder, L. A., S. A. Butcher and N. J. Saunders** (2001). Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp. *Microbiology* **147**(Pt 8): 2321-32.
- Staunton, J. and B. Wilkinson** (1997). Biosynthesis of Erythromycin and Rapamycin. *Chem Rev* **97**(7): 2611-2630.

- Steen, A., G. Buist, K. J. Leenhouts, M. El Khattabi, F. Grijpstra, A. L. Zomer, G. Venema, O. P. Kuipers and J. Kok** (2003). Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J Biol Chem* **278**(26): 23874-81.
- Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu and I. T. Paulsen** (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**(6799): 959-64.
- Talbot, U. M., A. W. Paton and J. C. Paton** (1996). Uptake of *Streptococcus pneumoniae* by respiratory epithelial cells. *Infect Immun* **64**(9): 3772-7.
- Tanaka, T., Y. Kuroda and S. Yokoyama** (2003). Characteristics and prediction of domain linker sequences in multi-domain proteins. *J Struct Funct Genomics* **4**(2-3): 79-85.
- Teichmann, S. A., J. Park and C. Chothia** (1998). Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements. *Proc Natl Acad Sci U S A* **95**(25): 14658-63.
- Teng, F., M. Kawalec, G. M. Weinstock, W. Hryniewicz and B. E. Murray** (2003). An *Enterococcus faecium* secreted antigen, SagA, exhibits broad-spectrum binding to extracellular matrix proteins and appears essential for *E. faecium* growth. *Infect Immun* **71**(9): 5033-41.
- Thibessard, A.** (2002). Défense de *Streptococcus thermophilus* contre le stress oxydatif :
-existence d'un système de réponse,
-construction et sélection d'une collection de mutants,
-identification de gènes impliqués. Laboratoire de Génétique et Microbiologie, UMR-INRA 1128. Université Henri Poincaré, Nancy 1. Vandoeuvre-lès-Nancy. Pages.
- Thibessard, A., F. Borges, A. Fernandez, B. Gintz, B. Decaris and N. Leblond-Bourget** (2004). Identification of *Streptococcus thermophilus* CNRZ368 Genes Involved in Defense against Superoxide Stress. *Appl Environ Microbiol* **70**(4): 2220-9.
- Thibessard, A., A. Fernandez, B. Gintz, B. Decaris and N. Leblond-Bourget** (2002a). Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can J Microbiol* **48**(5): 473-8.
- Thibessard, A., A. Fernandez, B. Gintz, N. Leblond-Bourget and B. Decaris** (2002b). Effects of *rodA* and *pbp2b* disruption on cell morphology and oxidative stress response of *Streptococcus thermophilus* CNRZ368. *J Bacteriol* **184**(10): 2821-6.
- Thibodeau, E. A. and C. M. Ford** (1991). Chain formation and de-chaining in *Streptococcus sobrinus* SL-1. *Oral Microbiol Immunol* **6**(5): 313-5.
- Turner, M. S., L. M. Hafner, T. Walsh and P. M. Giffard** (2004). Identification and characterization of the novel LysM domain-containing surface protein Sep from *Lactobacillus fermentum* BR11 and its use as a peptide fusion partner in *Lactobacillus* and *Lactococcus*. *Appl Environ Microbiol* **70**(6): 3673-80.
- van Belkum, A.** (1999). Short sequence repeats in microbial pathogenesis and evolution. *Cell Mol Life Sci* **56**(9-10): 729-34.
- van Belkum, A., S. Scherer, L. van Alphen and H. Verbrugh** (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**(2): 275-93.
- Vogel, C., M. Bashton, N. D. Kerrison, C. Chothia and S. A. Teichmann** (2004). Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* **14**(2): 208-16.
- Waters, C. M., M. H. Antiporta, B. E. Murray and G. M. Dunny** (2003). Role of the *Enterococcus faecalis* GelE protease in determination of cellular chain length, supernatant pheromone levels, and degradation of fibrin and misfolded surface proteins. *J Bacteriol* **185**(12): 3613-23.

- Watt, V. M., C. J. Ingles, M. S. Urdea and W. J. Rutter** (1985). Homology requirements for recombination in *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**(14): 4768-72.
- Wenner, T., V. Roth, G. Fischer, C. Fourrier, B. Aigle, B. Decaris and P. Leblond** (2003). End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*. *Mol Microbiol* **50**(2): 411-25.
- Whatmore, A. M., V. A. Barcus and C. G. Dowson** (1999). Genetic diversity of the streptococcal competence (*com*) gene locus. *J Bacteriol* **181**(10): 3144-54.
- Whatmore, A. M. and M. A. Kehoe** (1994). Horizontal gene transfer in the evolution of group A streptococcal *emm-like* genes: gene mosaics and variation in Vir regulons. *Mol Microbiol* **11**(2): 363-74.
- Wilton, J. L., A. L. Scarman, M. J. Walker and S. P. Djordjevic** (1998). Reiterated repeat region variability in the ciliary adhesin gene of *Mycoplasma hyopneumoniae*. *Microbiology* **144** (Pt 7): 1931-43.
- Wootton, J. C. and M. H. Drummond** (1989). The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng* **2**(7): 535-43.
- Wootton, J. C. and S. Federhen** (1996). Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* **266**: 554-71.
- Wuenscher, M. D., S. Kohler, A. Bubert, U. Gerike and W. Goebel** (1993). The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity. *J Bacteriol* **175**(11): 3491-501.
- Yang, Y. and D. W. Gabriel** (1995). Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *J Bacteriol* **177**(17): 4963-8.
- Yang, Z., L. Shipman, M. Zhang, B. P. Anton, R. J. Roberts and X. Cheng** (2004). Structural characterization and comparative phylogenetic analysis of *Escherichia coli* HemK, a protein (N5)-glutamine methyltransferase. *J Mol Biol* **340**(4): 695-706.
- Yeramian, E. and H. Buc** (1999). Tandem repeats in complete bacterial genome sequences: sequence and structural analyses for comparative studies. *Res Microbiol* **150**(9-10): 745-54.
- Zhang, Q. and K. S. Wise** (1996). Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent *vaa* genes. *Infect Immun* **64**(7): 2737-44.

Annexes

Article 1

Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress

Thibessard A, Borges F, Fernandez A, Gintz B, Decaris B, Leblond-Bourget N
Appl Environ Microbiol. 2004 Apr;70(4):2220-9.

Article 2

cse*, a chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus

Borges F, Layec S, Thibessard A, Fernandez A, Gintz B, Hols P, Decaris B et Leblond-Bourget N
J bacteriol. 2005 Apr;187(8):2737-46.

Acte de congrès :

Le gène mosaïque et variable *cse* est impliqué dans la segregation cellulaire chez *Streptococcus thermophilus*

Borges F, Layec S, Thibessard A, Fernandez A, Gintz B, Decaris B et Leblond-Bourget N
Communication orale au CBL 2004, Aurillac.

Article 3

Characterization of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368

Fernandez A, Thibessard A, Borges F, Gintz B, Decaris B, Leblond-Bourget N.
Arch Microbiol. 2004 Nov;182(5):364-72.

Article 4

Characterisation of *Streptococcus thermophilus* CNRZ368 oxidative stress-resistant mutants: involvement of a potential Rgg-like transcriptional regulator

Fernandez A, Borges F, Thibessard A, Gintz B, Decaris B, Leblond-Bourget N
Lait 2004; 84:77-85.

Identification of *Streptococcus thermophilus* CNRZ368 Genes Involved in Defense against Superoxide Stress

Annabelle Thibessard, Frédéric Borges, Annabelle Fernandez, Brigitte Gintz, Bernard Decaris,* and Nathalie Leblond-Bourget

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences de l'Université Henri Poincaré Nancy 1, 54506 Vandoeuvre-lès-Nancy, France

Received 18 September 2003/Accepted 11 December 2003

To better understand the defense mechanism of *Streptococcus thermophilus* against superoxide stress, molecular analysis of 10 menadione-sensitive mutants, obtained by insertional mutagenesis, was undertaken. This analysis allowed the identification of 10 genes that, with respect to their putative functions, were classified into five categories: (i) those involved in cell wall metabolism, (ii) those involved in exopolysaccharide translocation, (iii) those involved in RNA modification, (iv) those involved in iron homeostasis, and (v) those whose functions are still unknown. The behavior of the 10 menadione-sensitive mutants exposed to heat shock was investigated. Data from these experiments allowed us to distinguish genes whose action might be specific to oxidative stress defense (*tgt*, *ossF*, and *ossG*) from those whose action may be generalized to other stressful conditions (*mreD*, *rodA*, *php2b*, *cpsX*, and *iscU*). Among the mutants, two harbored an independently inserted copy of pGh9:ISS1 in two loci close to each other. More precisely, these two loci are homologous to the *sufD* and *iscU* genes, which are involved in the biosynthesis of iron-sulfur clusters. This region, called the *suf* region, was further characterized in *S. thermophilus* CNRZ368 by sequencing and by construction of Δ *sufD* and *iscU*₉₇ nonpolar mutants. The streptonigrin sensitivity levels of both mutants suggest that these two genes are involved in iron metabolism.

Oxygen is an essential element for the viability of most living organisms through its involvement in oxidative phosphorylation that takes place during the aerobic respiration process. However, oxygen can also be harmful due to the production, especially during respiration, of toxic derivatives, namely, reactive oxygen species (ROS) such as hydrogen peroxide, hydroxyl radicals, and superoxide radicals (for a review, see reference 25). The accumulation of these molecules triggers lipid, protein, and nucleic acid oxidation, causing lethal cellular damage (for reviews, see references 7, 25, and 53). In response to such attacks, aerobic microorganisms have evolved antioxidant mechanisms (e.g., catalase, peroxidase, and superoxide dismutase activities) to protect and adapt cells against ROS excess (6, 18, 24, 65, 68).

Although they do not achieve aerobic respiration, anaerobic microorganisms can be subject to ROS. The strict anaerobes *Bacteroides fragilis* and *Clostridium perfringens* are able to survive in the presence of a small amount of oxygen due to the induction of an adaptive oxidative stress response that is activated under aerobic conditions (13, 46–49). The peroxide response of *B. fragilis* shares some characteristics with that of *Escherichia coli*, such as the involvement of the redox-sensitive transcriptional activator OxyR (46) and some antioxidant activities under its control (47, 49).

The lactic acid bacterium (LAB) *Streptococcus thermophilus* is a gram-positive bacterium widely used in dairy factories, mainly as a starter of fermentation in yogurt and cheese pro-

duction. During fermentation processes, *S. thermophilus* is exposed to oxygen and to its deleterious effects, which can lead to a reduction of cell viability and consequently may have repercussions on the texture, flavor, and safety of the final product.

S. thermophilus is an anaerobic aerotolerant organism that can grow in the presence of oxygen and survive to low concentrations of superoxide and hydroxyl radicals (59). This ability suggests that *S. thermophilus* displays a defense system that provides protection against the effects of ROS. In the literature, data about the oxidative stress response of this organism are scarce. The existence of at least one ROS defense system and its inducibility by hydrogen peroxide are established for *S. thermophilus* CNRZ368 (58). This organism possesses a single manganese-containing superoxide dismutase (19) and a glutathione reductase (43). However, no catalase activity was found. In other streptococci, antioxidant enzymes were reported to play roles in defense against oxidative stress (27, 33, 44, 66) and in pathogenesis (45, 67, 69).

The aim of this work was to identify genes involved in the oxidative stress defense of *S. thermophilus* CNRZ368 without presuming their function. For this purpose, a collection of mutants was created by random insertional mutagenesis and the mutants were screened for their sensitivity to menadione (a superoxide-generating agent). The locus disrupted in 10 *S. thermophilus* menadione-sensitive clones was identified, and the potential physiological role of the mutated genes in oxidative stress defense is discussed. The implications of these 10 loci during heat shock were also evaluated to assess their specificity to oxidative stress defense. Moreover, the characterization of two of these mutants, disrupted in the *sufD* and *iscU* genes, which may participate in [Fe-S] cluster assembly or repair, is more detailed. The results presented in this paper

* Corresponding author. Mailing address: Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, UHP, Faculté des Sciences, 54506 Vandoeuvre-lès-Nancy, France. Phone: (33)3 83 68 42 12. Fax: (33)3 83 68 44 99. E-mail: decaris@nancy.inra.fr.

indicate that the inactivation of *sufD* and *iscU* genes caused higher sensitivity to superoxide stress and suggest that [Fe-S] cluster repair or assembly could play an important role in the oxidative stress defense of anaerobic bacteria.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *S. thermophilus* CNRZ368 was purchased from the Institut National de la Recherche Agronomique-Centre National de Recherches Zootechniques (CNRZ) (Jouy en Josas, France) strain collection. Depending on the experiments, *S. thermophilus* CNRZ368 and its derivatives were cultivated at 30 or 42°C in milk medium (for precultures), TPPY medium (12), or M17 medium (55) without shaking. Erythromycin was added at 2 µg · ml⁻¹ when required. Streptomycin sensitivity was monitored by using a TPPY medium which contained a reduced concentration of metallic dications (including Fe²⁺). TPPY was prepared according to the protocol described by Visca et al. (62) and modified as follows: Chelex beads (20 g/liter; Sigma) were added to TPPY medium and stirred at 4°C for 48 h, and the medium was then supplemented with 1 mM MgCl₂ and 1 mM CaCl₂.

Recombinant plasmids were transformed into *E. coli* EC101, a TG1 strain [*supE hsd-5 thi Δ(lac-proAB) F' (traD6 proAB⁺ lacI^r lacZΔM15)*] containing a chromosomal copy of the pWV01 *repA* gene (34) and selected at 37°C on Luria-Bertani (LB) (50) agar plates containing 150 µg of erythromycin/ml.

Insertional mutagenesis. The collection of mutants was obtained by insertional mutagenesis with the plasmid pGh9:ISS1, which carries the Em^r gene (35). The mutagenesis procedure and the cloning of the fragments flanking the pGh9:ISS1 insertion locus were carried out as previously described (56).

Selection of mutants impaired in their ability to tolerate menadione. For the selection of menadione-sensitive clones, two different procedures, differing mainly in the length of the menadione exposure, were assayed. Both assays were preceded by a preculture in milk medium containing 2 µg of erythromycin/ml; precultures of the 2,112 mutants were performed with 96-well microtiter plates.

For long-term menadione exposure, each preculture was transferred onto M17 agar plates containing 2 µg of erythromycin/ml with increasing concentrations of menadione (from 0 to 120 µg/ml). After 20 h of incubation at 42°C, the presence (or absence) of colonies was observed.

For short-term menadione exposure, each preculture was diluted (at a 1/100 ratio) in five different 96-well microtiter plates containing liquid TPPY medium supplemented with 2 µg of erythromycin/ml (in a final volume of 100 µl, resulting in an approximately 1:1 liquid-to-air space ratio). After a 3.5-h growth period at 42°C, five different concentrations of menadione (from 0 to 15 mg/ml) were added to each of the five microtiter plates. Microtiter plates were incubated at 42°C for 3 h. The cultures were then transferred onto TPPY agar plates containing erythromycin (2 µg/ml) by using a hedgehog. The ability of the mutants to survive the menadione stress was assessed after 20 h of incubation at 42°C by observation of the presence or absence of colonies.

For each experiment, clones were considered sensitive when they did not grow on a plate on which more than 90% of the clones were growing. Both procedures were applied twice, and the results were compiled to reveal the mutants exhibiting the more severe phenotype.

DNA manipulations. The preparation of chromosomal, plasmid, and recombinant DNA and Southern analysis were performed according to standard protocols (50). All cloning junctions and cloned PCR products were checked by sequencing with Dye-Terminator chemistry on an ABI Prism 377 genetic analyzer (PE Biosystems). Sequence data were analyzed by using GenBank, BLAST (2, 3), TMHMM (51), and HMMTOP (61) software.

Construction of the Δ*sufD* and *iscU*₉₇ mutants. Oligonucleotides used to construct the Δ*sufD* and *iscU*₉₇ mutants were purchased from MWG Biotech AG (Ebersberg, Germany). To construct the nonpolar truncated Δ*sufD* mutant, the 5' and 3' ends of the *sufD* gene were independently amplified by PCR. The 5' end was amplified by using the primers GAGCGTGTGAATTCACCG, containing an EcoRI site (in boldface type) and AACTTAACGACGACCGTCT, containing a PstI site (in boldface type). The 3' end was amplified by using the primers GGTGTTCCAGAATTCAAATAACC, containing an EcoRI site, and TGGTCTGTCATCAGGTCG. The latter primer does not contain a PstI site, but the PCR product generated does contain a PstI restriction site. These two PCR products were digested by PstI and ligated to each other. The resulting fragment was amplified by PCR and then digested by EcoRI and cloned into pGh9. The final construct (pGh9Δ*sufD*) was used to transform *S. thermophilus* CNRZ368. For IscU protein inactivation, the strategy consisted of replacing the TGT codon encoding the conserved Cys₉₇ with a TAT codon encoding a Tyr. This was achieved by replacing, in oligonucleotides, the three underlined nucle-

otides of the sequence AGG TGT ACA, generating an EcoRV restriction site (AGA TAT CCA). Thus, the three amino acids Gly₉₆, Cys₉₇, and Tyr₉₈ were replaced by Arg₉₆, Tyr₉₇, and Thr₉₈, respectively, in the mutated protein. Two internal and complementary primers containing the mutation (GCTGGTGAT GATATCACCATTTCAACA and GAAATGGTGATATCATCACCAGCAA AGG) were used with two additional external primers (AAGCGGAATTCAG CATCCACTCAGGC, containing an EcoRI site, and TGTGACAAGCTGTGA GAGGCAATTCTT, containing a HindIII site), respectively, to amplify the 5' and 3' ends of *iscU*. The two PCR products, of 515 and 262 bp, were digested by EcoRI and EcoRV and by HindIII and EcoRV, respectively, before being ligated to each other. The resulting fragment was cloned into pGh9, which was digested by EcoRI and HindIII. The final construct, pGh9*iscU*₉₇, was used to transform *S. thermophilus* CNRZ368. Preparation of *S. thermophilus* competent cells was performed as previously described by O'Sullivan and Fitzgerald (41). The following gene replacements were done as previously described (8) but in TPPY medium and with 30 and 42°C as the permissive and restrictive temperatures, respectively.

Physiological characterization of *S. thermophilus* mutant strains. For short-term methyl viologen exposure, cultures of CNRZ368 and its derivative strains were grown in TPPY medium. At an optical density at 600 nm (OD₆₀₀) of 0.6, cultures were divided into two subcultures (at a ratio of liquid to air space of approximately 1:1). Methyl viologen (paraquat; Sigma-Aldrich) was added to one of the subcultures at a final concentration of 50 mM. After 3 h of incubation at 42°C, cultures were diluted to appropriate concentrations and spread onto TPPY agar plates. For long-term methyl viologen exposure, strains were grown in TPPY medium until they reached an OD₆₀₀ of 0.6. Cultures were then diluted to a final theoretical OD₆₀₀ of 0.001 in TPPY medium (at a ratio of liquid to air space of approximately 5:1) supplemented with 1.7 mM methyl viologen or not supplemented (control experiment). After 18 h of incubation at 42°C, cultures were diluted in TPPY to appropriate concentrations and spread onto TPPY agar plates. In both procedures, methyl viologen sensitivity was estimated by comparing CFU-per-milliliter values obtained for treated cultures with those obtained for nontreated cultures.

Heat shock. The ability of *S. thermophilus* and its derivatives to respond to a heat stress was assessed as follows. Each strain was grown at 42°C in TPPY medium until the early log phase (OD₆₀₀ = 0.3). Then, the culture was divided into two parts. One part was kept at 42°C (control experiment); the other one was immediately transferred to 60°C for 15 min. Appropriate dilutions were spread on TPPY agar plates. The fraction of surviving cells was calculated by dividing the surviving number of CFU per milliliter by the count for the control experiment.

Streptomycin sensitivity. Streptomycin sensitivity was monitored for cells grown overnight in TPPY medium at 42°C. Cells were harvested by centrifugation, washed twice in TPPY medium, resuspended in TPPY, diluted to an OD₆₀₀ of 0.01, and cultivated in TPPY for 9 h at 42°C. Then, the cultures were diluted again to an OD₆₀₀ of 0.005 and divided into four parts, which were all incubated for 21 h at 42°C. One part was supplemented with 0.4% (vol/vol) dimethyl formamide (control experiment), and the other three parts were supplemented with different concentrations of streptomycin (1, 2, and 4 µg/ml, dissolved in dimethyl formamide). As a control experiment, cultures were also carried out with TPPY supplemented with 1 mM of 2'-dipyridyl (an Fe chelator). In each experiment, the OD₆₀₀ value was used to estimate cellular density after 21 h of incubation. The relative optical density was determined by comparing the cellular densities obtained with and without streptomycin. All experiments were performed in triplicate.

Nucleotide sequence accession numbers. For all analyzed sequences, the DNA sequences reported in this paper have been deposited in GenBank under accession numbers AY386239, AY386240, AY386241, AY386242, AY386243, AY386244, and AY386245.

RESULTS

Selection of *S. thermophilus* CNRZ368 mutants with altered oxidative stress response. A collection of 2,112 mutants of *S. thermophilus* CNRZ368 was generated by insertional mutagenesis with the plasmid pGh9:ISS1 (35, 56). This plasmid carries an erythromycin resistance gene for its selection and the insertional element ISS1, which allows random integration of the plasmid into the *S. thermophilus* genome. The 2,112 mutants obtained were screened to select clones sensitive to menadione

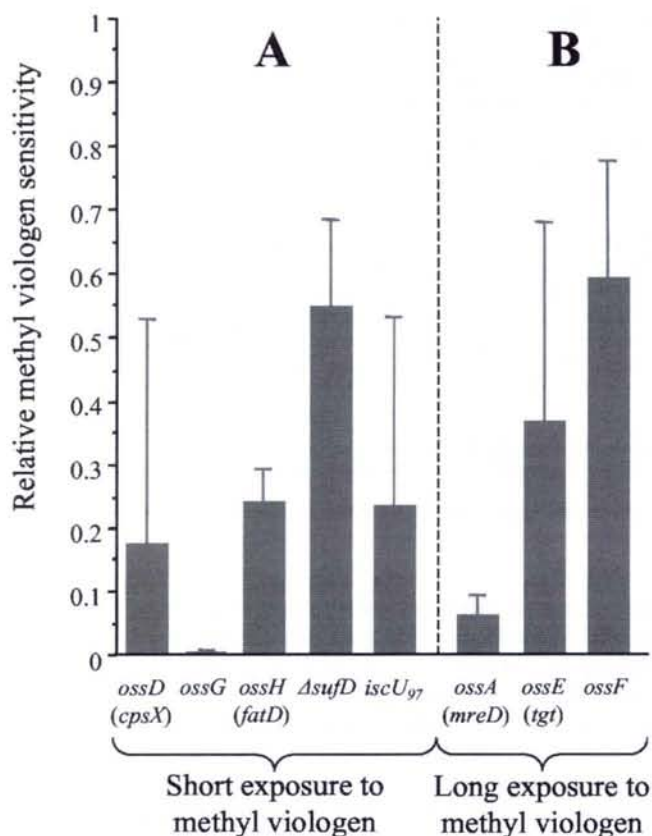


FIG. 1. Relative methyl viologen sensitivity. (A) Results of short-term methyl viologen exposure. (B) Results of long-term methyl viologen exposure (see Materials and Methods). The relative methyl viologen sensitivity was determined by comparing viabilities with and without treatment. Levels are shown as values relative to the survival level of the wild-type strain. All experiments were performed at least in duplicate. Error bars show two standard deviations of the data.

(a superoxide-generating compound). Two distinct selection procedures were used to identify mutants susceptible to menadione (see Materials and Methods). The characterization of 10 mutants, disrupted in loci originally named *ossA* to *ossJ* (for oxidative stress-sensitive locus), is described in this paper.

Sensitivity of *S. thermophilus* CNRZ368 mutants to methyl viologen. Previously, we demonstrated that *ossB* and *ossC* mutants (previously named 6G4 and 18C3, respectively) are compromised in their ability to survive superoxide stress (57). To confirm that the other eight mutants selected were sensitive to superoxide stress, their sensitivity to methyl viologen was examined (Fig. 1). In this experiment, we used the nonpolar mutants Δ *sufD* and *iscU*₉₇ constructed in this work (see below) rather than the potential polar clones *ossI* and *ossJ*. Short-term methyl viologen exposure revealed the extreme sensitivity of *ossG*, as well as the four- to fivefold increased sensitivity of the *ossD*, *ossH*, and *iscU*₉₇ mutants (Fig. 1A). Although its sensitivity was less severe, the Δ *sufD* mutant exhibited about twofold less survival following such stress. Under these conditions, the other three mutants (*ossA*, *ossE*, and *ossF*) did not exhibit any significant sensitivity; however, long-term methyl viologen exposure confirmed their sensitivity. Indeed, long-term methyl viologen exposure (Fig. 1B) showed that the *ossA* mutant strain displayed a relative survival rate more than 10-fold lower than that of the wild-type strain. The *ossE* and *ossF* mutants also showed a twofold lower survival rate than the wild-type strain.

Survival of the mutants exposed to heat stress. To assess the specificity of these 10 loci to oxidative stress defense, the abilities of the mutants to survive to heat shock were estimated. Five out of the 10 mutants (*ossA*, *ossB*, *ossC*, *ossD*, and *ossJ*) isolated as oxidative stress-sensitive clones also showed reduced survival with heat shock compared with the wild type (Fig. 2). These data suggest that the genes impaired in these mutants are not specific to the oxidative stress defense process but are probably involved in a more global response to stresses.

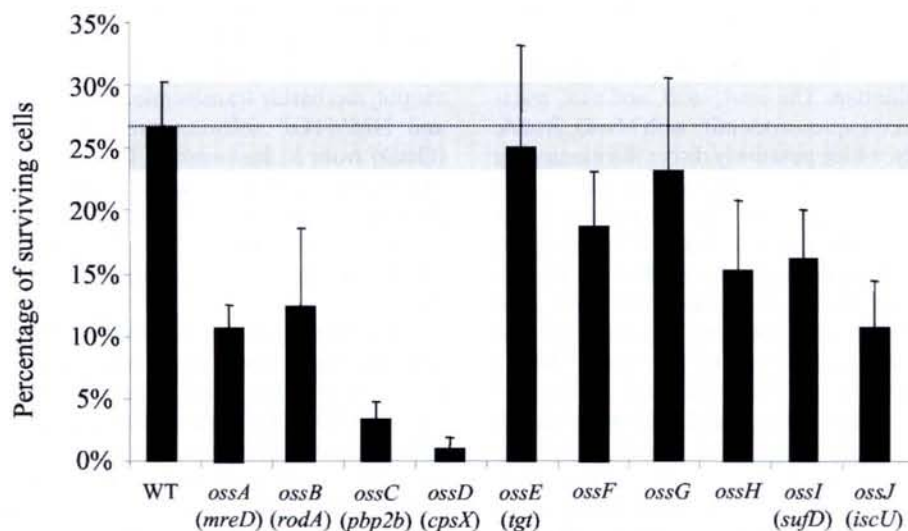


FIG. 2. Survival after heat shock. The heat shock treatment consisted of 15 min at 60°C. The percentage of surviving cells corresponds to the number of CFU counted on plates after the heat shock exposure divided by the number of CFU before treatment. Error bars show two standard deviations of the data. The dotted line represents the surviving percentage observed for the wild-type (WT) strain, and the gray band comprises two standard deviations of the data.

TABLE 1. Products of *oss* genes: database similarities and putative functions

Gene name ^a	Accession no.	Gene encoding the most similar product ^c	Amino acid identity ^d	Insertion site ^e	Putative function	Class
<i>ossA</i> (<i>mreD</i>)	AY386242	<i>S. mutans mreD</i> (168)	30 (1–168)	127	Cell shape maintenance	I
<i>ossB</i> (<i>rodA</i>)	AF399832	<i>S. mutans rodA</i> (408)	65 (1–391)	304	Peptidoglycan elongation	I
<i>ossC</i> (<i>pbp2b</i>) ^b	AF399833	<i>S. thermophilus</i> Sfi6 <i>pbp2b</i> (704)	98 (641–704)	640	Peptidoglycan elongation	I
<i>ossD</i> (<i>cpsX</i>) ^b	AY386241	<i>S. thermophilus cpsX</i> (378)	98 (212–378)	264	Membrane translocation of exopolysaccharide	II
<i>ossE</i> (<i>tgt</i>) ^b	AY386243	<i>S. pneumoniae tgt</i> (380)	92 (1–333)	99	tRNA-guanine transglycosylase	III
<i>ossF</i> ^b	AY386244	<i>S. agalactiae gbs0448</i> (451)	77 (1–204)	51	RNA methyltransferase	III
<i>ossG</i>	AY386239	<i>S. mutans smu.1428c</i> (284)	67 (1–280)	170	Unknown conserved function	IV
<i>ossH</i> (<i>fatD</i>)	AY386240	<i>S. pneumoniae fatD</i> (324)	55 (22–323)	119	Iron ABC transporter	V
<i>ossI</i> (<i>stfD</i>)	AY386245	<i>S. agalactiae sag0142</i> (420)	76 (1–420)	144	Assembly of Fe/S clusters	V
<i>ossJ</i> (<i>iscU</i>)	AY386245	<i>S. agalactiae gbs0140</i> (147)	78 (1–141)	57	Assembly of Fe/S clusters	V

^a Initial gene name with the definitive gene name in parentheses where applicable.

^b Disrupted ORF of *S. thermophilus* CNRZ368 that has been partially sequenced.

^c Organism, gene name, and (in parentheses) length (in amino acids) of the protein sequence giving the best alignment score.

^d Percentage of sequence identity with the protein showing the best alignment score with the *S. thermophilus* Oss protein. In parentheses is the portion of the protein sharing this identity (numbers correspond to amino acid positions).

^e Deduced amino acid position of the pGh9:ISS1-generated disruption in the full-length protein showing the best alignment score with the Oss protein.

In contrast, the *ossE*, *ossF*, and *ossG* mutants did not differ significantly from the wild type, indicating that the genes disrupted in these strains did not influence the ability of cells to respond to heat shock. Interestingly, the *ossG* mutant, which was highly sensitive to methyl viologen, showed no significant decrease of survival when subjected to heat shock compared with that of the wild type, suggesting that the *ossG* gene, essential for oxidative stress defense, could be specific to such a stress.

Identification of genes implicated in oxidative stress tolerance. The disrupted locus was identified for each of the 10 mutants by cloning and sequencing the junctions as previously described (56). All of the insertions were mapped within coding sequences. Sequence similarity searches and analysis revealed similar sequences in other bacteria (see Table 1), with an E-value lower than 10^{-10} for BLASTX searches. In 9 of 10 cases, data about the functions of these genes were available. To facilitate their description, these genes were grouped into five classes (classes I to V) according to their putative functions and renamed when relevant (Table 1).

(i) Class I: genes potentially involved in cell wall structure and cell shape determination. The *ossA*, *ossB*, and *ossC* putative gene products shared sequence identity with MreD, RodA, and Pbp2b, respectively, which putatively direct the elongation of the lateral cell wall in *E. coli*, conferring on the cell its characteristic rod shape (5, 28, 29). The *ossA* mutant was defective in a gene encoding a protein with 30% identity to MreD of *Streptococcus mutans*. Sequence analysis of the surrounding region indicated that the *S. thermophilus mreD* gene (*ossA*) was clustered with a gene encoding a putative protein similar to MreC. Moreover, hydrophathy profiles of MreD (*OssA*) determined by using TMHMM and HMMTOP showed five regions of sufficient length and hydrophobicity to be membrane-spanning segments (data not shown), suggesting that this protein is the counterpart of MreD. The relatively low identity (30%) observed between *S. thermophilus* MreD and *S. mutans* MreD was in agreement with the general idea that the structure of this protein is more highly conserved than its primary sequence (17). To study the effect of *mreD* disruption in *S. thermophilus*, TPPY cultures of the wild-type and *mreD*

mutant strains were observed with light microscopy. In contrast to wild-type cells, *mreD* mutant cells formed very long chains, with up to 180 cells per chain (Fig. 3). Moreover, a reduction of about 25% of the cell size was observed in the mutant. The growth of the *mreD* mutant and that of the wild type in TPPY medium were monitored by measuring the OD₆₀₀. Both strains displayed doubling times of 30 min, but the maximal OD₆₀₀ of the *mreD* mutant (1.7 ± 0.1) was lower than that of the wild type (2.4 ± 0.2), showing that *mreD* mutant cells entered the stationary phase earlier than wild-type cells. The characterization of *rodA* and *pbp2b* mutants of *S. thermophilus* CNRZ368 was described in a previous paper (57).

(ii) Class II: genes potentially involved in exopolysaccharide translocation. The sequence disrupted in the *ossD* mutant was partially determined. Its analysis indicated that the C-terminal portion of the predicted encoded protein showed 98% identity to CpsX, a putative membrane protein from *S. thermophilus* MR-1C. The role of CpsX is still unknown, but the conserved location of *cpsX* near *eps* genes of *S. thermophilus* (14) and the presence of a putative gene related to an ABC transport system in its vicinity argue in favor of CpsX being involved in exopolysaccharide translocation (15). Moreover, the TMHMM and HMMTOP software predictions indicated that CpsX (*OssD*) from *S. thermophilus* CNRZ368 would contain highly hydrophobic domains compatible with a location in the cytoplasmic membrane. Although the function of *cpsX* is not known yet, our work indicates that its disruption is associated with sensitivity not only to superoxide stress but also to heat shock.

(iii) Class III: genes involved in RNA maturation. In the *ossE* and *ossF* mutants, the plasmid inserted into genes whose predicted products share sequence identity with enzymes involved in RNA maturation. The *ossE* gene encodes a putative tRNA guanine transglycosylase (TGT) which shares 92% identity with the putative *Streptococcus pneumoniae* TGT. TGT enzyme takes part in the process that removes the guanine residue at position 34 (the wobble position of the anticodon loop) of tRNA^{His}, tRNA^{Asp}, tRNA^{Asn}, and tRNA^{Tyr} and replaces it with the modified residue queuosine (39, 40).

ossF potentially encodes a protein that shows 77% identity

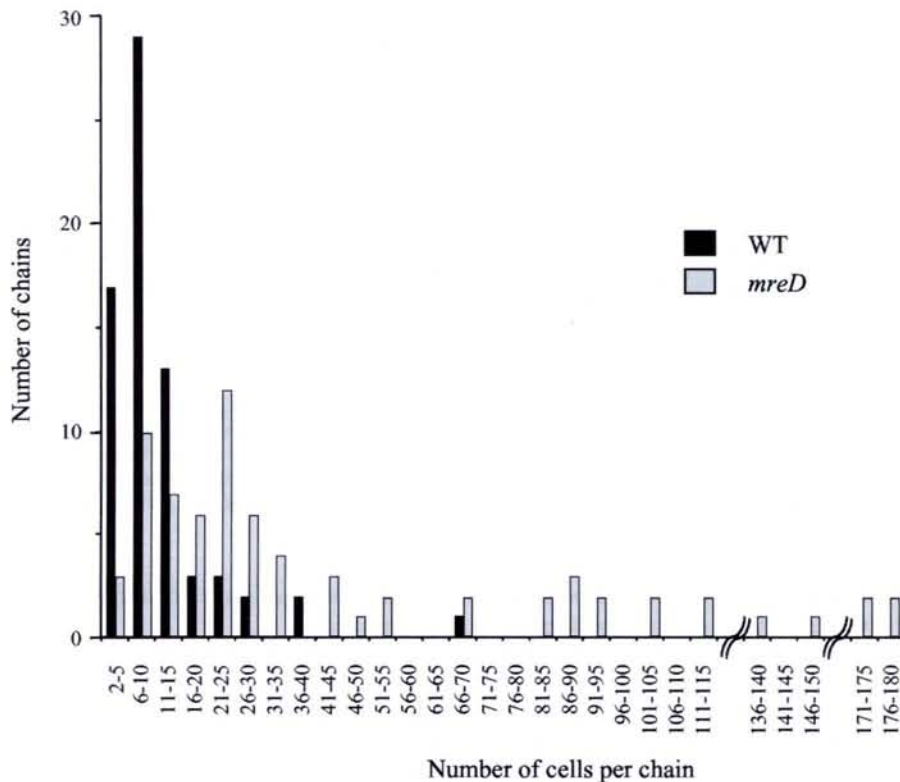


FIG. 3. Influence of *mreD* disruption on chain length. Cell number per chain was determined for 70 chains of the wild-type (WT) and *mreD* mutant strains grown in TPTY liquid medium.

to a putative RNA methyltransferase of *Streptococcus agalactiae* (Gbs0448) belonging to the TrmA family. Enzymes of the TrmA family catalyze the formation of the 5-methyluridine in position 54 (localized in the T Ψ C loop) of tRNA (9). The presence of a TRAM-like domain (RNA binding domain) in the N-terminal part of OssF supports the idea of an interaction between this protein and RNA.

(iv) **Class IV: genes of unknown function.** The *ossG* product shows 67% identity with the product of an *S. mutans* gene of unknown function (*smu.1428c*) (Table 1). Although the function of *ossG* is not known yet, the mutation of this gene is now associated with a phenotype of sensitivity to superoxide stress. The TMHMM and HMMTOP software predictions suggest the existence of two transmembrane domains in the N-terminal region of OssG, making OssG a potential membrane protein.

(v) **Class V: genes involved in iron metabolism.** (a) **Ferric iron transport.** The putative product of *ossH* shows 55% identity to FatD from *S. pneumoniae*, a potential membrane-spanning permease of an Fe³⁺ ABC transporter. Analysis of the adjacent sequence reveals the presence of a gene encoding a putative protein similar to FatC downstream of *fatD* (*ossH*) in the same orientation. The *fatC* gene is also clustered with *fatD* in the *S. pneumoniae* genome. The hydrophathy profile of FatD (OssH) from *S. thermophilus*, assessed with TMHMM and HMMTOP, reveals that this protein displays 10 putative-membrane spanning segments compatible with its expected cytoplasmic membrane location.

(b) **[Fe-S] cluster biosynthesis.** The sequence affected in the *ossI* mutant encodes a product that shows 75% identity to

Sag0142, a putative protein deduced from the complete genome sequence of *S. agalactiae* and similar to SufD-like proteins. Analysis of totally sequenced microbial genomes reveals that *sufD*-like genes are conserved in bacteria and also in some archaea, such as *Halobacterium* sp. NRC-1, *Thermoplasma acidophilum*, and *Aeropyrum pernix* (23). In *E. coli* and *Erwinia chrysanthemi*, the *sufD* gene belongs to the *sufABCDSE* cluster identified as being involved in iron metabolism and, more precisely, in [Fe-S] cluster biogenesis and/or repair (37, 42). The *ossJ* gene product shows sequence identity to the *gbs0140* product from *S. agalactiae* (Table 1), probably encoding a IscU protein, which is a NifU family protein. NifU proteins have been proposed to serve as scaffold proteins in [Fe-S] cluster assembly in *Synechocystis* PCC6803 and would function as a mediator for iron-sulfur cluster delivery (38).

In some genomes, including that of *Streptococcus pyogenes*, *nifU*-like genes are clustered with the *suf* genes (23). To determine whether *sufD* (*ossI*) and *iscU* (*ossJ*) from *S. thermophilus* were genetically linked together, the chromosomal DNA regions flanking the pGhost9:ISS1 insertion sites of both mutants were sequenced until they overlapped. Examination of the genes in the neighborhood of *sufD* and *iscU* revealed that they both belong to a cluster of five genes all orientated in the same direction. This cluster, named the *sufCDS iscU sufB* cluster, is shown in Fig. 4. The highest sequence identities found for these gene products are those to putative proteins from *S. agalactiae* (Table 2). They also share sequence similarity with proteins encoded by *E. coli* and *Erwinia chrysanthemi* *suf* gene clusters. The putative SufC, SufD, SufS, and SufB proteins

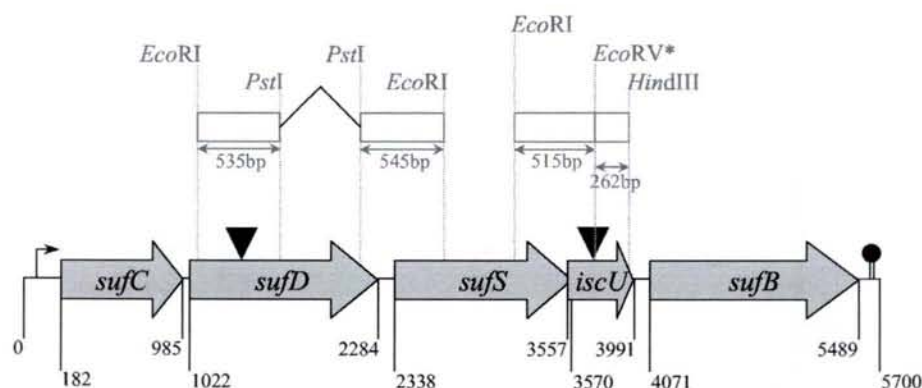


FIG. 4. Schematic map of the *S. thermophilus* *sufCDS iscU sufB* cluster. Genes are indicated by gray arrows, and the positions of putative start and stop codons are given (in base pairs) below the map. The sites of the insertion of the pGh9:ISS1 plasmid into *sufD* (*ossJ*)- and *iscU* (*ossJ*)-disrupted mutants are indicated by filled triangles. The broken arrow indicates the position of the *suf* cluster putative promoter. The putative rho-independent terminator is represented as hairpin loop. Fragments used for the generation of the nonpolar mutants are represented by open boxes. An asterisk indicates the position of the EcoRV restriction site causing the replacement of Cys₉₇ with a Tyr in the *iscU*₉₇ mutant.

encoded by the *S. thermophilus* cluster present, respectively, 72, 48, 65, and 57% similarity to their *E. coli* counterparts. Moreover, the *S. thermophilus* IscU presents 47% similarity with IscU from *E. coli*.

Sequences corresponding to promoter regions were searched by inspection of the total sequence of the *sufCDS iscU sufB* cluster. A unique extended -10 sequence (5'-TTT GTTATAAT-3') was found upstream of *sufC*, the first open reading frame (ORF) of the cluster. A potential rho-independent transcriptional terminator was identified 63 bp downstream of *sufB*, the last ORF of the cluster ($\Delta G = -10$ kcal/mol). These features, together with the short or missing intergenic regions and the same orientation of the genes, suggest that these five ORFs form an operon, as is the case for the *Erwinia chrysanthemi* *suf* locus.

Molecular characterization of the *sufCDS iscU sufB* cluster.

(i) **Construction of Δ *sufD* and *iscU*₉₇ mutants.** Since *S. thermophilus* *suf* genes are potentially clustered in an operon, nonpolar Δ *sufD* and *iscU*₉₇ mutants were constructed and analyzed. The *sufD* mutation consists of an internal in-frame deletion leading to the loss of 43% of the whole protein (see Materials and Methods and Fig. 4). Sequence comparison of NifU family proteins from various organisms indicates that the *S. thermophilus* IscU protein contains the three highly conserved cysteine residues that have been shown to be the assembly site of a transitory [2Fe-2S] center within NifU (70). Thus, the construction of a null *iscU* mutant was achieved by replacing one of these conserved residues, the Cys₉₇, with a Tyr (see Materials and Methods and Fig. 4). Sequencing of these

two constructs confirmed that they did not contain any secondary mutation.

(ii) **Role of *sufD* and *iscU* in iron metabolism.** To investigate the possibility that Δ *sufD* and *iscU*₉₇ mutations impair growth due to the lack of functional [Fe-S] cluster biosynthesis, the mutants were grown in TPPY medium and compared to the wild type. The growth of the Δ *sufD* and *iscU*₉₇ mutants showed no significant difference from that of the wild-type strain. Assuming that the high iron concentration of the TPPY medium could allow self-assembly of the [Fe-S] clusters (71) and consequently could be responsible for the absence of a difference in growth rates between strains, the same experiment was conducted in TPPYC, a divalent cation-depleted medium. Under these conditions, the mutants displayed a growth rate similar to that of the wild-type strain, but their maximal OD₆₀₀ was lower than that of the wild type, showing their premature entry into the stationary phase (Table 3). Moreover, this difference disappeared when the cultures were grown on TPPYC supplemented with 100 μ M FeSO₄. Taken together, these results suggest (i) that *iscU* and *sufD* genes of *S. thermophilus* are involved in iron metabolism and (ii) that the presence of these genes would be of particular importance under iron-limiting conditions.

To confirm the involvement of *sufD* and *iscU* in iron metabolism, the iron intracellular concentrations of the Δ *sufD* and *iscU*₉₇ mutants were checked by measuring the sensitivities of these strains to streptonigrin (an iron-activated antibiotic). As shown in Fig. 5A, the *iscU* mutant was more sensitive to streptonigrin than the wild-type strain was, whereas the Δ *sufD* mu-

TABLE 2. *suf* cluster gene products of *S. thermophilus* CNRZ368

Gene product	Closest similar protein ^a	Identity, similarity (%)	Function and/or nature of protein
SufC	CAD45782 (<i>S. agalactiae</i> NEM316)	88, 94	Putative ABC transporter (ATP-binding protein)
SufD	AAM99050 (<i>S. agalactiae</i> 2603V/R)	76, 87	Unknown, conserved protein
SufS	CAD45784 (<i>S. agalactiae</i> NEM316)	79, 89	Similar to SufS
IscU	CAD45785 (<i>S. agalactiae</i> NEM316)	78, 90	Similar to NifU family proteins
SufB	CAD45786 (<i>S. agalactiae</i> NEM316)	89, 94	Unknown, conserved protein

^a Accession number of the protein giving the best alignment score. The name of the organism is given in parentheses.

TABLE 3. Growth of wild-type and mutant *S. thermophilus* strains in TPPYC and in TPPYC supplemented with iron^a

Strain	Final OD ₆₀₀ resulting from growth in:	
	TPPYC	TPPYC + FeSO ₄
Wild type	0.96	1.32
<i>ΔsufD</i>	0.58	1.48
<i>iscU₉₇</i>	0.58	1.15

^a Cells were grown in milk medium for 15 h at 42°C and diluted in FeSO₄-deficient medium (TPPYC medium supplemented with MgCl₂ [1 mM] and CaCl₂ [1 mM]). After 1 night of growth, cells were washed twice and resuspended either in FeSO₄-deficient medium or in medium supplemented with 100 μM FeSO₄. The growth of the cells was then followed by the measurement of the OD₆₀₀. The OD₆₀₀ was measured after 24 h of incubation at 42°C. This experiment was done in triplicate.

tant was more resistant. Moreover, the addition of 2,2'-dipyridyl (a ferrous iron chelator) abolished the difference in growth rates between the strains (Fig. 5B), indicating that iron was responsible for the streptonigrin sensitivity observed without 2,2'-dipyridyl. Taken altogether, these results suggest that both *iscU* and *sufD* of *S. thermophilus* play a role in iron metabolism.

DISCUSSION

To investigate the oxidative stress defense system of *S. thermophilus*, a collection of random insertion mutants was screened to select clones with high sensitivity to menadione. Using this approach, we aimed to identify genes involved in superoxide stress defense without presuming their function. It can be noted that no gene encoding a protein considered a major antioxidant enzyme in other organisms was found by using this approach. Possible explanations for this outcome are (i) that the collection of mutants was nonexhaustive, (ii) that the corresponding mutants could have a weak phenotype, (iii) that such mutations could be lethal, and (iv) that most antiox-

idant enzymes, except those from SOD, are protective against H₂O₂ and not O₂^{•-}. The genes identified in the present work were classified according to their functions into five groups: genes potentially involved in cell wall metabolism, genes potentially involved in exopolysaccharide translocation, genes potentially involved in tRNA modification, genes potentially involved in iron metabolism, and genes of unknown function.

Perhaps the most striking observation of this study is that 5 out of the 10 sensitive mutants selected (*ossH*, *mreD*, *rodA*, *pbp2b*, and *cpsX*) are disrupted in a gene encoding a membrane protein. This finding could reflect that the cell envelope is the main target of the stress applied. Among these mutants, three (*mreD*, *rodA* and *pbp2b*) are impaired in genes potentially encoding proteins described as playing a role in peptidoglycan biosynthesis during cell elongation in *E. coli* (28, 52, 63). Phenotypic analysis of these three *S. thermophilus* mutants showed that they all present a reduced cell size associated with a modification of the cell shape, as observed in *rodA*, *pbp2*, and *mre* mutants of *E. coli* or *Salmonella enterica* serovar Typhimurium (5, 20, 28, 29, 63, 64). Moreover, the three *S. thermophilus* mutants also presented increased sensitivity to heat shock compared to the wild-type strain, suggesting that the protective role of MreD, RodA, and PBP2b is not specific to oxidative stress but could be enlarged to a more global stress response. In support of this hypothesis is the existence in *Lactococcus lactis* of mutations in *pbp* genes that negatively affect the acid tolerance of a *guaA* mutant (16).

The major function assigned to the cell wall is the maintenance of cell integrity through the protection of cells not only against osmotic stress but also against other stress conditions. For instance, it has already been shown that in LAB, compromised biosynthesis of peptidoglycan results in increased sensitivity to either UV (22) or acid (11) stress. Additionally, the disruption of *mreD*, *rodA*, or *pbp2b* may cause a modification of the peptidoglycan architecture in *S. thermophilus*, reducing cell protection against stresses such as oxidative and heat stresses.

tRNA modification is a universal feature of all living organisms, although some described modifications have no assigned function yet. Modifying enzymes affect the tRNA precursor either by chemical modification of an existing residue (in the base or in the sugar) or by nucleotide substitution. These modifications, particularly those present in the anticodon region, can affect translation efficiency and fidelity (for a review, see reference 10). In this study, two clones disrupted in the *tgt* and *ossF* loci, encoding proteins showing high identity to tRNA-modifying enzymes, were selected for their high sensitivity to oxidative stress. This sensitivity could suggest that the expression of genes involved in oxidative stress defense may require tRNA maturation or that such modifications could help tRNA stabilization under stress conditions. In agreement with the former hypothesis, *Shigella flexneri* pathogenicity is regulated by *virF* mRNA translation efficiency under TGT activity control (21). Additionally, several environmental conditions are suggested to influence tRNA modification. In the case of queuosine (whose synthesis and substitution are carried out by several enzymes, among which is TGT), synthesis is sensitive to the presence of iron (31) and TGT requires oxygen for its full activity (54). Therefore, further investigations on this topic should provide an interesting new regulatory model and could contribute to the emerging observation that trans-

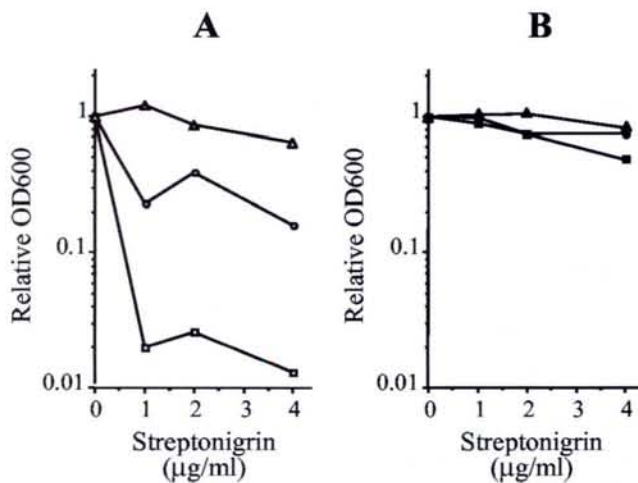


FIG. 5. Streptonigrin tolerance. The tolerance of *S. thermophilus* strains to streptonigrin was assessed by measuring the OD₆₀₀ (OD₆₀₀) after growth in TPPYC medium containing various concentrations of streptonigrin unsupplemented (A) or supplemented with 1 mM of 2,2'-dipyridyl (B) (see Material and Methods for details). The relative OD₆₀₀ was determined by comparing the OD₆₀₀s obtained with and without streptonigrin. ○ and ●, wild-type strain; □ and ■, *iscU₉₇* mutant; △ and ▲, *ΔsufD* mutant.

lational regulation plays an important role in oxidative stress response, a role ascribed mainly to transcriptional regulation thus far.

Iron is essential for cellular metabolism due to its requirement as a cofactor for a large number of enzymes. However, an excess of iron is toxic because of its involvement in the Fenton reaction, which enhances the formation of ROS. In this work, a gene homologous to *fatD*, encoding a putative Fe³⁺ ABC transporter, was found to be involved in the *S. thermophilus* oxidative stress response. To test whether *S. thermophilus* FatD is indeed an iron transporter, the sensitivity of the *fatD* mutant to streptonigrin (an iron-activated antibiotic) was assessed. No significant difference between the mutant and the wild-type strains was detected (data not shown), suggesting that the disruption of *fatD* would have no effect on intracellular iron concentration. Thus, either FatD is not a major iron transporter under the experimental procedure we used or *S. thermophilus fatD* encodes an ABC transporter for another metallic cation. A possible cation could be Mn²⁺, known to have a protective function against oxidative damage either via direct dismutase activity, such as that reported in other organisms (4, 26, 60), or via its role as a cofactor of the only SOD reported in *S. thermophilus* (19). Thus, the inactivation of an Mn²⁺ ABC transporter could trigger sensitivity to oxidative stress.

[Fe-S] clusters are oxygen-sensitive prosthetic groups incorporated into proteins via covalent bonds to cysteines. Iron-sulfur proteins, which are found in more than 120 classes of enzymes and proteins (30), are known to play important physiological roles in electron transfer, metabolic reactions, and transcriptional regulation (30, 32). Indeed, because of their intrinsic redox properties, they are likely to sense oxygen or its derivatives, as is the case for SoxR and FNR transcriptional regulators, responding to oxidative stress and oxygen availability (for a review, see reference 32). IscR is also an iron-sulfur transcriptional regulator that is suggested to respond to [Fe-S] cluster-limiting conditions and potentially to hydrogen peroxide in *E. coli* (32). Although it has been shown that [Fe-S] cluster assembly can occur spontaneously in vitro under anaerobic conditions (36), this process can also be catalyzed enzymatically in vivo. Genes of the *suf* operon and of the *isc* locus (iron-sulfur cluster) are conserved and encode [Fe-S] cluster biosynthesis and/or repair pathways (23). More precisely, it has recently been proposed that Isc proteins are necessary for biosynthesis, while Suf proteins are essentially involved in the repair of oxidatively damaged [Fe-S] clusters (37).

Study of Δ *sufD* and *iscU*₉₇ nonpolar mutants indicates that SufD and IscU proteins are required for growth in an iron-deficient medium. SufD and IscU inactivation conferred increased resistance and sensitivity, respectively, to streptonigrin. These results indicate that *S. thermophilus* SufD and IscU are involved in intracellular iron balance. Additionally, the SufD and IscU of *S. thermophilus*, similar to SufD and NifU family proteins from other microorganisms, are encoded by genes clustered with *sufB*, *sufC*, and *sufS* homologues, also potentially involved in [Fe-S] cluster repair. Hence, the participation of *S. thermophilus* SufD and IscU proteins in iron intracellular balance is probably the result of their involvement in [Fe-S] cluster repair, as in *Erwinia chrysanthemi* (37). The NifU family protein function is to provide a scaffold for the assembly of clusters mediated by NifS activity, and such clusters can then

be transferred into apo forms of [Fe-S] cluster-containing proteins (1, 70). The role of SufD in [Fe-S] cluster biosynthesis is not well established. However, in *S. thermophilus*, the increased resistance to streptonigrin of the Δ *sufD* mutant suggests that its intracellular iron concentration is lower than that of the wild type. At least two alternative hypotheses can be made: (i) SufD could be directly involved in the uptake or assimilation of extracellular iron, or (ii) SufD could inhibit the synthesis and/or repair of the [Fe-S] clusters. In the latter case, SufD inactivation would result in an increase in the synthesis of [Fe-S] clusters and, consequently, in a reduced intracellular free iron concentration.

Additionally, *S. thermophilus sufD* and *iscU* mutants displayed increased sensitivity to superoxide radicals compared to the wild type. Thus, SufD and IscU are likely to play a major role in oxidative stress defense, probably resulting from their predicted role in [Fe-S] cluster assembly and/or repair. The involvement of the *suf* operon in oxidative stress resistance during the early steps of infection by *Erwinia chrysanthemi* has also been reported (37), as has the induction of the *suf* gene cluster of *E. coli* by the hydrogen peroxide response regulator OxyR (72). In spite of a possible spontaneous formation, our data suggest that under oxidative stress, [Fe-S] cluster assembly and/or repair becomes a limiting step in the formation of iron-sulfur proteins and that the cell requires an additional, dedicated system, encoded by the *suf* gene cluster, to optimize [Fe-S] cluster formation or repair.

Conclusion. The strategy used in this work allowed the identification of genes involved in the *S. thermophilus* oxidative stress defense system. Data concerning the functions of some loci are already available in the literature, whereas the roles of other loci are not known. In both cases, the exact roles of these genes in the defense of *S. thermophilus* against oxidative stress remain to be elucidated.

ACKNOWLEDGMENTS

A.T. and A.F. were supported by grants from the Ministère de la Recherche. F.B. was supported by a grant from the Institut National de la Recherche Agronomique.

We are grateful to Paul Hoskinson for his advice regarding the English formulation of the manuscript.

REFERENCES

1. Agar, J. N., C. Krebs, J. Frazzon, B. H. Huynh, D. R. Dean, and M. K. Johnson. 2000. IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* **39**: 7856–7862.
2. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
3. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
4. Archibald, F. S., and I. Fridovich. 1981. Manganese, superoxide dismutase, and oxygen tolerance in some lactic acid bacteria. *J. Bacteriol.* **146**:928–936.
5. Begg, K. J., and W. D. Donachie. 1985. Cell shape and division in *Escherichia coli*: experiments with shape and division mutants. *J. Bacteriol.* **163**:615–622.
6. Benov, L. T., and I. Fridovich. 1994. *Escherichia coli* expresses a copper- and zinc-containing superoxide dismutase. *J. Biol. Chem.* **269**:25310–25314.
7. Berlett, B. S., and E. R. Stadtman. 1997. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**:20313–20316.
8. Biswas, I., A. Gruss, S. D. Ehrlich, and E. Maguin. 1993. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J. Bacteriol.* **175**:3628–3635.
9. Bjork, G. R. 1975. Transductional mapping of gene *tmA* responsible for the production of 5-methyluridine in transfer ribonucleic acid of *Escherichia coli*. *J. Bacteriol.* **124**:92–98.
10. Bjork, G. R., J. M. Durand, T. G. Hagervall, R. Leipuviene, H. K. Lundgren,

- K. Nilsson, P. Chen, Q. Qian, and J. Urbanavicius. 1999. Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett.* **452**:47–51.
11. Boyd, D. A., D. G. Cvitkovitch, A. S. Bleiweis, M. Y. Kiriukhin, D. V. DeBabov, F. C. Neuhaus, and I. R. Hamilton. 2000. Defects in α -anilypoliteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J. Bacteriol.* **182**:6055–6065.
 12. Bracquart, P. 1981. An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yogurt. *J. Appl. Bacteriol.* **51**:303–305.
 13. Briolat, V., and G. Reysset. 2002. Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *J. Bacteriol.* **184**:2333–2343.
 14. Broadbent, J., D. McMahon, C. Oberg, and D. Welker. 2001. Use of exopolysaccharide-producing cultures to improve the functionality of low fat cheese. *Int. Dairy J.* **11**:433–439.
 15. Broadbent, J., D. McMahon, D. Welker, C. Oberg, and S. Moineau. 2003. Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J. Dairy Sci.* **86**:407–423.
 16. Budin, A., V. Pichereau, E. Domakova, J. Tremblay, R. Dervyn, E. Auffray, S. D. Ehrlich, S. Kulakauskas, and E. Maguin. 2002. Effect of a *guaA* mutation on the acid tolerance of *L. lactis*. *Sci. Aliments* **22**:67–74.
 17. Burger, A., K. Sighler, G. Kelemen, M. Buttner, and W. Wohlleben. 2000. Identification and characterization of the *mre* gene region of *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet.* **263**:1053–1060.
 18. Chae, H. Z., K. Robison, L. B. Poole, G. Church, G. Storz, and S. G. Rhee. 1994. Cloning and sequencing of thiol-specific antioxidant from mammalian brain: alkyl hydroperoxide reductase and thiol-specific antioxidant define a large family of antioxidant enzymes. *Proc. Natl. Acad. Sci. USA* **91**:7017–7021.
 19. Chang, S. K., and H. M. Hassan. 1997. Characterization of superoxide dismutase in *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **63**:3732–3735.
 20. Costa, C. S., and D. N. Anton. 1993. Round-cell mutants of *Salmonella typhimurium* produced by transposition mutagenesis: lethality of *rodA* and *mre* mutations. *Mol. Gen. Genet.* **236**:387–394.
 21. Durand, J. M., B. Dagberg, B. E. Uhlin, and G. R. Bjork. 2000. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. *Mol. Microbiol.* **35**:924–935.
 22. Duwat, P., A. Cochu, S. D. Ehrlich, and A. Gruss. 1997. Characterization of *Lactococcus lactis* UV-sensitive mutants obtained by *ISS1* transposition. *J. Bacteriol.* **179**:4473–4479.
 23. Ellis, K. E., B. Clough, J. W. Saldanha, and R. J. Wilson. 2001. *Nifs* and *Sufs* in malaria. *Mol. Microbiol.* **41**:973–981.
 24. Faguy, D. M., and W. F. Doolittle. 2000. Horizontal transfer of catalase-peroxidase genes between archaea and pathogenic bacteria. *Trends Genet.* **16**:196–197.
 25. Farr, S. B., and T. Kogoma. 1991. Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* **55**:561–585.
 26. Fridovich, I. 1983. Superoxide radical: an endogenous toxicant. *Annu. Rev. Pharmacol. Toxicol.* **23**:239–257.
 27. Gibson, C. M., T. C. Mallett, A. Claiborne, and M. G. Caparon. 2000. Contribution of NADH oxidase to aerobic metabolism of *Streptococcus pyogenes*. *J. Bacteriol.* **182**:448–455.
 28. Ishino, F., W. Park, S. Tomioka, S. Tamaki, I. Takase, K. Kunugita, H. Matsuzawa, S. Asoh, T. Ohta, B. G. Spratt, et al. 1986. Peptidoglycan synthetic activities in membranes of *Escherichia coli* caused by overproduction of penicillin-binding protein 2 and RodA protein. *J. Biol. Chem.* **261**:7024–7031.
 29. Iwaya, M., R. Goldman, D. J. Tipper, B. Feingold, and J. L. Strominger. 1978. Morphology of an *Escherichia coli* mutant with a temperature-dependent round cell shape. *J. Bacteriol.* **136**:1143–1158.
 30. Johnson, M. K. 1998. Iron-sulfur proteins: new roles for old clusters. *Curr. Opin. Chem. Biol.* **2**:173–181.
 31. Kersten, H., and W. Kersten. 1990. Biosynthesis and function of queuine and queuosine tRNAs, p. B69–B108. In C. W. Gehrke and K. C. T. Kuo (ed.), *Chromatography and modification of nucleosides*. Elsevier, Amsterdam, The Netherlands.
 32. Kiley, P. J., and H. Beinert. 2003. The role of Fe-S proteins in sensing and regulation in bacteria. *Curr. Opin. Microbiol.* **6**:181–185.
 33. King, K. Y., J. A. Horenstein, and M. G. Caparon. 2000. Aerotolerance and peroxide resistance in peroxidase and PerR mutants of *Streptococcus pyogenes*. *J. Bacteriol.* **182**:5290–5299.
 34. Leenhouts, K. 1995. Integration strategies and vectors. *Dev. Biol. Stand.* **85**:523–530.
 35. Maguin, E., H. Prevost, S. D. Ehrlich, and A. Gruss. 1996. Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J. Bacteriol.* **178**:931–935.
 36. Merchant, S., and B. W. Dreyfuss. 1998. Metalloprotein and cytochrome assembly in chloroplasts. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**:25–51.
 37. Nachin, L., M. El Hassouni, L. Loiseau, D. Expert, and F. Barras. 2001. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.* **39**:960–972.
 38. Nishio, K., and M. Nakai. 2000. Transfer of iron-sulfur cluster from NifU to apoferredoxin. *J. Biol. Chem.* **275**:22615–22618.
 39. Noguchi, S., Y. Nishimura, Y. Hirota, and S. Nishimura. 1982. Isolation and characterization of an *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J. Biol. Chem.* **257**:6544–6550.
 40. Okada, N., S. Noguchi, H. Kasai, N. Shindo-Okada, T. Ohgi, T. Goto, and S. Nishimura. 1979. Novel mechanism of posttranscriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. *J. Biol. Chem.* **254**:3067–3073.
 41. O'Sullivan, T., and G. Fitzgerald. 1999. Electroporation of industrial strains of *Streptococcus thermophilus*. *J. Appl. Microbiol.* **86**:275–283.
 42. Patzer, S. I., and K. Hantke. 1999. SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in *Escherichia coli*. *J. Bacteriol.* **181**:3307–3309.
 43. Pébay, M., A. C. Holl, J. M. Simonet, and B. Decaris. 1995. Characterization of the *gor* gene of the lactic acid bacterium *Streptococcus thermophilus* CNRZ368. *Res. Microbiol.* **146**:371–383.
 44. Poole, L. B., M. Higuchi, M. Shimada, M. L. Calzi, and Y. Kamio. 2000. *Streptococcus mutans* H₂O₂-forming NADH oxidase is an alkyl hydroperoxide reductase protein. *Free Radic. Biol. Med.* **28**:108–120.
 45. Poyart, C., E. Pellegrini, O. Gaillot, C. Boumaila, M. Baptista, and P. Trieu-Cuot. 2001. Contribution of Mn-cofactored superoxide dismutase (SodA) to the virulence of *Streptococcus agalactiae*. *Infect. Immun.* **69**:5098–5106.
 46. Rocha, E. R., G. Owens, Jr., and C. J. Smith. 2000. The redox-sensitive transcriptional activator OxyR regulates the peroxide response regulon in the obligate anaerobe *Bacteroides fragilis*. *J. Bacteriol.* **182**:5059–5069.
 47. Rocha, E. R., T. Selby, J. P. Coleman, and C. J. Smith. 1996. Oxidative stress response in an anaerobe, *Bacteroides fragilis*: a role for catalase in protection against hydrogen peroxide. *J. Bacteriol.* **178**:6895–6903.
 48. Rocha, E. R., and C. J. Smith. 1998. Characterization of a peroxide-resistant mutant of the anaerobic bacterium *Bacteroides fragilis*. *J. Bacteriol.* **180**:5906–5912.
 49. Rocha, E. R., and C. J. Smith. 1999. Role of the alkyl hydroperoxide reductase (*ahpCF*) gene in oxidative stress defense of the obligate anaerobe *Bacteroides fragilis*. *J. Bacteriol.* **181**:5701–5710.
 50. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 51. Sonnhammer, E. L. L., G. Von Heijne, and A. Krogh. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, p. 175–182. In J. Glasgow, T. Littlejohn, F. Major, L. R. D. Sankoff, and C. Sensen (ed.), *Proceedings of the 6th International Conference on Intelligent Systems for Molecular Microbiology*. AAAI Press, Menlo Park, Calif.
 52. Spratt, B. G. 1975. Distinct penicillin binding proteins involved in the division, elongation, and shape of *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* **72**:2999–3003.
 53. Stadtman, E. R. 1993. Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. *Annu. Rev. Biochem.* **62**:797–821.
 54. Szabo, L., S. Nishimura, and W. R. Farkas. 1988. Possible involvement of queuine in oxidative metabolism. *Biofactors* **1**:241–244.
 55. Terzaghi, B. E., and W. E. Sandine. 1975. Improved medium for lactic streptococci and their bacteriophages. *Appl. Environ. Microbiol.* **29**:807–813.
 56. Thibessard, A., A. Fernandez, B. Gintz, B. Decaris, and N. Leblond-Bourget. 2002. Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can. J. Microbiol.* **48**:473–478.
 57. Thibessard, A., A. Fernandez, B. Gintz, N. Leblond-Bourget, and B. Decaris. 2002. Effects of *rodA* and *pbp2b* disruption on cell morphology and oxidative stress response of *Streptococcus thermophilus* CNRZ368. *J. Bacteriol.* **184**:2821–2826.
 58. Thibessard, A., A. Fernandez, B. Gintz, N. Leblond-Bourget, and B. Decaris. 2001. Hydrogen peroxide effects on *Streptococcus thermophilus* CNRZ368 cell viability. *Res. Microbiol.* **152**:593–596.
 59. Thibessard, A., N. Leblond-Bourget, A. Fernandez, B. Gintz, and B. Decaris. 2001. Response of *Streptococcus thermophilus* CNRZ368 and its colonial variants to oxidative stress: evidence for an inducible defence system. *Lait* **81**:311–316.
 60. Tseng, H., Y. Srihanta, A. McEwan, and M. Jennings. 2001. Accumulation of manganese in *Neisseria gonorrhoeae* correlates with resistance to oxidative killing by superoxide anion and is independent of superoxide dismutase activity. *Mol. Microbiol.* **40**:1175–1186.
 61. Tusnady, G. E., and I. Simon. 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **283**:489–506.
 62. Visca, P., A. Ciervo, V. Sanfilippo, and N. Orsi. 1993. Iron-regulated salicylate synthesis by *Pseudomonas* spp. *J. Gen. Microbiol.* **139**:1995–2001.
 63. Wachi, M., M. Doi, Y. Okada, and M. Matsushashi. 1989. New *mre* genes

- mreC* and *mreD*, responsible for formation of the rod shape of *Escherichia coli* cells. *J. Bacteriol.* **171**:6511–6516.
64. **Wachi, M., M. Doi, S. Tamaki, W. Park, S. Nakajima-Iijima, and M. Matsubashi.** 1987. Mutant isolation and molecular cloning of *mre* genes, which determine cell shape, sensitivity to mecillinam, and amount of penicillin-binding proteins in *Escherichia coli*. *J. Bacteriol.* **169**:4935–4940.
65. **Wu, J., and B. Weiss.** 1992. Two-stage induction of the *soxRS* (superoxide response) regulon of *Escherichia coli*. *J. Bacteriol.* **174**:3915–3920.
66. **Yamamoto, Y., M. Higuchi, L. B. Poole, and Y. Kamio.** 2000. Role of the *dpr* product in oxygen tolerance in *Streptococcus mutans*. *J. Bacteriol.* **182**:3740–3747.
67. **Yesilkaya, H., A. Kadioglu, N. Gingles, J. E. Alexander, T. J. Mitchell, and P. W. Andrew.** 2000. Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **68**:2819–2826.
68. **Youn, H. D., H. Youn, J. W. Lee, Y. I. Yim, J. K. Lee, Y. C. Hah, and S. O. Kang.** 1996. Unique isozymes of superoxide dismutase in *Streptomyces griseus*. *Arch. Biochem. Biophys.* **334**:341–348.
69. **Yu, J., A. P. Bryant, A. Marra, M. A. Lonetto, K. A. Ingraham, A. F. Chalker, D. J. Holmes, D. Holden, M. Rosenberg, and D. McDevitt.** 2001. Characterization of the *Streptococcus pneumoniae* NADH oxidase that is required for infection. *Microbiology* **147**:431–438.
70. **Yuvaniyama, P., J. N. Agar, V. L. Cash, M. K. Johnson, and D. R. Dean.** 2000. NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. *Proc. Natl. Acad. Sci. USA* **97**:599–604.
71. **Zheng, L., R. H. White, V. L. Cash, R. F. Jack, and D. R. Dean.** 1993. Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc. Natl. Acad. Sci. USA* **90**:2754–2758.
72. **Zheng, M., X. Wang, L. J. Templeton, D. R. Smulski, R. A. LaRossa, and G. Storz.** 2001. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.* **183**:4562–4570.

cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein Involved in Cellular Segregation in *Streptococcus thermophilus*

Frédéric Borges,¹ Séverine Layec,¹ Annabelle Thibessard,¹ Annabelle Fernandez,¹ Brigitte Gintz,¹ Pascal Hols,² Bernard Decaris,¹ and Nathalie Leblond-Bourget^{1*}

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences et Techniques de l'Université Henri Poincaré Nancy 1, Vandoeuvre-lès-Nancy, France,¹ and Unité de Génétique, Institut des Sciences de la Vie, Université catholique de Louvain, Louvain-La-Neuve, Belgium²

Received 2 September 2004/Accepted 4 January 2005

The isolation of a *Streptococcus thermophilus* CNRZ368 mutant displaying a long-chain phenotype allowed us to identify the *cse* gene (for cellular segregation). The N terminus of Cse exhibits high similarity to *Streptococcus agalactiae* surface immunogenic protein (SIP), while its C terminus exhibits high similarity to *S. thermophilus* PcsB. In CNRZ368, deletion of the entire *cse* open reading frame leads to drastic lengthening of cell chains and altered colony morphology. Complementation of the Δcse mutation with a wild-type allele restored both wild-type phenotypes. The central part of Cse is a repeat-rich region with low sequence complexity. Comparison of *cse* from CNRZ368 and LMG18311 strains reveals high variability of this repeat-rich region. To assess the impact of this central region variability, the central region of LMG18311 *cse* was exchanged with that of CNRZ368 *cse*. This replacement did not affect chain length, showing that divergence of the central part does not modify cell segregation activity of Cse. The structure of the *cse* locus suggests that the chimeric organization of *cse* results from insertion of a duplicated sequence deriving from the *pcsB* 3' end into an ancestral *sip* gene. Thus, the *cse* locus illustrates the module-shuffling mechanism of bacterial gene evolution.

Among gram-positive cocci, some species grow as single cells and others grow as grouped cells, linked together into various arrangements. In particular, cocci from several genera, such as *Streptococcus*, *Lactococcus*, *Peptostreptococcus*, *Ruminococcus*, and *Coprococcus*, grow as cell chains (48). The number of cells per chain varies considerably among species and strains (14). Indeed, chain length can range from two cells, as in *Streptococcus pneumoniae*, to around 100 cells in *Streptococcus mitior* (14). Few genes have been implicated in short-chain phenotypes, since their inactivation leads to an increased number of cells per chain. One of the most striking phenotypes is that of an *S. pneumoniae* *lytA lytB* double mutant, whose chain length can reach more than 100 cells, while the wild type typically grows as pairs (12). The shortening of cell chains was ascribed to the cell wall hydrolytic activity of LytA and LytB proteins in *S. pneumoniae* (25) and of AcmA in *Lactococcus lactis* (6). The increase of cell number per chain could be accompanied by obvious changes in cellular morphology. Thus, *rodA*, *pbp2*, and *mreD* mutants from *Streptococcus thermophilus* display not only long cell chains but also rod-shaped cells instead of ovoid wild-type cells (50, 52, 54). Also, an *S. pneumoniae* mutant depleted for expression of the essential *pcsB* gene forms long chains with irregularly shaped cells (33). Although the biological function of chains is unknown, it was recently demonstrated that the adhesion ability is reduced by cell chain lengthening in *L. lactis* (30).

In this study, we focused on a gene called *cse* for its role in

cellular segregation, identified in *S. thermophilus*. This bacterium is a lactic acid bacterium used as a starter of fermentation for the conversion of milk into yogurt and many cheeses (e.g., Emmental, Gruyère, mozzarella, and cheddar). We describe the chimeric structure of *cse* and its central region, which is repeat rich and exhibits intraspecies sequence variability. The construction of several mutants showed that this chimeric and variable gene encodes a functional protein involved in cellular segregation and colony morphology and allowed assessment of the impact of the central part variability on cell segregation activity.

MATERIALS AND METHODS

Bacterial strains, growth conditions, and plasmids. All strains and plasmids used in this work are presented in Table 1. Depending on the experiments, *S. thermophilus* and its derivatives were cultivated in milk medium, TTPY (tryptose, 7 g · liter⁻¹; Proteose peptone, 7 g · liter⁻¹; yeast extract, 2 g · liter⁻¹; lactose, 20 g · liter⁻¹) (5), or M17 (51) medium without shaking. Milk medium was used for strain storage, M17 was used for mutant generation, and TTPY was employed for phenotypic analysis. Phenotypic analyses were performed at 42°C, the optimal growth temperature of *S. thermophilus*. Erythromycin was added at 2 $\mu\text{g} \cdot \text{ml}^{-1}$ when required. *S. thermophilus* derivative strains containing pGh9 (27) or pNST260+ plasmid derivatives were cultivated at 30°C when plasmid self-maintenance was required and at 42°C for selection of clones with the chromosome's integrated plasmid. *S. thermophilus* strains containing pFUN or its derivatives were cultivated at 30°C. *L. lactis* strains were grown at 30°C in M17 medium (51) supplemented with 0.5% glucose. Erythromycin was added at 5 $\mu\text{g} \cdot \text{ml}^{-1}$ when required.

Recombinant plasmids derived from pGh9 were transformed into *Escherichia coli* EC101, a TG1 strain containing a chromosomal copy of the pWV01 *repA* gene (6), and selected at 37°C on Luria-Bertani (LB) (47) containing 150 μg of erythromycin ml⁻¹. Recombinant plasmids derived from pFUN, which was kindly provided by I. Poquet from INRA (Jouy en Josas, France) (38) were transformed into *E. coli* DH5 α and selected on LB medium containing 100 μg of ampicillin ml⁻¹.

DNA manipulations. Preparation of chromosomal and plasmid DNA and Southern analysis were performed according to standard protocols (47). Se-

* Corresponding author. Mailing address: Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences et Techniques de l'Université Henri Poincaré Nancy 1, BP 239, 54506 Vandoeuvre-lès-Nancy, France. Phone: 33 3 83 68 42 10. Fax: 33 3 83 68 44 99. E-mail: bourget@nancy.inra.fr.

TABLE 1. Strains and plasmids used in this work

Strain or plasmid	Relevant phenotype(s) or genotype(s)	Source or reference
Strains		
<i>S. thermophilus</i>		
CNRZ368	Wild-type strain	INRA-CNRZ, strain collection
16D10	Em ^r <i>S. thermophilus</i> CNRZ368 containing integrated pGh9:ISS1 into <i>cse</i>	This work
CNRZ368-Δ <i>cse</i>	<i>S. thermophilus</i> CNRZ368 with Δ <i>cse</i> mutation	This work
CNRZ368-Δ <i>cse</i> -T	Em ^r , CNRZ368-Δ <i>cse</i> with pNST260+ integrated into the chromosome	This work
CNRZ368-Δ <i>cse</i> -C	Em ^r , CNRZ368-Δ <i>cse</i> with pNST260+:: <i>cse</i> integrated into the chromosome	This work
LMG18311	Wild-type strain	BCCM/LMG, strain collection
LMG18311-Δ <i>cse</i>	<i>S. thermophilus</i> LMG18311 with Δ <i>cse</i> mutation	This work
LMG18311- <i>cse</i> _{varCNRZ368}	<i>S. thermophilus</i> LMG18311 with the chimeric allele <i>cse</i> _{LMG18311varCNRZ368}	This work
<i>L. lactis</i> subsp. <i>cremoris</i>	Derivative of MG1363 wild-type strain containing 701-bp SacI-SpeI chromosomal deletion in <i>acmA</i> gene	6
MG1363 <i>acmA</i> Δ1		
<i>E. coli</i>		
EC101	<i>supE hsd-5 thi</i> Δ(<i>lac-proAB</i>) F' (<i>traD6 proAB⁺ lac^f lacZΔM15</i>) <i>repA⁺</i> , derivative of TG1 strain (47)	23
DH5α	<i>supE44 ΔlacU169</i> (φ80 <i>lacZ</i> ΔM15) <i>hsdR17 endA1 gyrA96 thi-1 relA1</i>	47
BL21 (DE3)	<i>hsdS gal</i> (λ <i>cl</i> s857 <i>ind1</i> Sam7 <i>nin5 lacUV5-T7</i> gene 1)	47
Plasmids		
pGh9	Em ^r , thermosensitive replication origin from pVE6002	27
pGh9:ISS1	pGh9 with ISS1	28
pGh9::Δ <i>cse</i> _{CNRZ368}	pGh9 with Δ <i>cse</i> and CNRZ368 surrounding regions	This work
pGh9::Δ <i>cse</i> _{LMG18311}	pGh9 with Δ <i>cse</i> and LMG18311 surrounding regions	This work
pGh9:: <i>cse</i> _{LMG18311varCNRZ368}	pGh9 with chimeric allele <i>cse</i> _{LMG18311varCNRZ368}	This work
pNST260+	pGh9 with <i>int</i> and <i>attI</i> from ICESiI	G. Guédon, personal communication
pNST260+:: <i>cse</i>	pNST260+ with CNRZ368 <i>cse</i> ORF, its putative promoter and terminator	This work
pFUN	Em ^r and Amp ^r , derivative of pIL252, with Δ <i>Spnuc</i> gene	38
pFUN:: <i>cse</i>	pFUN with <i>cse</i> -Δ <i>Spnuc</i> translational fusion	This work
pET15b	Protein expression vector	Novagen
pET15b:: <i>cse</i>	pET15b with hexa-His-Cse translational fusion	This work
pET15b:: <i>cse</i> ₁₋₁₈₃	pET15b with hexa-His-Cse ₁₋₁₈₃ translational fusion	This work
pET15b:: <i>cse</i> ₁₈₄₋₄₆₁	pET15b with hexa-His-Cse ₁₈₄₋₄₆₁ translational fusion	This work

quencing was performed by using dye terminator chemistry on an ABI Prism 377 genetic analyzer (PE Biosystems). Sequence data were analyzed with BLAST (1, 2), SignalP (35), Mfold (59), Dot plot (<http://arbl.cvmbs.colostate.edu/molkit/dnadot/>), SEG (56), and PSIPRED (29) software. GenBank, the Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), and the Codon Usage Database (<http://www.kazusa.or.jp/codon/>) (32) were consulted. Sequence data of *S. thermophilus* LMG18311 strain were obtained from the UCL Life Sciences Institute website at <http://www.biol.ucl.ac.be/genome/>. Sequencing of *S. thermophilus* LMG18311 was supported by the Walloon region (BIOVAL grant no. 9813866).

Nuclease assay on agar plates. Extracellular production of nuclease activity by *E. coli* and *S. thermophilus* were detected by the metachromatic agar diffusion method, as previously described (21, 24). Briefly, following bacterial growth on solid media, plates were overlaid with toluidine blue-DNA agar and incubated at 37°C for 2 h. The presence of a pink halo around colonies indicates extracellular nuclease activity. *E. coli* strains were not incubated for more than 2 h, since false positives could appear after this time (38).

Recombinant DNA and mutant construction. Oligonucleotides used in this study were purchased from MWG Biotech AG (Ebersburg, Germany) and are listed Table 2.

(i) **Translational fusion with Δ*Spnuc*.** To obtain the plasmid pFUN::*cse*, the region covering the putative promoter region and the *cse* open reading frame (ORF) except its stop codon, was amplified by PCR and introduced into pFUN, leading to *cse*-Δ*Spnuc* fusion. For pFUN::Δ*Spnuc*, the regions flanking the sequence encoding the putative signal peptide (the 81 bp following the ATG putative initiation site of translation) were amplified by PCR and introduced into pFUN. This led to a Δ*Spnuc*-*cse*-Δ*Spnuc* fusion where the sequence encoding the putative signal peptide of Cse was replaced by a HindIII restriction site. Following selection in *E. coli* DH5α, both constructs were introduced into *S. thermophilus* by electroporation.

(ii) **Deletion of *cse* in *S. thermophilus* CNRZ368 and LMG18311 strains.**

In-frame *cse* deletion mutants were constructed as previously described (52). Briefly, the two regions flanking the locus to be deleted were independently amplified by PCR, digested by appropriate restriction enzymes, and joined by ligation together and with pGh9. After introduction of the recombinant plasmid into *S. thermophilus*, two crossovers, upstream and downstream of the deleted region, were selected (52). Thus, only the first and the last three codons of the *cse*

TABLE 2. Oligonucleotides used in this work

Oligonucleotide	Sequence ^a (5'-3')
delcse.E5'	CCCCCCCCAAGCTTACAAGATCAGTTTGGAAAC
delcse.I5'	CCCCCGAATTCGTGATAACATGTATAAATTCAT
delcse.I3'	CCCCCGAATTCATTTATCCATAATGAGTACA
delcse.E3'	CCCCCCTGCAGCAGCCTGTGGTTGAAG
fcompl	TTGTTGGCGCGCTTATAGCCGATTTGGCTTTTGAAG
rcompl	TTGTTGGCGCGCTTATAGCCGATTTGGCTTTTGAAG
racse.E5'	CCCCCAAGCTTGAATGTTTGGCTAATATC
racse.I5'	CCCCCGATATCAATAGCGTCTGCAACT
racse.I3'	CCCCCGATATCAACTCTAATACATATC
racse.E3'	CCCCCTGCAGTTATGGATAAATAATATATC
ravar.5'	CATGTTACTTCAGCAACAAC
ravar.3'	GCATATTCGTATGTAGCTGC
fnuc.cse.5'	CCCCCGAATTCGTATAGGGATAGGGCTT
fnuc.cse.15'	CCCCCAAGCTTCATGTATAATTCATTCCTT
fnuc.cse.13'	CCCCCAAGCTTGATGAACTTCTCACTG
fnuc.cse.3'	CCCCCGGATCCCTGGATAAATAATATACAG
sex.E5'	CCCCCCCCATGTTATCAAAATCTAAAC
sex.I5'	CCCCGGATCCTTAAGCTGGAGTATCTGATGC
sex.I3'	CCCCCCCCATGTCACGCAAGACTGAAACAA
sex.E3'	CCCCGGATCCTTATGGATAAATAATATATC

^a Restriction enzyme sites are underlined.

ORF were kept, and the remainder of the ORF was replaced by an EcoRI restriction site. The *cse* deletion was checked by PCR and Southern hybridization (data not shown).

(iii) ***Δcse* complementation.** Complementation of the *Δcse* mutation was carried out by inserting the wild-type allele in *trans*. For this purpose, the pNST260+ plasmid (G. Guédon, personal communication), carrying a gene encoding the ICES_{St1} integrase and its *attI* attachment site, was used (7). This plasmid can integrate into the *S. thermophilus* CNRZ368 chromosome at the *attR* site of ICES_{St1} (G. Guédon, personal communication). The entire *cse* ORF with its putative promoter and terminator was amplified by PCR and ligated into pNST260+. Following selection in *E. coli* EC101, the construct was introduced into *S. thermophilus* by electroporation. As a control, *S. thermophilus* was also transformed with empty pNST260+. Integration of pNST260+ and pNST260+::cse into the chromosome was selected at 42°C, the replication-restrictive temperature, in the presence of erythromycin. Plasmid integration into the chromosomal attachment site was checked by Southern hybridization (data not shown).

(iv) **Allelic replacement of the *var-cse* region.** The approach for *var-cse* replacement involved the wild-type sequence containing GAT and ATC sequences localized, respectively, upstream and downstream of *var-cse*. These sequences allow an EcoRV restriction site formation if they are assembled. The plasmid used for replacement of the *var-cse* region of LMG18311 by that of CNRZ368 was constructed in a two-step cloning procedure. The *var-cse* flanking regions from the LMG18311 strain were amplified by PCR. After digestion by appropriate restriction enzymes, PCR fragments were ligated together, generating an EcoRV restriction site, and inserted into pGh9. This EcoRV restriction site was generated by assembling the GAT and ATC sequences that flank the wild-type *var-cse* sequence. The plasmid carrying the two PCR fragments was selected in *E. coli* EC101. Then the region localized between the GAT and ATC sequences including the *var-cse* region from CNRZ368 (as defined in Fig. 2D) was amplified by PCR. The PCR product was next treated at 72°C for 30 min with *Pfu* polymerase for generation of blunt ends. This last step removes the 3' adenosine overhang added during the PCR. The polished fragment was inserted into the plasmid carrying the *var-cse* flanking regions previously digested by EcoRV to generate blunt ends. The resulting plasmid was selected in *E. coli* EC101. The correct insert orientation was checked by sequencing. Then the plasmid was introduced into *S. thermophilus* LMG18311 by electroporation. The transformed strain was checked by Southern hybridization with a specific probe for LMG18311 genomic DNA (37). Two crossover events, surrounding the *var-cse* region, were selected to generate the LMG18311-*cse*_{varCNRZ368} mutant.

(v) **Plasmid constructs for protein overproduction.** The whole *cse* ORF and the regions encoding Cse parts from amino acid 1 to 183 and from amino acid 184 to 461 were amplified by PCR. The PCR products were independently introduced into the pET15b plasmid (Novagen), ending in an in-frame fusion downstream of the hexa-His encoding sequence. After selection in *E. coli* DH5 α , the constructs were introduced into *E. coli* BL21(DE3) for protein overproduction.

Microscopy. Colonies and cells were observed with a Nikon OPTIPHOT microscope mounted with phase contrast equipment (Ph). Colony examination was done at a magnification of $\times 100$. Cells were observed at a magnification of $\times 100$, with the condenser turret at position Ph4, or at a magnification of $\times 1,000$ by phase contrast.

Protein overproduction. *E. coli* BL21(DE3) strains transformed with pET15b and derivatives were grown in LB medium at 37°C supplemented with 50 μ g of ampicillin ml⁻¹. At an optical density at 600 nm (OD₆₀₀) of 0.6, isopropyl- β -D-thiogalactopyranoside (IPTG) was added at a final concentration of 100 μ M during 4 h to induce expression of N-terminally hexa-His-tagged proteins. To extract proteins, cell pellets resuspended in sample electrophoresis buffer were heated for 5 min at 95°C. For cell lysis testing in nondenaturing conditions, proteins were extracted by sonication.

SDS-PAGE and renaturing SDS-PAGE. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as previously described (22) with 10% (wt/vol) polyacrylamide. Lytic activity was tested by SDS-PAGE, performed with gels containing 10% polyacrylamide (wt/vol) and 0.2% (wt/vol) *S. thermophilus* cells or *Micrococcus lysodeikticus* cells, followed by a protein renaturation step. The *S. thermophilus* cells used as substrates for lytic activity were prepared as previously described (39) with the following modification: after lyophilization, cells were resuspended in distilled water at a 10% final concentration and autoclaved for 20 min at 120°C. *S. thermophilus* proteins were extracted by glass bead disruption of cells as previously described (17). After electrophoresis, gels were gently shaken in 100 ml of water at 4°C for 1 h. Then water was replaced by 100 ml of 20 mM Tris-HCl (pH 7) containing 1% (vol/vol)

Triton X-100 for overnight incubation at 42°C. Bands of lytic activity were visualized as previously described (6).

Nucleotide sequence accession numbers. DNA sequences reported in this paper have been deposited in GenBank under accession numbers AY695844, AY730642, and AY730643.

RESULTS

***cse* involvement in colony morphology and cell segregation.** The 16D10 mutant, displaying a long-chain phenotype in addition to a variant colony morphology, was selected from an *S. thermophilus* CNRZ368 mutant collection. This collection was previously obtained by insertional mutagenesis (53) with the pGh9:ISS1 plasmid (28). The locus disrupted by pGh9:ISS1 in the 16D10 mutant was identified by cloning and sequencing the 2,697 bp flanking the disruption site. Sequence analysis revealed that pGh9:ISS1 disrupted an ORF of 1,386 bp, named *cse* for cell segregation. The CNRZ368 *cse* ORF encodes a putative 462-amino-acid protein of unknown function with a deduced molecular mass of approximately 48.5 kDa. The first 28 amino acids display Sec-dependent signal peptide characteristics as predicted by signalP software (35), suggesting that Cse is exported. The Cse central part, containing repeats (see below), is highly enriched in alanine, glutamate, threonine, and serine (20.4, 32.9, 15.3, and 8.3%, respectively) residues compared to the mean content of *S. thermophilus* proteins (6.1, 5.9, 5.7, and 4.6%, respectively), according to the Codon Usage Database (32). This central region is potentially hydrophilic (due to its high glutamate, serine, and threonine content) and acidic (due to its glutamate enrichment).

To confirm the involvement of *cse* in these phenotypes, a *cse* null mutant was constructed by in-frame deletion of the *cse* ORF, resulting in a CNRZ368-*Δcse* strain. Colonies of this mutant were flatter, larger, and less opaque than wild-type ones (Fig. 1A). Phase-contrast photonic microscopy observations revealed that CNRZ368-*Δcse* chains were much longer than wild-type ones (Fig. 1B). This was reinforced by counting the number of cells per chain for each strain. Only 15% of wild-type cells formed chains containing more than 100 cells in stationary phase, whereas 100% of CNRZ368-*Δcse* cells did (Fig. 1C). During exponential growth (OD₆₀₀ = 0.2), 76.5% of the wild-type cells counted formed chains containing less than 40 cells and none of them formed chains containing more than 80 cells. On the contrary, 87.5% of the CNRZ368-*Δcse* cells formed chains containing more than 100 cells. These results show that the long-chain phenotype of the CNRZ368-*Δcse* mutant is not growth phase dependent. No major modification of the cell shape or size was observed in the CNRZ368-*Δcse* mutant. Moreover, mutant and wild-type strain growth curves showed no significant difference in either doubling time or in stationary phase entry OD₆₀₀ value (data not shown).

Complementation was performed by using the pNST260+ integrative plasmid (G. Guédon, personal communication). This plasmid can integrate into the *S. thermophilus* CNRZ368 chromosome in the *attR* site of ICES_{St1} (G. Guédon, personal communication), a resident integrative element (7). The *Δcse* mutation was complemented in *trans* by chromosomal integration of the *cse* wild-type gene carried by the pNST260+::cse plasmid, resulting in the CNRZ368-*Δcse*-C strain. The negative control consisted of the chromosomal insertion of empty pNST260+, resulting in the CNRZ368-*Δcse*-T strain. The

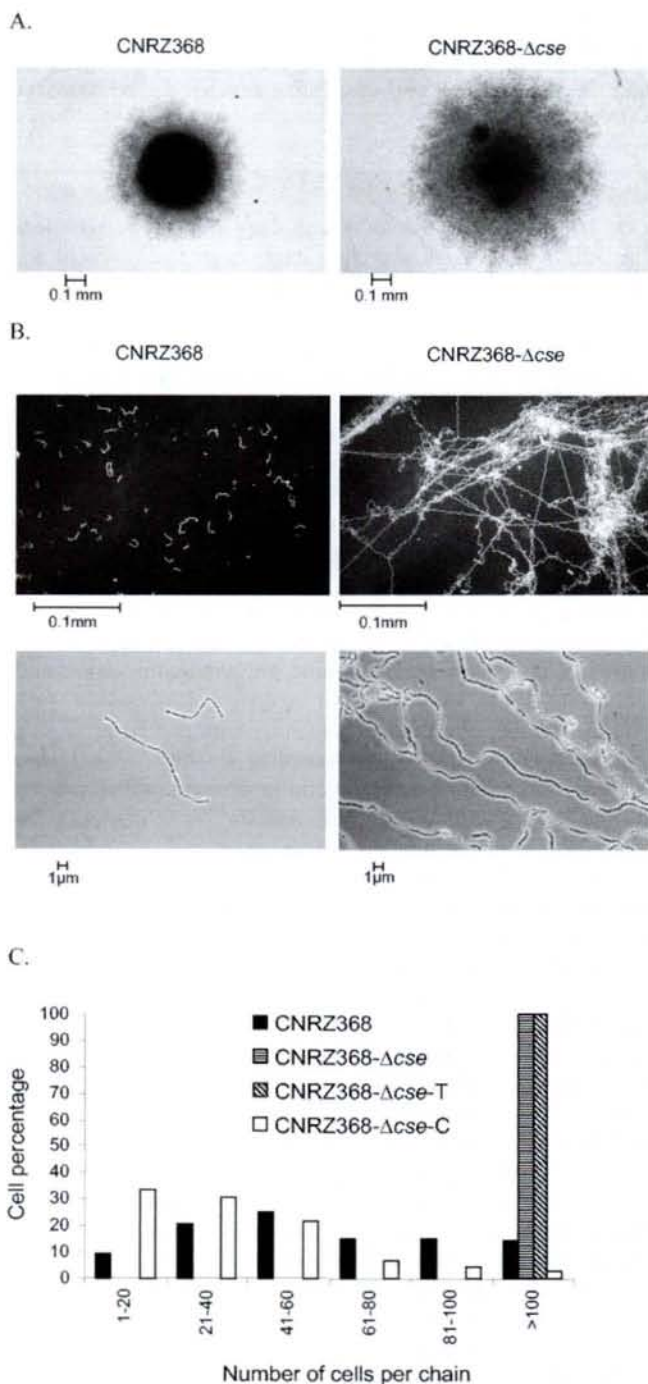


FIG. 1. Morphological consequences of *cse* deletion. (A) Colony morphology pictured after 20 h of growth; (B) cell chain morphology of *S. thermophilus* CNRZ368 strains; (C) results from counting number of cells per chain. A total of 1,000 cells were counted from three independent cultures. Cells were photographed and counted in stationary phase after 20 h of growth. After this time of growth, OD_{600} values reached 0.66 ± 0.074 , 0.56 ± 0.038 , 0.50 ± 0.029 , and 0.62 ± 0.022 , respectively, for strains CNRZ368, CNRZ368- Δcse , CNRZ368- Δcse -T, and CNRZ368- Δcse -C.

complemented strain exhibited a wild-type phenotype with respect to colony morphology and number of cells per chain, while the negative control maintained a mutant phenotype (Fig. 1C and colony morphology data not shown). These re-

sults demonstrate that *cse* is involved in colony morphology and cell segregation.

Central part of *cse* is repeat rich and variable. Alignment of the *cse* nucleotide sequence from strains CNRZ368 and LMG18311 (*cse*_{CNRZ368} and *cse*_{LMG18311}) showed that the first 600 bp and the last 346 bp of these ORFs display 94 and 96% identity, respectively (Fig. 2A). In contrast, both *cse* allele central parts, consisting of 443 bp in *cse*_{CNRZ368} and 503 bp in *cse*_{LMG18311}, show only 61% identity. Because of its variability, this region was called *var-cse*. Both *var-cse* regions show a high level of redundancy consisting of short direct repeats, with tandem or dispersed arrangement (Fig. 2B). As a consequence of its high content in short repeats, it is a low-complexity region as detected by the SEG program. Nine different repeat units (named A to I) that varied in size and sequence were distinguished (Fig. 2C). All of the repeated units have a size divisible by three and are in frame. The direct consequence is that all nucleic acid repeat sequences correspond to amino acid repeats in the *cse* product (Fig. 2B). The comparison of *var-cse* from *S. thermophilus* CNRZ368 and LMG18311 showed that *var-cse*_{LMG18311} is 60 bp longer than *var-cse*_{CNRZ368}. Additionally, the repeat content of both *var-cse* alleles is qualitatively different (Fig. 2D). In conclusion, the central part of *cse* is almost exclusively built with repeated sequences and displays a high degree of intraspecies variability.

Secondary structures were predicted for Var-Cse by using the PSIPRED method (19). Almost the entire Var-Cse_{CNRZ368} (93%) region and the whole Var-Cse_{LMG18311} region are predicted to adopt coil structures, suggesting that this region is a nonglobular domain. However, the confidence value of the prediction is low (3.56 ± 1.79 for Var-Cse_{CNRZ368} and 3.45 ± 1.83 for Var-Cse_{LMG18311}, on a 0 to 10 scale). The predicted nonglobular structure of this region could explain the tolerance for its high genetic variability.

***cse* is chimeric.** The nucleotide sequence of *cse* and its flanking regions in *S. thermophilus* LMG18311 was used in a BLASTN search against the genome of the same strain. Results revealed a region of 619 bp, named HRC1 for homologous region, copy 1, sharing 93% identity with another region of 618 bp, localized elsewhere in the genome and named HRC2 for homologous region, copy 2 (Fig. 3A). HRC1 includes the last 350 bp of *cse*, the following intergenic region, and the first 58 bp of the downstream *orf1*. HRC2 includes the last 350 bp of *pcsB*, the following intergenic region, and the first 58 bp of the downstream *rppk* ORF (encoding a putative ribose-phosphate pyrophosphokinase) (Fig. 3A). Thus, the C-terminal part of Cse is homologous to the C-terminal part of PcsB from *S. thermophilus* LMG18311 (Fig. 3A). BLAST searches within nonredundant databases showed that a PcsB orthologue is present in the *Streptococcus* and *Lactococcus* genera, whereas no protein with significant similarity to the whole Cse protein was found in these organisms. PcsB is essential in the maintenance of cell shape in *Streptococcus* species (9, 33, 34, 42, 43). Searches of the Conserved Domain Database show that both PcsB and Cse possess a region homologous to the CHAP domain (cysteine histidine-dependent aminohydrolase/peptidase) (Fig. 3A) which possesses a glutathionylspermidine amidase activity in *E. coli* (4). The CHAP domain is widely distributed in extracellular proteins, where it is supposed to be involved in peptidoglycan hydrolysis (3, 45).

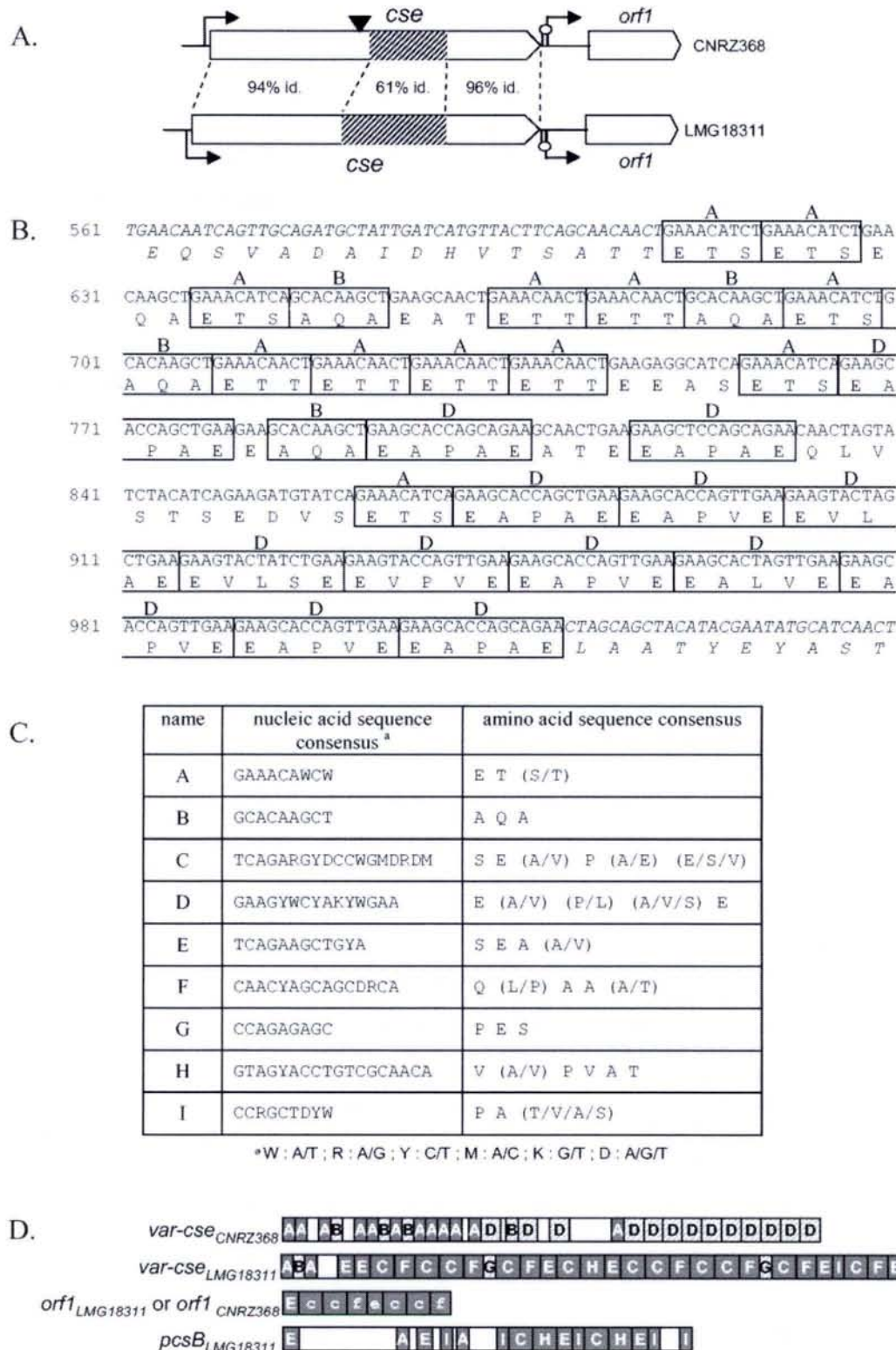


FIG. 2. *var-cse* variability and repeat content of *cse*, *pcsB*, and *orf1*. (A) Schematic representation of the *cse* locus from *S. thermophilus* strains CNRZ368 and LMG18311. Open arrows represent ORFs and indicate their reading direction, broken arrows indicate putative promoters, and hairpin loops symbolize putative rho-independent terminators. The black arrowhead indicates the insertion site of pGh9:ISS1 within the genome of the 16D10 mutant. The percentage of sequence identity (id.) is indicated between 5' ends, central parts, and 3' ends of *cse* from CNRZ368 and LMG18311. Hatched boxes represent repeat-containing regions (*var-cse*). (B) Nucleic acid and amino acid sequences of *var-cse*_{CNRZ368} and the proximal region. The regions flanking *var-cse* are italicized. Each repeat unit is boxed, and the name of the repeat unit is indicated above the box. (C) Table of consensus repeat sequences represented in *cse*, *pcsB*, and *orf1*. (D) Schematic representation of repeat-containing regions. Each repeat unit is represented by a letter-containing box. Repeat units, in accordance with consensus sequences presented in panel C, are represented by capital letters, while those slightly divergent from the consensus (17% of maximum divergence) are represented by lowercase letters. Empty boxes represent regions without any repeats. Repeats in grey are common between *cse* from *S. thermophilus* strains CNRZ368 and LMG18311, *orf1*, and *pcsB* from *S. thermophilus* strain LMG18311. Hatched boxes represent repeats only found in *cse*.

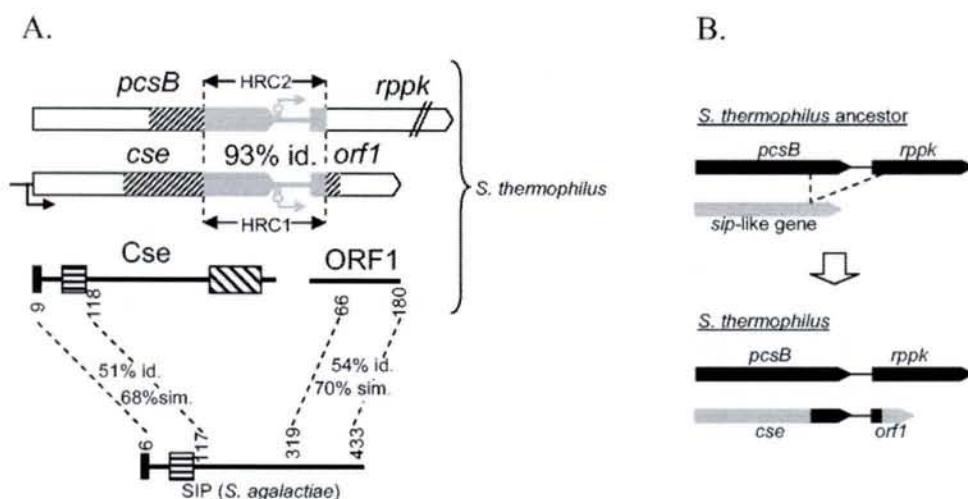


FIG. 3. *cse*, *orf1*, and their putative products are chimeric and may result from a duplication event. (A) Results from identity searches. Open arrows represent ORFs and indicate their reading directions, broken arrows indicate putative promoters, and hairpin loops symbolize putative rho-independent terminators. Hatched boxes represent repeat-containing regions, and grey color points out sequences from the *cse* and *pcsB* regions displaying high identity. Thick lines represent proteins, black boxes show putative signal peptides, and boxes with horizontal and oblique hatching symbolize putative LysM and CHAP domains, respectively. Homologous regions between proteins are linked, and positions of amino acids delimiting homologous regions are indicated. Identity (id.) and similarity (sim.) results are mentioned. (B) Duplication hypothesis. Grey and black arrows represent ORFs. Dashed lines both delimit the duplicated region and indicate the insertion site.

In contrast, the N-terminal part of Cse is homologous to the N-terminal part of the surface immunogenic protein (SIP) from *Streptococcus agalactiae* (Fig. 3A). Searches of the Conserved Domain Database showed that the N-terminal parts of both SIP and Cse contain regions homologous to the LysM domain (Fig. 3A), found in extracellular proteins and involved in their attachment to the cell wall (49). This domain is probably responsible for preferential localization of the SIP surface protein at the cell poles (46, 49).

Interestingly, the C-terminal region of SIP is homologous to the C-terminal part of a putative protein (ORF1) (Fig. 3A) encoded by the gene immediately downstream of *cse* (Fig. 3A).

The CDART database contains two hypothetical proteins from *Streptococcus mutans* and *Streptococcus intermedius* with one LysM domain and one CHAP domain (GenBank accession numbers NP_720819 and BAB61101, respectively). These proteins have a LysM domain at the N terminus and a CHAP domain at the C terminus. Both proteins show significant similarity (48% identity and 63% similarity) over 92% of their length. Their CHAP domains show significant similarity with Cse (51% identity and 62% similarity for NP_720819, 60% identity and 72% similarity for BAB61101), contrary to that of the LysM-containing region, which does not exhibit significant similarity with either Cse or SIP.

***pcsB* and *orf1* contain common repeats with *cse*.** Considering partial homology between *pcsB* and *cse* loci, *pcsB*_{LMG18311}, *orf1*_{LMG18311}, and *rppk*_{LMG18311} were scanned for repeats. The analysis revealed that the *pcsB*_{LMG18311} central region and *orf1*_{LMG18311} 5' end also contain repeats (Fig. 3A) that are common to both *cse* alleles (Fig. 2D). More precisely, all repeats found in *pcsB*_{LMG18311} (A, C, E, H, and I) are also represented in *var-cse*_{LMG18311}, and one repeat (A) found in *pcsB*_{LMG18311} is represented in *var-cse*_{CNRZ368}. The *orf1* 5' end from CNRZ368 and LMG18311 strains (99.6% identity between *orf1* alleles) contains C, E, and F repeated sequences,

although all but one are degenerate. Additionally, the succession ECCF is found twice in *orf1* and displayed once in *var-cse*_{LMG18311} (Fig. 2D).

Interestingly, *var-cse*_{LMG18311} and the *orf1*_{LMG18311} repeat-rich regions are located closed to HRC1 boundaries (Fig. 3A). Thus, HRC1 is flanked by two repeat-rich regions which contain repeated motifs common to each other and to the *pcsB*_{LMG18311} repeat-rich region.

Extracellular localization of Cse. According to SignalP (35) analysis, Cse contains a canonical putative signal peptide, suggesting that it is exported from the cell. To test this hypothesis, *cse* was fused to the reporter gene $\Delta_{sp}nuc$ carried by the plasmid pFUN (38). The $\Delta_{sp}nuc$ gene encodes a *Staphylococcus aureus* nuclease lacking its signal peptide (10). The region containing the putative promoter and the entire *cse* ORF (except the stop codon) was cloned into the pFUN vector to generate a *cse*- $\Delta_{sp}nuc$ fusion. The resulting pFUN::*cse* vector was introduced into *E. coli* DH5 α and *S. thermophilus* CNRZ368. Following growth, the transformants were overlaid with TBD agar (21), allowing detection of extracellular nuclease activity by visualization of a pink halo. A pink halo was visualized around colonies of each transformed strain (data not shown), indicating that the fusion protein was extracellular. As a negative control, the same region lacking the putative signal peptide encoding sequence was cloned into pFUN to generate a $\Delta_{sp}cse$ - $\Delta_{sp}nuc$ fusion. Transformed by this construct, both the *E. coli* and *S. thermophilus* colony surroundings remained blue. These results indicate that Cse is exported and that its export is signal peptide dependent.

Impact of *var-cse* variability on Cse cell segregation activity. The variability of *var-cse* raised the question of a possible variability of Cse cell segregation activity. Analysis of *S. thermophilus* CNRZ368 and LMG18311 cell distribution in chains revealed intraspecies variability of cell chain length (Fig. 4A). Indeed, 85% of CNRZ368 cells formed chains from 1 to 100

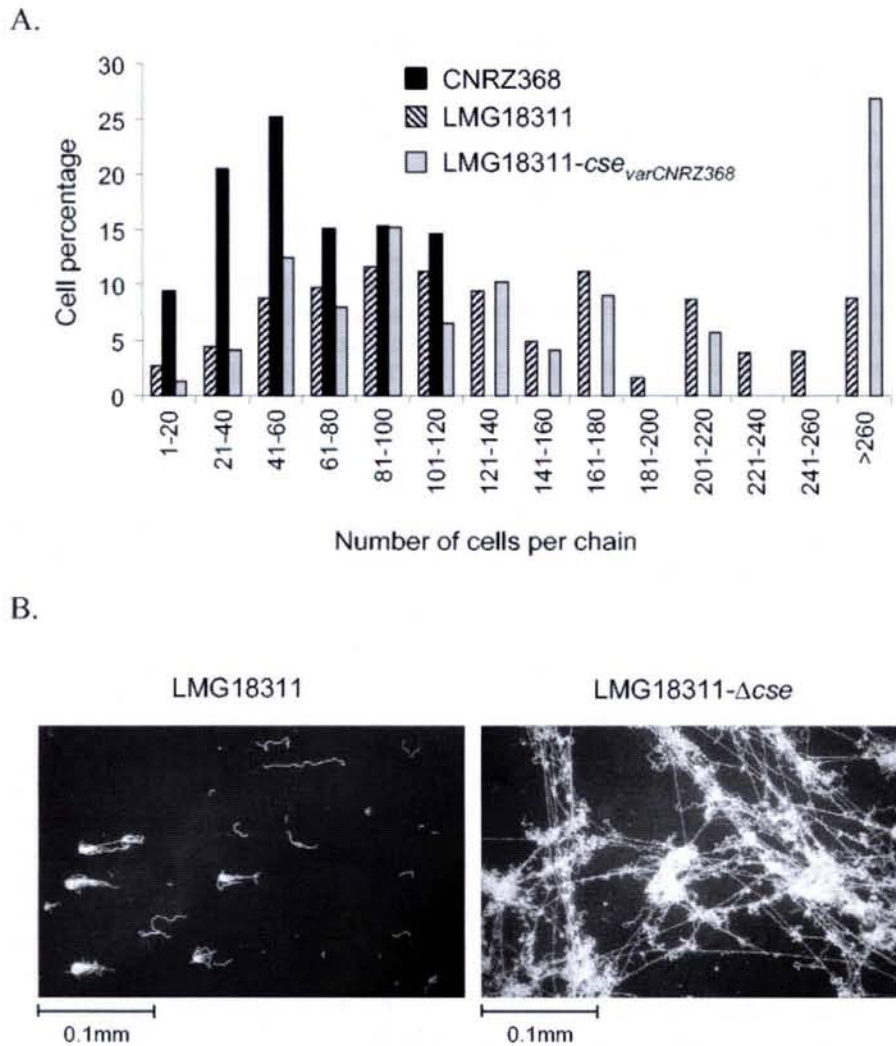


FIG. 4. Influence of *var-cse* variability on Cse cell segregation activity. (A) Results of counting numbers of cells per chain. A total of 1,000 cells were counted from three independent cultures. (B) Cell morphology of *S. thermophilus* strain LMG18311. Cells were photographed or counted in stationary phase after 20 h of growth. After this period of growth, OD_{600} values reached 0.66 ± 0.074 , 1.1 ± 0.014 , 1.1 ± 0.032 , and 0.87 ± 0.049 , respectively, for strains CNRZ368, LMG18311, LMG18311-*cse*_{varCNRZ368}, and LMG18311- Δ *cse*.

cells, whereas only 39% of LMG18311 cells are in this size range. Thus, a large proportion of LMG18311 cell chains were much longer than those of CNRZ368. Due to the involvement of *cse* in cell segregation, we hypothesized a correlation between variability of the cell chain length and *var-cse* variability. A Δ *cse* mutant of LMG18311 was found to exhibit a long-chain phenotype, as the CNRZ368- Δ *cse* strain does (Fig. 4B), confirming that the *cse*_{LMG18311} allele was functional. At the *cse* chromosomal locus, the *var-cse* region (Fig. 2D) was replaced by that of CNRZ368 in the LMG18311 genetic background. This was done by allelic replacement, taking into account that *var-cse* homologous flanking regions are not identical even if they show at least 94% identity (Fig. 2A). Thus, the plasmid used for allelic replacement should contain *var-cse* from strain CNRZ368 flanked by *cse* homologous regions from LMG18311 and LMG18311-*cse*_{varCNRZ368} strains with respect to chain length (Fig. 4A). This indicates that

the cell segregation activity of Cse is not affected by the high variability of its central region.

Cell wall hydrolase activity. Since it was previously reported that cell segregation results from cell wall hydrolytic activity detectable by zymography (6, 8, 18, 36, 41, 58), Cse was tested for this activity. Thus, wild-type and CNRZ368- Δ *cse* strain protein extracts were tested by zymography. Under our conditions, no difference was detected between the two lytic activity profiles, with either autoclaved *S. thermophilus* or *M. lysodeikticus* cells as a substrate (data not shown). Therefore, we tried to overproduce the Cse protein in *E. coli* with an N-terminal hexa-His tag. However, this approach was unsuccessful for the full-length Cse protein, as in the case of the full-length AcnB protein from *L. lactis* previously reported (16). The authors were able to circumvent this issue by overproducing AcnB fragments. Using the same alternative way, we were able to overproduce two Cse fragments in *E. coli*, one comprising the

region from amino acid 1 to 183 and the other from amino acid 184 to 461. Thus, the N-terminal region comprised the LysM putative domain, and the C-terminal region comprised both the Var-Cse region and the CHAP domain. Renaturing SDS-PAGE, performed with crude protein extract, either on autoclaved *S. thermophilus* or *M. lysodeikticus* cells, did not allow detection of any lytic activity attributable to either the N-terminal or the C-terminal fragment. To exclude that this negative result was not due to the incorrect refolding of the Cse fragments, a cell lysis test was performed under nondenaturing conditions. For this purpose, proteins of *E. coli* Cse fragment overproducer strains were extracted by sonication. The control crude protein extract was obtained from an *E. coli* strain carrying the pET15b plasmid without an insert. After checking that the overproduced proteins were recovered in the soluble fraction, 200 μ g of crude protein extract were added to a TPPY suspension of autoclaved *S. thermophilus* or *M. lysodeikticus* cells and incubated at 42°C. OD₆₀₀ measurements over time did not reveal any significant differences between crude protein extracts containing the overproduced Cse fragments and the control.

Finally, we tried an indirect way, by checking whether *cse* would be able to suppress a long-chain phenotype resulting from a murein hydrolase defect. For this purpose, *L. lactis* Δ *acmA* was transformed with pNST260+::*cse*, and the number of cells per chain was counted. No difference in chain length was noticed between Δ *acmA* and Δ *acmA* transformed with pNST260+::*cse* (data not shown).

DISCUSSION

Three genes of *S. thermophilus* (*rodA*, *pbp2b*, and *mreD*) have been previously shown to be involved in cell chain length (50, 52, 54). These genes are also involved in cell shape, since mutants in these genes display rod-shaped cells instead of ovoid cells. In this study, we identified a fourth gene involved in the shortening of cell chains, called *cse* for its role in cellular segregation. Deletion of *cse* and complementation experiments have shown that the *cse* chimeric gene is involved in cell segregation. The Cse extracellular protein contains a putative LysM domain and a putative CHAP domain. LysM domains of AcmA, the major peptidoglycan hydrolase of *L. lactis*, are directly involved in cell wall attachment (49), and CHAP domains are widely found in surface proteins where they are supposed to be involved in peptidoglycan hydrolysis (3, 45). Therefore, Cse may be attached to the cell wall and may exhibit peptidoglycan hydrolase activity.

To test whether Cse has a peptidoglycan hydrolase activity, protein extracts of *S. thermophilus* wild type, *S. thermophilus* Δ *cse*, and *E. coli* Cse fragment overproducer strains were tested by zymography with embedded *S. thermophilus* or *M. lysodeikticus* cells as a substrate. In addition, cell lysis activity was tested in liquid medium, with native crude protein extracts of *E. coli* Cse fragment overproducer strains. In both experiments, no Cse lytic activity was detected. Finally, *cse* was not able to complement the long-chain phenotype resulting from a peptidoglycan hydrolase defect of an *L. lactis* Δ *acmA* mutant. The absence of a detectable lytic activity was previously reported for PcsB from *S. agalactiae* (42, 43), another protein with a CHAP domain. These results suggest that either Cse has

no peptidoglycan hydrolase activity or Cse activity could not be detected by these assays.

The results reported here show that *cse* is a chimeric gene and suggested that its 5' and 3' ends originated from *sip*-like and *pcsB* genes, respectively. Indeed, the region comprising the 3' end of *cse*, the following intergenic region, and the beginning of *orf1* (region named HRC1) shows 93% identity with the region comprising the *pcsB* 3' end, the following intergenic region, and the beginning of *rppk* (region named HRC2) (Fig. 3A). Moreover, HRC1 is flanked by two regions encoding amino acid sequences homologous to parts of the SIP protein from *S. agalactiae* (Fig. 3A). Furthermore, SIP-encoding genes had been found in the genomes of all streptococci except that of *S. thermophilus* and *S. mutans*, and a PcsB-encoding gene had been found in the genomes of all streptococci, including *S. thermophilus*. Taken together, all of these data suggest that HRC2 had been subjected to a duplication event. The resulting duplicated sequence would have been inserted into a *sip* ancestral gene, thus creating two chimeric genes, *cse* and *orf1* (Fig. 3B).

Interestingly, *S. mutans* and *S. intermedius* possess one LysM-CHAP protein. The *S. mutans* genome, whose complete sequence is available, does not encode a protein with significant similarity to SIP of other streptococci and encodes a protein, named GbpB, homologous to whole PcsB (9). Thus, the genomic context of *S. mutans* would be similar to that of *S. thermophilus*, suggesting that Cse and the *S. mutans* LysM-CHAP protein have a common ancestor that would have been generated by module shuffling. However, the *S. mutans* LysM-CHAP protein does not show significant similarity to SIP and shows significant similarity with Cse only from its putative CHAP domain. Moreover, the C terminus of the LysM-CHAP protein harbors only 61% identity to the C terminus of GbpB, whereas the C terminus of Cse shows 93% identity with the C terminus of *S. thermophilus* PcsB (Fig. 3A). If the LysM-CHAP protein of *S. mutans* and Cse had been generated by the same DNA duplication event, it would be expected that the percentage of divergence between these LysM-CHAP proteins and their PcsB/GbpB genomic counterparts would be the same. These data support an alternative hypothesis where the LysM-CHAP protein from *S. mutans* and Cse have different origins.

The consequence of the suspected DNA rearrangement that originated *cse* is the association of a putative LysM-encoding sequence to a putative CHAP-encoding sequence, giving rise to a new protein generation (Cse) by domain shuffling. Domain shuffling has been proposed as a mechanism of gene evolution in bacteria (13, 15, 25, 44) and implies considering genes to be associations of modules. Each gene module would encode a protein domain which has been defined as "part of a protein that can fold up independently of neighboring sequences" (11). These modules could undergo rearrangements, ending in the creation of new module associations and, consequently, new protein domain associations. As an example, extracellular proteins from *S. pneumoniae* that bind the choline component of the cell wall appear to evolve by domain shuffling. These multidomain proteins consist of a choline-binding domain, triggering protein association to the cell wall, and a catalytic domain that differs from one protein to another (25). Another example is the peptidoglycan hydrolase domain of *Bacillus anthracis*

AmiA that is also found in other proteins where it is fused to different attachment signals (31).

The *cse* central region and the central region of its product are variable and repeat rich. The existence of variable tandemly repeated sequences is a common characteristic of extracellular proteins of gram-positive bacteria (20). This variability may facilitate adaptation to environmental changes (20), for instance, variability in the number of repeats in the alpha C protein allows group B streptococci to escape host immunity (26). In other cases, not restricted to extracellular proteins, genetic variability of repeat-rich regions directly modifies intrinsic biochemical activity of the protein, such as in restriction and modification enzymes encoded by EcoR124 and EcoR124/3 genes (40). These type I restriction enzymes recognize sites GAA(N6)RTCG and GAA(N7)RTCG, respectively, differing only in the length of a spacer. This difference in their specificity is due to two or three copies of a 12-bp sequence localized in their respective gene central parts (40). Therefore, we could not exclude that variations in the Cse repeat-rich region have consequences on cell segregation activity. However, replacement of *var-cse*_{LMG18311} by *var-cse*_{CNRZ368} at the LMG18311 *cse* chromosomal locus did not induce significant changes in cell segregation activity on the criterion of cell number per chain. Thus, it is likely that the high variability of the *cse* central region was not counter-selected because it does not affect the cell segregation activity of Cse.

The question of the role of the Cse central part in cell segregation remains open. This region could be a linker joining the N-terminal part to the C-terminal part of Cse. Such a function is assumed, for instance, by Q-linkers that join functionally distinct domains in nitrogen regulatory proteins (55). Variations in length and sequence of NifA and NtrC Q-linkers have no consequence on the activity of these proteins (55). Similarly, the replacement of *var-cse*_{LMG18311} by *var-cse*_{CNRZ368} ending in a 60-bp shortening of *var-cse*, since *var-cse*_{CNRZ368} is 60 bp shorter than *var-cse*_{LMG18311}, did not have consequences on Cse cell segregation activity. A linker role of the Cse central part could explain why its high variability, and especially its length variability, does not affect cellular segregation. This possible function of the Cse central part would also be consistent with its predicted non-globular structure.

Two repeat-rich regions sharing common repeat motifs flank HRC1. This suggests that the insertion event of the duplicated region occurred within an ancestral repeat-rich region localized in the central part of the *sip*-like gene ancestor. Thus, SIP-encoding genes of other streptococci could also contain a repeat-rich region. Therefore, repeat content and sequence complexity level were analyzed in the *S. agalactiae* SIP protein and orthologous counterparts in *S. pneumoniae*, *Streptococcus pyogenes*, *Streptococcus gordonii*, and *Streptococcus suis*. No repeat was found in any of these sequences. However, analysis of the amino acid composition showed that these regions have a low sequence complexity. Moreover, as observed for *cse*, their alignment revealed strong interspecies divergence of the central region (data not shown). Thus, although SIP proteins do not contain repeats, their central part showed low sequence complexity and interspecies variability. These data are in agreement with the existence of a low-complexity region in the central part of the *sip*-like gene ancestor of *cse*. We hypothesize that this low-complexity region could have been the site of

an HRC1 insertion event responsible for *cse* creation. Interestingly, other multidomain proteins exhibit junctions with low-complexity sequences between domains (44, 55). For instance, the multiphosphoryl transfer protein of *Rhodobacter capsulatus*, a permease from a phosphotransferase system, is composed of three domains connected by two similar linker regions of 17 residues. These two linkers are rich in glycine, alanine, and proline residues, lowering the complexity of these regions (57). It has been proposed that the evolution of PTS permease occurred by interdomain shuffling and that this shuffling was allowed by genetic recombination between linker-encoding sequences (57). This example, together with *cse*, raises the question of a possible evolutionary advantage, at the genetic level, of DNA regions with a low-complexity sequence. Recombination events occurring inside domain-encoding sequences will probably inactivate, in many cases, the domain functionality and will therefore give rise to inefficient domain associations. We speculate that low-complexity regions can be more tolerant targets for genetic recombination events, responsible for domain shuffling, because of their low requirement in sequence and size.

ACKNOWLEDGMENTS

We thank Isabelle Poquet and Emmanuelle Maguin for providing the pFUN and pGh9:ISS1 plasmids, respectively, and Valérie Legué for photographic contributions. We also thank Jan Kok for providing *L. lactis* MG1363*acmAD1* and Gérard Guédon for allowing us to use pNST260+ prior to publication. We are grateful to Paul Hoskisson for help in preparing the manuscript.

S.L. and A. F. were supported by grants from the Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche. F.B. was supported by a grant from the Institut National de la Recherche Agronomique. P.H. is Research Associate at FNRS.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bateman, A., and N. D. Rawlings. 2003. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem. Sci.* **28**:234–237.
- Bollinger, J. M., Jr., D. S. Kwon, G. W. Huisman, R. Kolter, and C. T. Walsh. 1995. Glutathionylspermidine metabolism in *Escherichia coli*. Purification, cloning, overproduction, and characterization of a bifunctional glutathionylspermidine synthetase/amidase. *J. Biol. Chem.* **270**:14031–14041.
- Bracqart, P. 1981. An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yogurt. *J. Appl. Bacteriol.* **51**:303–305.
- Buist, G., J. Kok, K. J. Leenhouts, M. Dabrowska, G. Venema, and A. J. Haandrikman. 1995. Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation. *J. Bacteriol.* **177**:1554–1563.
- Burrus, V., Y. Roussel, B. Decaris, and G. Guédon. 2000. Characterization of a novel integrative element, ICES1, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **66**:1749–1753.
- Carroll, S. A., T. Hain, U. Technow, A. Darji, P. Pashalidis, S. W. Joseph, and T. Chakraborty. 2003. Identification and characterization of a peptidoglycan hydrolase, MurA, of *Listeria monocytogenes*, a muramidase needed for cell separation. *J. Bacteriol.* **185**:6801–6808.
- Chia, J. S., L. Y. Chang, C. T. Shun, Y. Y. Chang, Y. G. Tsay, and J. Y. Chen. 2001. A 60-kilodalton immunodominant glycoprotein is essential for cell wall integrity and the maintenance of cell shape in *Streptococcus mutans*. *Infect. Immun.* **69**:6987–6998.
- Cuatrecasas, P., S. Fuchs, and C. B. Anfinsen. 1967. Catalytic properties and specificity of the extracellular nuclease of *Staphylococcus aureus*. *J. Biol. Chem.* **242**:1541–1547.
- Doolittle, R. F. 1995. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**:287–314.
- Garcia, P., M. P. Gonzalez, E. Garcia, R. Lopez, and J. L. Garcia. 1999.

- LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol. Microbiol.* **31**:1275–1277.
13. Gubler, M., D. Bragaglia, J. Meyer, A. Piekawicz, and T. A. Bickle. 1992. Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* **11**:233–240.
 14. Hardie, J. M. 1986. Genus *Streptococcus* Rosenbach 1884, 22nd, p. 1043–1047. In P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 2. William & Wilkins, Baltimore, Md.
 15. Houghton, J. E., G. A. O'Donovan, and J. R. Wild. 1989. Reconstruction of an enzyme by domain substitution effectively switches substrate specificity. *Nature* **338**:172–174.
 16. Huard, C., G. Miranda, F. Wessner, A. Bolotin, J. Hansen, S. J. Foster, and M. P. Chapot-Chartier. 2003. Characterization of AcmB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*. *Microbiology* **149**:695–705.
 17. Husson-Kao, C., J. Mengaud, L. Benbadis, and M. P. Chapot-Chartier. 2000. MurI, a *Streptococcus thermophilus* peptidoglycan hydrolase devoid of a specific cell wall binding domain. *FEMS Microbiol. Lett.* **187**:69–76.
 18. Ishikawa, S., Y. Hara, R. Ohnishi, and J. Sekiguchi. 1998. Regulation of a new cell wall hydrolase gene, *cwlF*, which affects cell separation in *Bacillus subtilis*. *J. Bacteriol.* **180**:2549–2555.
 19. Jones, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**:195–202.
 20. Kehoe, M. A. 1994. Cell-wall-associated proteins in Gram-positive bacteria, p. 217–261. In J.-M. G. A. R. Hakenbeck (ed.), *Bacterial cell wall*. Elsevier Science B.V., Amsterdam, The Netherlands.
 21. Lachica, R. V., C. Genigeorgis, and P. D. Hoepflich. 1971. Metachromatic agar-diffusion methods for detecting staphylococcal nuclease activity. *Appl. Microbiol.* **21**:585–587.
 22. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
 23. Leenhouts, K. 1995. Integration strategies and vectors. *Dev. Biol. Stand.* **85**:523–530.
 24. Le Loir, Y., A. Gruss, S. D. Ehrlich, and P. Langella. 1994. Direct screening of recombinants in gram-positive bacteria using the secreted staphylococcal nuclease as a reporter. *J. Bacteriol.* **176**:5135–5139.
 25. Lopez, R., M. P. Gonzalez, E. Garcia, J. L. Garcia, and P. Garcia. 2000. Biological roles of two new murein hydrolases of *Streptococcus pneumoniae* representing examples of module shuffling. *Res. Microbiol.* **151**:437–443.
 26. Madoff, L. C., J. L. Michel, E. W. Gong, D. E. Kling, and D. L. Kasper. 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc. Natl. Acad. Sci. USA* **93**:4131–4136.
 27. Maguin, E., P. Duwat, T. Hege, D. Ehrlich, and A. Gruss. 1992. New thermosensitive plasmid for gram-positive bacteria. *J. Bacteriol.* **174**:5633–5638.
 28. Maguin, E., H. Prevost, S. D. Ehrlich, and A. Gruss. 1996. Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J. Bacteriol.* **178**:931–935.
 29. McGuffin, L. J., K. Bryson, and D. T. Jones. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**:404–405.
 30. Mercier, C., C. Durrieu, R. Briandet, E. Domakova, J. Tremblay, G. Buist, and S. Kulakauskas. 2002. Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol. Microbiol.* **46**:235–243.
 31. Mesnage, S., and A. Fouet. 2002. Plasmid-encoded autolysin in *Bacillus anthracis*: modular structure and catalytic properties. *J. Bacteriol.* **184**:331–334.
 32. Nakamura, Y., T. Gojobori, and T. Ikemura. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**:292.
 33. Ng, W. L., K. M. Kazmierczak, and M. E. Winkler. 2004. Defective cell wall synthesis in *Streptococcus pneumoniae* R6 depleted for the essential PcsB putative murein hydrolase or the VicR (YycF) response regulator. *Mol. Microbiol.* **53**:1161–1175.
 34. Ng, W. L., G. T. Robertson, K. M. Kazmierczak, J. Zhao, R. Gilmour, and M. E. Winkler. 2003. Constitutive expression of PcsB suppresses the requirement for the essential VicR (YycF) response regulator in *Streptococcus pneumoniae* R6. *Mol. Microbiol.* **50**:1647–1663.
 35. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**:1–6.
 36. Ohnishi, R., S. Ishikawa, and J. Sekiguchi. 1999. Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J. Bacteriol.* **181**:3178–3184.
 37. Pluvinet, A., F. Charron-Bourgoin, C. Morel, and B. Decaris. 2004. Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. *Int. Dairy J.* **14**:627–634.
 38. Poquet, L., S. D. Ehrlich, and A. Gruss. 1998. An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *J. Bacteriol.* **180**:1904–1912.
 39. Potvin, C., D. Leclerc, G. Tremblay, A. Asselin, and G. Bellemare. 1988. Cloning, sequencing and expression of a *Bacillus* bacteriolytic enzyme in *Escherichia coli*. *Mol. Gen. Genet.* **214**:241–248.
 40. Price, C., J. Lingner, T. A. Bickle, K. Firman, and S. W. Glover. 1989. Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**:115–125.
 41. Qin, X., K. V. Singh, Y. Xu, G. M. Weinstock, and B. E. Murray. 1998. Effect of disruption of a gene encoding an autolysin of *Enterococcus faecalis* OG1RF. *Antimicrob. Agents Chemother.* **42**:2883–2888.
 42. Reinscheid, D. J., K. Ehlert, G. S. Chhatwal, and B. J. Eikmanns. 2003. Functional analysis of a PcsB-deficient mutant of group B streptococcus. *FEMS Microbiol. Lett.* **221**:73–79.
 43. Reinscheid, D. J., B. Gottschalk, A. Schubert, B. J. Eikmanns, and G. S. Chhatwal. 2001. Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J. Bacteriol.* **183**:1175–1183.
 44. Reizer, J., and M. H. Saier, Jr. 1997. Modular multidomain phosphoryl transfer proteins of bacteria. *Curr. Opin. Struct. Biol.* **7**:407–415.
 45. Rigen, D. J., M. J. Jedrzejak, and M. Y. Galperin. 2003. Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem. Sci.* **28**:230–234.
 46. Rioux, S., D. Martin, H. W. Ackermann, J. Dumont, J. Hamel, and B. R. Brodeur. 2001. Localization of surface immunogenic protein on group B streptococcus. *Infect. Immun.* **69**:5162–5165.
 47. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 48. Schleifer, K. H. 1986. Gram-positive cocci, p. 999–1002. In P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 2. William & Wilkins, Baltimore, Md.
 49. Steen, A., G. Buist, K. J. Leenhouts, M. El Khattabi, F. Grijpstra, A. L. Zomer, G. Venema, O. P. Kuipers, and J. Kok. 2003. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J. Biol. Chem.* **278**:23874–23881.
 50. Stingle, F., and B. Mollet. 1996. Disruption of the gene encoding penicillin-binding protein 2b (*pbp2b*) causes altered cell morphology and cease in exopolysaccharide production in *Streptococcus thermophilus* Sfl6. *Mol. Microbiol.* **22**:357–366.
 51. Terzaghi, B., and W. Sandine. 1975. Improved medium for lactic streptococci and their bacteriophages. *Appl. Environ. Microbiol.* **29**:807–813.
 52. Thibessard, A., F. Borges, A. Fernandez, B. Gintz, B. Decaris, and N. Leblond-Bourget. 2004. Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. *Appl. Environ. Microbiol.* **70**:2220–2229.
 53. Thibessard, A., A. Fernandez, B. Gintz, B. Decaris, and N. Leblond-Bourget. 2002. Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can. J. Microbiol.* **48**:473–478.
 54. Thibessard, A., A. Fernandez, B. Gintz, N. Leblond-Bourget, and B. Decaris. 2002. Effects of *rodA* and *pbp2b* disruption on cell morphology and oxidative stress response of *Streptococcus thermophilus* CNRZ368. *J. Bacteriol.* **184**:2821–2826.
 55. Wootton, J. C., and M. H. Drummond. 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.* **2**:535–543.
 56. Wootton, J. C., and S. Federhen. 1996. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**:554–571.
 57. Wu, L. F., J. M. Tomich, and M. H. Saier, Jr. 1990. Structure and evolution of a multidomain multiphosphoryl transfer protein. Nucleotide sequence of the fruB(HI) gene in *Rhodobacter capsulatus* and comparisons with homologous genes from other organisms. *J. Mol. Biol.* **213**:687–703.
 58. Wuenschel, M. D., S. Kohler, A. Bubert, U. Gerike, and W. Goebel. 1993. The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity. *J. Bacteriol.* **175**:3491–3501.
 59. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.

**LE GENE MOSAÏQUE ET VARIABLE *CSE* EST IMPLIQUÉ
DANS LA SEGREGATION CELLULAIRE CHEZ
*STREPTOCOCCUS THERMOPHILUS***

Frédéric BORGES, Séverine LAYEC, Annabelle THIBESSARD, Annabelle FERNANDEZ,
Brigitte GINTZ, Bernard DECARIS et Nathalie LEBLOND-BOURGET

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences de
l'Université Henri Poincaré Nancy 1, BP239, 54506 Vandoeuvre-lès-Nancy, France.

Les streptocoques se développent sous la forme de chaînettes de cellules dont la longueur influe la capacité à former un biofilm (1). Nous avons identifié chez *S. thermophilus* CNRZ368 un gène, nommé *cse* (Cell Segregation), dont la délétion entraîne une augmentation drastique du nombre de cellules par chaînettes. Le phénotype mutant est complétement corrigé lorsqu'une copie sauvage du gène est ajoutée en trans montrant ainsi que *cse* est impliquée dans la ségrégation cellulaire. La protéine Cse présente un peptide signal putatif suggérant la localisation extracellulaire de Cse. Une fusion entre *cse* et un gène rapporteur a été construite et montre que la protéine est exportée.

Cse présente en N-terminal un domaine de liaison à la paroi cellulaire putatif (domaine LysM) et en C-terminal un domaine amidase putatif (domaine CHAP). La comparaison de la séquence du locus *cse* avec celles disponibles dans les banques de données montre que *cse* est un gène mosaïque. Il aurait été créé par duplication du fragment codant un domaine CHAP d'un autre gène de *S. thermophilus*, le gène *pcsB*, et insertion de ce fragment de gène à l'intérieur d'un gène ancestral codant un domaine LysM. Cet événement aurait ainsi permis de réunir au sein de la même protéine un domaine LysM et un domaine CHAP et ainsi de générer une protéine fonctionnelle impliquée dans la ségrégation cellulaire.

Ces deux domaines sont séparés par une région riche en motifs répétés. La comparaison de la séquence du gène *cse* ou de la protéine Cse de différentes souches de *S. thermophilus*, montre que cette région centrale est variable. La délétion du gène *cse* de la souche *S. thermophilus* LMG18311, dont la région centrale est très différente de la souche CNRZ368, montre que l'allèle *cse* de cette souche code une protéine de ségrégation cellulaire également fonctionnelle. Ces résultats suggèrent que d'importantes modifications de la région centrale de la protéine Cse peuvent avoir lieu sans que la protéine soit inactivée. La fonction de la protéine Cse ainsi que le rôle de la variabilité de la région centrale du gène sont en cours d'étude.

1. Mercier, C., C. Durrieu, R. Briandet, E. Domakova, J. Tremblay, G. Buist, and S. Kulakauskas. 2002. Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol Microbiol* **46**:235-43.

Annabelle Fernandez · Annabelle Thibessard
Frédéric Borges · Brigitte Gintz · Bernard Decaris
Nathalie Leblond–Bourget

Characterisation of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368

Received: 19 April 2004 / Revised: 1 July 2004 / Accepted: 8 July 2004 / Published online: 17 September 2004
© Springer-Verlag 2004

Abstract During industrial processes, the dairy organism *Streptococcus thermophilus* is exposed to stress conditions. Its ability to survive and grow in an aerobic environment indicates that it must possess defensive mechanisms against reactive oxygen species. To identify the genes involved in oxidative stress defence, a collection of mutants was generated by random insertional mutagenesis and screened for menadione sensitivity and resistance. Results obtained for resistant clones allowed the identification of eight loci. The insertions affected genes whose homologues in other bacteria were previously identified as being involved in stress response (*deoB*, *gst*) or transcription regulation (*rggC*) and five ORFs of unknown function. The tolerance of the eight mutants to air-exposure, methyl viologen and H₂O₂ was studied. Real-time quantitative PCR was used to analyse the transcript level of mutated genes and revealed that most were down-regulated during oxidative stress.

Keywords *Streptococcus thermophilus* · Oxidative stress · Real-time quantitative RT-PCR

Introduction

Streptococcus thermophilus is widely used as a starter strain in the production of yogurt, Swiss-type cheeses and Mozzarella cheese. During the production or conservation of these products, *S. thermophilus* is exposed to potentially stressful environmental conditions, including contact with oxygen. To thrive in such an environment, *S. thermophilus* must possess defensive mechanisms

against the deleterious effects of oxygen caused by the production of reactive oxygen species.

The key enzymes of the oxidative stress response are generally well conserved throughout the bacterial world (Koonin et al. 2000) and, as in other bacteria, streptococci are equipped with superoxide dismutase (SOD), NADH oxidase, peroxidase and glutathione reductase. SOD activity represents the major mechanism of defence against superoxide stress by converting superoxide anions to molecular oxygen and H₂O₂ (Fridovich 1997). Four types of SOD can be distinguished in bacteria, depending on the metal cofactor (Cu–Zn, Fe, Mn or Ni). Most streptococci possess only a MnSOD encoded by the *sodA* gene (Poyart et al. 2001; Jakubovics et al. 2002), whereas in *S. pneumoniae* two co-factored SOD (MnSOD, FeSOD) have been found (Yesilkaya et al. 2000). To remove O₂, streptococci also utilise the action of NADH oxidase or pyruvate oxidase (Gibson et al. 2000; Pericone et al. 2003). In *S. mutans*, two distinct NADH oxidases have been identified, catalysing either the two-electron reduction of O₂ to H₂O₂ (Nox1) or the four-electron reduction of O₂ to H₂O (Nox2; Higuchi et al. 1999).

In the genus *Streptococcus*, H₂O₂ generated by SOD or NADH oxidase is metabolised by various peroxidase enzymes. For example, protection against H₂O₂ and organic peroxides in *S. pyogenes* involves a glutathione peroxidase and an alkyl hydroperoxide reductase (King et al. 2000) and, in *S. parasanguis*, a thiol peroxidase (Spatafora et al. 2002). Moreover, streptococci contain Dpr that provides a means to tolerate H₂O₂, by binding free intracellular iron and thus inhibiting the Fenton chemistry-catalysed formation of toxic hydroxyl radicals (Yamamoto et al. 2000).

Streptococcus, however, lacks some proteins, such as catalase and the global regulator OxyR, that have been shown, through the analysis of multiple bacterial species, to protect against oxidative stress. Moreover, in streptococci the pool of enzymes involved in oxidative stress defence can vary from one species to another. For instance, the H₂O₂ scavenger NADH peroxidase is

A. Fernandez · A. Thibessard · F. Borges · B. Gintz
B. Decaris · N. Leblond–Bourget (✉)
Laboratoire de Génétique et Microbiologie,
UMR INRA 1128, IFR 110,
Faculté des Sciences de l'Université Henri Poincaré Nancy 1,
BP 238, 54506 Vandoeuvre-lès-Nancy, France
E-mail: bourget@nancy.inra.fr

absent in *S. pneumoniae* (Hoskins et al. 2001; Tettelin et al. 2001) and *S. suis* (Niven and Ekins 2001). Thus, although the *Streptococcus* oxidative-stress defence mechanism shares common characteristics with other bacteria, it also provides a good model for novel oxidative stress response mechanisms in bacteria.

In order to better understand the ability of *S. thermophilus* to cope with reactive oxygen species, mutants from the strain CNRZ368 were generated, screened and characterised. In a previous study, menadione-sensitive mutants were analysed and revealed that genes involved in cell wall metabolism, exopolysaccharide translocation, tRNA modification and iron metabolism also play a role in cell survival during oxidative stress (Thibessard et al. 2004). In this work, a genetic and phenotypic analysis of menadione-resistant mutants is presented.

Materials and methods

Bacterial strains, plasmid and growth conditions

Bacterial strains used in this study were derived from *S. thermophilus* CNRZ368 purchased from the INRA-CNRZ368 strain collection (Jouy en Josas, France). Depending on the experiments, *S. thermophilus* CNRZ368 and its derivatives were cultivated at 30°C or 42°C, in milk medium (for pre-cultures), TPPY (Bracquart 1981) or M17 (Terzaghi and Sandine 1975) media. When required, 2 µg/ml erythromycin was added.

The plasmid pGh9:ISS1 (Maguin et al. 1996) used for insertional mutagenesis served to generate recombinant plasmids. *E. coli* EC101 (Leenhouts 1995), a TG1-derived strain [*supE hsd-5 thi Δ(lac-proAB) F' (traD6pro-AB⁺ lacI^q lacZΔM15)*] containing the pWV01 *repA* gene, was transformed with pGh9:ISS1 and derivatives. The resulting transformants were selected at 37°C on Luria-Bertani medium containing 150 µg/ml erythromycin.

Insertional mutagenesis and isolation of menadione-resistant mutants

Mutagenesis with pGh9:ISS1 and cloning of the chromosomal target of the plasmid were performed as described by Thibessard et al. (2002). The screening procedure used to select menadione-sensitive and resistant mutants is detailed by Thibessard et al. (2004) and is briefly described below. Two experimental procedures were performed, differing mainly in the length of menadione exposure. For long-term menadione exposure, mutant pre-cultures were inoculated onto M17 agar plates containing 2 µg/ml erythromycin and various concentrations of menadione (0–120 µg/ml). Plates were incubated for 20 h at 42°C. For short-term menadione exposure, mutant pre-cultures were inoculated in liquid TPPY medium and grown for 3.5 h at 42°C. Then, each culture was exposed to various menadione concentra-

tions (0–15 mg/ml). After 3 h incubation at 42°C, cells were dropped on TPPY agar plates containing erythromycin and incubated for 20 h at 42°C. For each procedure, clones were considered resistant when they grew on a plate on which less than 10% of the clones were growing.

General DNA techniques

DNA manipulations were performed according to Sambrook et al. (1989). Sequencing was carried out with the DNA sequencing kit (Applied Biosystems) on an ABI Prism 377 genetic analyser (Applied Biosystems). Sequence data were analysed with BLAST, ProDom, the TIGRFam model, TMHMM and HMMTOP software. DNA sequences obtained in this work were deposited in GenBank under accession numbers from AY598747 to AY598752.

RNA manipulation

To extract total RNA, *S. thermophilus* CNRZ368 was cultivated in TPPY medium containing 200 mM MOPS and incubated at 42°C to an optical density at 600 nm (OD₆₀₀)=0.6. The culture was subdivided into three aliquots. One was kept at 42°C without addition of H₂O₂ (control experiment); and in the other two, 1 mM H₂O₂ was added for 20 min or 45 min. Total RNA was extracted with a guanidinium thiocyanate and phenol-chloroform step (Chomczynski and Sacchi 1987) using the Tri-Reagent (Sigma-Aldrich). DNA-free RNA was obtained by the action of DNase I (0.5 units/µg RNA) for 15 min at 37°C. The absence of contaminating genomic DNA was checked by PCR, directly on RNA samples.

cDNA was generated from 2.5 µg of RNA by the action of 100 units of MMLV-reverse transcriptase at 37°C for 60 min in the presence of hexamer oligonucleotide random primers at a concentration of 100 ng/µl (Amersham Pharmacia Biotech).

Primers for real-time RT-PCR (Table 1) were designed with Primer3 software and purchased from MWG Biotech. Amplification of specific products was performed with the iCycler (BioRad) with the following program: 1 cycle at 50°C for 2 min, 1 cycle at 95°C for 10 min, 35 cycles at 95°C for 15 s and at 55°C for 1 min. The amplification reactions were carried out with qPCR Mastermix for Sybr Green I (Eurogentec) in a final volume of 25 µl containing reaction buffer (1×), a 1:66,000 dilution of Sybr Green I, 400 nM (each) forward and reverse primer, 10 nM fluorescein and cDNA template (83 ng or 28 ng, depending on the experiment). A standard curve for efficiency of PCR determination was plotted for each primer set with a serially diluted cDNA template. Melting curve analysis was performed with 0.5°C increments every 10 s from 55°C to 90°C to check that the cDNA amplification did not lead to

Table 1 Real-time RT-PCR primers

PCR template gene	Primer used for real-time RT PCR	
	Name	Gene sequence
<i>ldh</i>	ldh.F	AAGCTATCCTTGACGATGAA
	ldh.R	AATAGCAGGTTGACCGATAA
<i>osrA (deoB)</i>	osrA.F	ACCCAACCTTATGCTGGTACA
	osrA.R	AAATGTCCAACCTGGAATCAA
<i>osrB (gst)</i>	osrB.F	GCGCAATTGTAGTTCTACTTG
	osrB.R	AAGGACATTCAAAGCTGTTCT
<i>osrC (rggC)</i>	osrC.F	ACTTGTTTTCTGCGGAATACT
	osrC.R	TCCGAACGATGACACATT
<i>osrD</i>	osrD.F	ACCGTCGTGAATTTCTTAAA
	osrD.R	CTTGAACGATCCTGATTAGC
<i>osrE</i>	osrE.F	ACATGGTGTCTTCTGGTAA
	osrE.R	TTCTCCAATCAATGGTTGT
<i>osrF</i>	osrF.F	TTGGAATAGCTTCGGATTTA
	osrF.R	TAAGCCTCCAATAACAGCAT
<i>osrG</i>	osrG.F	TCCCAAATAGCCCAATAGTA
	osrG.R	ATAGGCGGGAACCTCTCTAAA
<i>osrH (orfP)</i>	osrH.F	TGAGGATGACTCTGATGACA
	osrH.R	AAAGGCATCCTCTCCATAAT

secondary products. Quantification of tested gene expression was done using the comparative $\Delta\Delta C_t$ method (Livak and Schmittgen 2001). Cycle threshold (C_t) values were defined as the cycle number at which the fluorescence exceeded a fixed threshold value above the baseline.

Determination of tolerance to other oxidative stress

Exposure to air in liquid medium *S. thermophilus* and its derivatives were grown in TPPY at 42°C to the stationary phase. Each culture was divided into five parts. Four of them were shaken (200 rpm) in a ratio of liquid to air space of approximately 3/1 and the fifth was not shaken (control experiment). Appropriate dilutions of each aliquot were spread on TPPY. The fraction of surviving colony-forming units (CFU) was calculated by dividing the CFU/ml count after growth under aerated conditions by the CFU/ml count for the control experiment.

Exposure to methyl viologen Overnight cultures in milk medium were diluted 100-fold into TPPY. After 8 h growth, cultures were diluted 150-fold in fresh TPPY medium, split and exposed to different concentrations of methyl viologen (Sigma-Aldrich). After 14 h incubation at 42°C, the OD₆₀₀ was measured and the growth ratio was determined by comparing the final OD obtained with and without treatment.

Exposure to H₂O₂ Overnight cultures in milk medium were diluted 1,000-fold in TPPY containing different concentrations of H₂O₂. Following 14 h incubation at 42°C, the OD₆₀₀ was measured and the growth ratio was determined by comparing the final OD with and without treatment.

Results

Selection of menadione-resistant mutants

A collection of 2,112 mutants was constructed by insertional mutagenesis (Thibessard et al. 2002). Two different procedures (see Materials and methods) assessing resistance to menadione, a superoxide-generating compound, allowed the selection of eight clones exhibiting a menadione-resistance phenotype. These clones were named *osrA* to *osrH* (for oxidative stress resistance locus) and were further characterised.

Identification of genes involved in menadione-resistant phenotype

Replicative transposition of *ISS1*, carried by the pGh9:*ISS1* plasmid, resulted in the integration of the entire plasmid in the chromosome, flanked by two direct copies of *ISS1*. Chromosomal targets of pGh9:*ISS1* insertion were recovered on replicative plasmids as described by Thibessard et al. (2002) and sequenced. In each case, the plasmid interrupted an ORF and the corresponding sequence was used to establish DNA and protein identity with database sequences.

Six out of the eight target gene products (OsrA to OsrF) showed sequence similarities with proteins with an *E*-value lower than 10⁻¹⁰ for BLASTX searches. The last two sequences (OsrG, OsrH) showed no significant identity to any database sequence. According to their putative function, target genes were separated into four classes (from I to IV) and renamed where relevant (Table 2).

Class I: genes already known to be involved in lactic acid bacteria stress defence mechanism

The first class comprised *osrA* and *osrB* mutants, both disrupted in ORFs whose products share homology with proteins already involved in stress response.

The sequence disrupted in the *osrA* mutant encodes a protein exhibiting 100% identity with DeoB from *S. thermophilus* NCBF2393, and is renamed *deoB*. The DeoB protein is a phosphopentomutase, that catalyses the phosphotransfer between the C1 and C5 carbon atoms of pentose. This enzyme is involved in the purine salvage pathway and assimilation of exogenous free bases or nucleosides from the environment (Nygaard 1993). In *Lactococcus lactis*, *deoB* mutations confer higher sensitivity to UV stress (Duwat et al. 1997) and increased resistance to acid and heat stress, demonstrating the importance of purine metabolism in stress defence (Rallu et al. 2000).

The disrupted ORF in the *osrB* mutant corresponds to *gst*, a gene localised in ICE *S11*, an integrative and conjugative element of *S. thermophilus* CNRZ368 (Burrus et al. 2002). The Gst (OsrB) protein shows 33%

Table 2 Characterisation of *osr* ORFs. The definite gene name is in brackets when applicable. Organism, gene name and, in brackets, accession number and length are given for the protein sequence showing the best alignment score. Percentage of identity and the portion of the protein sharing this identity (numbers correspond to

the amino acid positions) are indicated for the protein showing the best alignment score to each *S. thermophilus* *Osr* protein. Insertion site corresponds to the deduced amino acid position of the pGh9:IS *SI*-generated disruption in the full-length protein showing the best alignment score to the *Osr* proteins

Gene name	Accession number	Gene encoding the most similar product	Amino acid identity	Insertion site	Putative ORF function	Class
<i>osrA</i> (<i>deoB</i>)	AY598747	<i>S. thermophilus</i> NCFB2393 <i>deoB</i> (AAM93387, 403)	98% (133–403)	133	Phosphopentomutase	I
<i>osrB</i> (<i>gst</i>)	AJ278471	<i>Yersinia pestis</i> <i>glpP</i> (AH0031, 438)	33% (9–403)	404	Proton/sodium-dicarboxylate symporter	
<i>osrC</i> (<i>rggC</i>)	AY598748	<i>S. mutans</i> <i>rgg</i> (AAN59160, 285)	57% (3–284)	242	Transcriptional regulator	II
<i>osrD</i>	AY598749	<i>S. pneumoniae</i> R6 <i>spr1179</i> (AAK99982, 334)	78% (1–174)	48	Conserved hypothetical protein	III
<i>osrE</i>	AY598750	<i>Lactococcus lactis</i> ORF00050 (T43121, 614)	23% (3–399)	59	Conserved hypothetical protein	
<i>osrF</i>	AY598751	<i>P. horikoshii</i> PH1052 (H71098, 307)	35% (3–303)	169	Conserved hypothetical protein	
<i>osrG</i>	AY598752	–	778 nt ^a	–	–	IV
<i>osrH</i> (<i>orfP</i>)	AJ278471	–	1317 nt ^a	–	–	

^aSequence analysed length (in nucleotides, nt) for the mutants showing no homology.

identity with a proton-glutamate/aspartate symport protein from *Yersinia pestis* and with other proteins belonging to the proton/sodium-dicarboxylate symporter protein family (SDF). Hydropathy profiles of *Gst*, determined using TMHMM (Sonnhammer et al. 1998) and HMMTOP (Tusnady and Simon 1998), revealed the presence of ten trans-membrane regions, as observed in proteins belonging to the SDF family (Storck et al. 1992). The link between glutamate transport and stress response has already been suggested in *L. lactis* (Rallu et al. 2000). The short intergenic region and the lack of a rho-independent transcriptional terminator between *osrB* and the downstream ORF suggest that these two ORFs form an operon. This adjacent ORF corresponds to the already-named *orfX* gene of ICE *St1* (Burrus et al. 2002) and encodes a protein showing 49% identity with a hypothetical protein BAC15033 from *Oceanobacillus ihyensis*. These proteins belong to the RimK family, comprising a potential alpha-L-glutamate ligase, whose function is to add glutamate residues to maturing proteins. Thus, *gst* and *orfX*, like most operon-organised genes, seem to be involved in the same function: *Gst* could import glutamate residues needed for *OrfX* activity.

Class II: gene involved in transcriptional regulation

In the *osrC* mutant, pGh9:ISS1 disrupted an ORF (*rggC*₂, 699 bp) whose hypothetical product shares 54% identity with the C-terminal region of *Rgg* from *S. mutans*. This ORF is preceded by another ORF (*rggC*₁, 198 bp) in the same orientation, whose hypothetical product shares 77% identity with the N-terminal part of the same protein. This apparent frameshift was confirmed by sequencing this region directly on the *S. thermophilus* CNRZ368 chromosome. A putative Shine–Dalgarno sequence located 8 bp upstream of

*rggC*₁ was observed, suggesting this ORF was probably translated. In contrast, no canonical Shine–Dalgarno sequence preceding *rggC*₂ was detected. A full-length protein (*RggC*) might be translated, implying a translational or transcriptional frameshifting. This event could occur between positions 136 and 144, where a 9-mer (TTT TTT TTT), could serve as a substrate, as described by Larsen et al. (2000). The sequence of *RggC*₁ displays a characteristic helix-turn-helix DNA-binding motif (Schell 1993) expected for *Rgg*-like proteins. These proteins constitute a family of transcriptional regulators that coordinate the expression of various functions in Gram-positive bacteria.

Class III: genes of unknown function

The *osrD* mutant was found to be disrupted in a gene encoding a putative protein that shares 78% identity with the conserved hypothetical protein *spr1179* from *S. pneumoniae* R6. *OsrD* and *spr1179* proteins show significant identity with predicted iron-dependent peroxidases belonging to the family of dye-decolorising (or DyP-type) peroxidases. DyP peroxidases are characterised by an atypical heme-binding region (Sugano et al. 2000). Sequence analysis of *OsrD* indicated the presence of a twin-arginine translocation (TAT) signal sequence followed by a membrane-spanning hydrophobic region. The TAT domain allows translocation of a folded protein with a bound cofactor across the membrane (Berks et al. 2000). Analysis of the surrounding sequence in the *S. thermophilus* LMG18311 genome suggests that the *osrD* gene and the downstream ORF are part of an operon. This downstream ORF encodes a putative lipoprotein involved in the transport of iron, necessary for peroxidase activity. Thus, once again, both *osrD* and the downstream ORF seem to contribute to the same function.

The OsrE protein shows 23% identity with a hypothetical protein encoded by the *L. lactis* pMRC01 plasmid. These two proteins present a putative catalytic domain typical of kinase protein.

The *osrF* product shows 33% identity with the hypothetical protein PH1052 of the archaean, *Pyrococcus horikoshii*.

Class IV: genes with no known homologue

No sequence identities were detected for the *osrG* and *osrH* genes, which are thus considered as specific to *S. thermophilus*. Therefore, no hypothesis can be drawn about their function. The *osrH* gene is carried by ICE *St1* from *S. thermophilus* CNRZ368 and was named *orfP* by Burrus et al. (2002).

Menadione-resistant mutant growth rates

Growth rate is an important parameter for the industrial use of *S. thermophilus*. In TPPY medium, seven out of the eight mutants exhibited a doubling time similar to that of the parental CNRZ368 strain (32.9 ± 2.1 min). Only the *deoB* mutant grew faster (24.8 ± 2.8 min). Therefore, these mutations do not negatively affect growth.

Entry into the stationary phase was also analysed for the eight mutants. Only the *deoB* and *orfP* mutants showed a delayed entry into the stationary phase ($OD_{600} = 1.94 \pm 0.07$, 1.95 ± 0.06 , respectively), compared with the reference strain ($OD_{600} = 1.66 \pm 0.11$). Under the conditions used, entry into the stationary phase was caused by culture medium acidification. In order to check whether such a delay could be correlated with impairment of lactic acid production, the pH decrease in the TPPY medium was monitored during growth of both *deoB* and *orfP* mutants. Results of this analysis showed no difference, compared with the reference strain. This point suggested a better resistance of *deoB* and *orfP* mutants to acid stress and was supported by the pH values observed during entry into the stationary phase (pH 4.98 for the reference strain, pH 4.78 for the *deoB* mutant, pH 4.82 for the *orfP* mutant).

Viability during the stationary phase in TPPY medium was estimated for each mutant 2.5 h and 24 h after entry into the stationary phase. The reference strain exhibited $23.7 \pm 3.9\%$ and $0.31 \pm 0.14\%$ survival after 2.5 h and 24 h, respectively. No significant difference could be shown between mutant and reference strains. Thus, *deoB*, *gst*, *rggC*, *osrD*, *osrE*, *osrF*, *osrG* and *orfP* gene integrity was not essential for survival during the stationary phase.

Survival of menadione-resistant mutants with other oxidising agents

The *osr* mutant survival with various oxidising agents was examined. Survival curves of stationary-phase cells exposed to air were compared (Fig. 1). The mutants

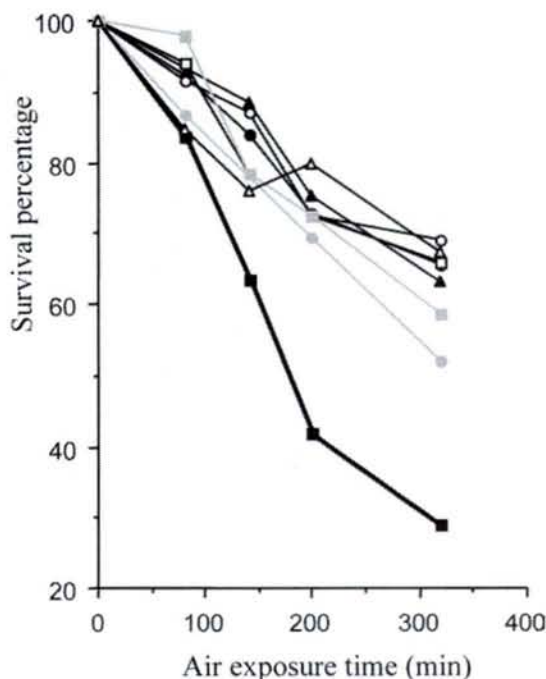
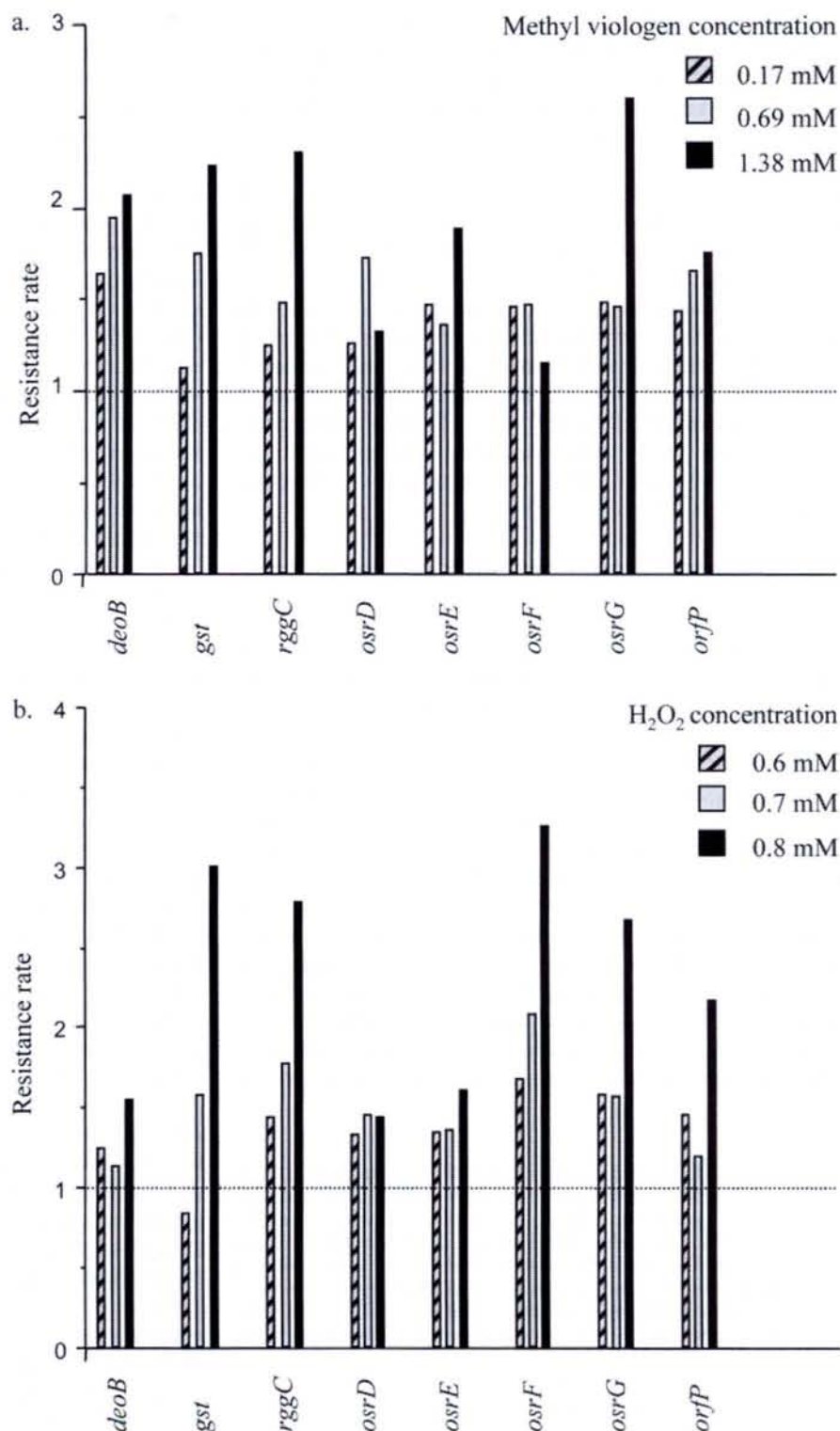


Fig. 1 Survival of *osr* mutant and reference strains to air exposure. During air exposure, all strains are grown in TPPY liquid medium at 42°C under agitation (200 rpm). Each experiment was performed in triplicate and typical results are presented. The *osrG* mutant, whose response was not reproducible, is missing in the graph. Filled squares Reference strain, open circles *deoB* mutant, open triangles *gst* mutant, grey circles *rggC* mutant, grey squares *osrD* mutant, filled triangles *osrE* mutant, filled circles *osrF* mutant, open squares *orfP* mutant

showed an increased survival compared with the reference strain, all of them displaying at least a 2-fold higher resistance after a 320-min air exposure. Since survival rates in the stationary phase without air exposure were similar (see above), the observed difference may be due to air resistance.

The survival of the mutants with various concentrations of methyl viologen (Fig. 2a) and H_2O_2 (Fig. 2b) was also estimated. The resistance phenotype of the *osr* mutants with menadione was extended not only to another superoxide-generating compound (methyl viologen) but also to H_2O_2 . Four of the mutants (*gst*, *rggC*, *osrG*, *orfP*) were markedly more resistant than the reference strain, especially at high doses of both oxidising compounds. Considering the spontaneous or catalysed dismutation of superoxide radicals into H_2O_2 , the observed superoxide resistance of these mutants could be at least partially due to H_2O_2 resistance. Although they also exhibited an increased resistance to both oxidising agents, the *osrD* and *osrE* mutants survived less well, compared with the other strains. In contrast, *deoB* displayed a strong relative resistance to methyl viologen and only a weak resistance to H_2O_2 , whereas *osrF* showed a high resistance to H_2O_2 and only a slightly increased survival with methyl viologen. These data suggest that DeoB and OsrF affect the sensitivity level by two distinct pathways.

Fig. 2 Effects of oxidising agents on the survival of *S. thermophilus* strains. All strains were grown in TPPY liquid medium supplemented with the indicated concentration of **a** methyl-viologen and **b** H₂O₂. The growth ratio was calculated as the ratio of the OD₆₀₀ obtained with and without oxidative stress exposure. To obtain the resistance rate, the growth ratio of each mutant was divided by the growth ratio of the reference strain. Each experiment was done in duplicate and typical results are presented. The reference strain OD₆₀₀ was 0.72 without oxidative stress and was, respectively, 0.42, 0.33, 0.27 with 0.17, 0.69, 1.38 mM methyl viologen for the experiment shown in **a**. The reference strain OD₆₀₀ was 1.10 without oxidative stress and was, respectively, 0.46, 0.24, 0.11 with 0.6, 0.7, 0.8 mM H₂O₂ for the experiment shown in **b**



Expression of *osr* genes under stress conditions

Real-time quantitative RT-PCR was used to analyse changes in the levels of *deoB*, *gst*, *rggC*, *osrD*, *osrE*, *osrF*, *osrG* and *orfP* gene transcripts after exposure to H₂O₂ for 20 min and 45 min in *S. thermophilus* CNRZ368.

The *S. thermophilus* *ldh* gene encoding lactate dehydrogenase was used as a reference gene. To confirm that

ldh was constitutively expressed in the presence of H₂O₂, the amount of *ldh* cDNA was measured in at least two separate RT-PCR experiments with three independent RNA samples from cells treated with H₂O₂ or not treated (control experiment). No significant difference in gene expression level was detected and the *ldh* gene was then considered as a relevant reference. PCR efficiencies were approximately the same for all genes studied,

ranging from 95% to 100% (data not shown). The quantity of cDNA for each gene was normalised to the quantity of *ldh* cDNA present in each RNA preparation and was ranked relative to the quantity of cDNA in the non-stressed conditions.

Transcripts were detected for each of the *osr* genes, demonstrating that all the genes were active. The transcriptional expression of these genes was analysed in the absence or presence of 1 mM H₂O₂ at two different exposure times (20 min, 45 min). During analysis, only 2-fold and better differences were considered as significant. Following 45 min exposure, the *deoB*, *gst*, *osrD*, *osrE* and *osrF* transcript levels were about 5-fold lower than in non-stress conditions and about 2- to 3-fold lower for the *osrP* gene (Table 3). Therefore, these genes were down-regulated during oxidative stress. For the *rggC* and *osrG* genes, the differences were not significant.

Discussion

In this study, we identified eight novel loci involved in *S. thermophilus* behaviour under various oxidative stress conditions, named *osr* (for oxidative stress resistance). In each case, the plasmid disrupted an ORF. Analysis of the region surrounding the *deoB*, *gst*, *osrD*, *osrE*, *osrF* and *osrG* genes suggests that these genes belong in operons. Thus, the resistance phenotype of the corresponding disrupted mutants can be the result of inactivation of the listed genes or/and the result of a polar effect on neighbouring genes. Nevertheless, genes clustered in operons are generally involved in a common process and, from the sequence data available, the operons disrupted in this study do not seem to be an exception. Thus, our mutants should be considered as mutants impaired in a process rather than mutants of one particular ORF. The link between the disrupted loci

Table 3 Relative *osr* gene expression under H₂O₂ stress. The $\Delta\Delta Ct$ method (Livak and Schmittgen 2001) was used to calculate gene expression changes relative to the non-stressed condition, normalised against *ldh*. For each assayed gene, a gene expression level of 100 was assigned for the non-stressed condition. Gene expression in stress conditions was given as a fraction of the expression level in the non-stressed condition. The amount of each cDNA was measured in two separate RT-PCR experiments with independent RNA samples

Gene name	Relative gene expression		
	0 min	20 min	45 min
<i>deoB</i>	100	13.6 ± 4.5	16.2 ± 3.3
<i>gst</i>	100	34.0 ± 17.3	19.8 ± 1.2
<i>rggC</i>	100	69.6 ± 22.4	59.1 ± 22.7
<i>osrD</i>	100	17.2 ± 8.9	13.8 ± 2.3
<i>osrE</i>	100	30.1 ± 17.6	6.4 ± 1.5
<i>osrF</i>	100	19.6 ± 12.8	13.6 ± 0.1
<i>osrG</i>	100	55.8 ± 44.0	58.8 ± 64.2
<i>orfP</i>	100	63.8 ± 10.8	35.8 ± 11.7

and behaviour under oxidative stress conditions is discussed below.

deoB (encoding a phosphopentomutase) was previously shown in *L. lactis* to be involved in defence against various stress conditions, e.g. UV (Duwat et al. 1997), acid and thermal (Rallu et al. 2000). Moreover, disruption of *deoD*, potentially co-transcribed with *deoB*, confers increased heat tolerance in *S. thermophilus* (Varcamonti et al. 2003). A common feature of *deoB* and *deoD* mutations is the modulation of the guanine nucleotide pool involved in the biosynthesis of (p)ppGpp. This molecule, known to be involved in the stringent response, is proposed to play a role in other stress-signal transduction by Duwat et al. (1999) and Rallu et al. (2000). The *deoB* disruption could lead to an increased ppGpp level, as shown for a *S. thermophilus* *deoD* mutant (Varcamonti et al. 2003). Such a ppGpp increase would induce a constitutive pre-adapted state, which could explain stress resistance.

In *S. thermophilus* CNRZ368, disruption of *gst* led to an oxidative stress-resistant phenotype and transcription of this gene was down-regulated under oxidative stress. The involvement of glutamate in stress response has already been reported in *L. lactis* (Rallu et al. 2000). Moreover, in *S. mutans*, the transcription of genes involved in glutamate synthesis (aconitase hydratase, isocitrate dehydrogenase, citrate synthase) is repressed under acid and osmotic stresses (Chia et al. 2001). It is possible that glutamate acts as a stress sensor, a reduction of its concentration inducing stress resistance, maybe via the establishment of a stringent response. Furthermore, the gene downstream of *gst*, *orfX*, belongs to the RimK family whose *E. coli* member adds glutamate residues to maturing proteins. The *E. coli* *rimK* gene has already been linked to oxidative stress defence in different studies (Kang et al. 1989; Zheng et al. 2001), although its action in stress response is still not determined.

The *rggC* gene, encoding a Rgg-like transcriptional regulator, contains a frameshift mutation. The production of a full-length RggC protein could arise by a recoding event, requiring either a ribosome or RNA polymerase slippage to an alternative frame in a region composed of nine thymines, as observed in the *Thermus thermophilus* *dnaX* gene (Larsen et al. 2000). Thus, disruption of the *rggC*₂ ORF could abolish RggC full-protein production. The Rgg-like protein family consists of activators and/or repressors that are, so far, confined to Gram-positive bacteria (Chaussee et al. 2001; Rawlinson et al. 2002) and are involved in the regulation of various functions (Sulavik and Clewell 1996; Chaussee et al. 1999). However, this study involves for the first time a Rgg-like protein in oxidative stress response. RggC probably acts in regulating, and more probably in repressing, the expression of oxidative-stress defence genes.

The last five genes are of unknown function. An oxidative stress-resistant phenotype can now be linked to the disruption of these genes. Moreover, the recognition

of specific domains allows the formulation of some hypotheses regarding their function and localisation. The OsrD protein sequence contains an iron-dependent peroxidase domain. The neighbouring gene encodes a predicted lipoprotein involved in iron transport necessary for the peroxidase activity. In the mutant disrupted in this locus, the inactivation of the putative peroxidase could result in an increased level of reactive oxygen species, which could lead to a constitutive pre-adapted state. Such a state would improve survival under oxidative stress conditions. The *osrD* phenotype could also be explained by a depletion of intracellular iron concentration, reducing the catalysis of the Fenton reaction (and consequently the reactive oxygen species production) and thus increasing cell survival.

Determination of the expression level of the studied genes under oxidative stress conditions indicated that most of them (*deoB*, *gst*, *osrD*, *osrE*, *osrF*, *orfP*) are down-regulated. The ratios in gene expression in the absence and presence of H₂O₂ range from 2 to 15. Small variation in gene expression seems to be a common feature in streptococci (Smoot et al. 2001; Voyich et al. 2003), suggesting a complex level of regulation in these organisms. Moreover, it would appear that *S. thermophilus* modulates gene expression in response to the cell oxidation state, probably thanks to proteins capable of sensing and responding to such signals. The *osr* genes could form part of an oxidative-stress response regulation, which regulator(s) might be a PerR homologue (King et al. 2000; Ricci et al. 2002), RggC or another unidentified regulator.

Acknowledgements A.F. and A.T. were supported by grants from the Ministère de la recherche. F.B. was supported by a grant from the Institut National de la Recherche Agronomique. We are grateful to Paul Hoskisson for his advice regarding the English formulation of the manuscript. Sequence data for *S. thermophilus* LMG18311 were obtained from the UCL Life Sciences Institute (ISV) website at <http://www.biol.ucl.ac.be/gene/genome/>. Sequencing of *S. thermophilus* was supported by the Walloon Region (BIOVAL grant 9813866).

References

- Berks BC, Sargent F, Palmer T (2000) The Tat protein export pathway. *Mol Microbiol* 35:260–274
- Bracquart P (1981) An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yogurt. *J Appl Bacteriol* 51:303–305
- Burris V, Pavlovic G, Decaris B, Guedon G (2002) The ICESt1 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
- Chaussee MS, Ajdic D, Ferretti JJ (1999) The *rgg* gene of *Streptococcus pyogenes* NZ131 positively influences extracellular SPE B production. *Infect Immun* 67:1715–1722
- Chaussee MS, Watson RO, Smoot JC, Musser JM (2001) Identification of Rgg-regulated exoproteins of *Streptococcus pyogenes*. *Infect Immun* 69:822–831
- Chia JS, Lee YY, Huang PT, Chen JY (2001) Identification of stress-responsive genes in *Streptococcus mutans* by differential display reverse transcription-PCR. *Infect Immun* 69:2493–2501
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159
- Duwat P, Cochu A, Ehrlich SD, Gruss A (1997) Characterization of *Lactococcus lactis* UV-sensitive mutants obtained by *ISS1* transposition. *J Bacteriol* 179:4473–4479
- Duwat P, Ehrlich SD, Gruss A (1999) Effect of metabolic flux on stress response pathways in *Lactococcus lactis*. *Mol Microbiol* 31:845–858
- Fridovich I (1997) Superoxide anion radical (O₂⁻), superoxide dismutases, and related matters. *J Biol Chem* 272:18515–18517
- Gibson CM, Mallett TC, Claiborne A, Caparon MG (2000) Contribution of NADH oxidase to aerobic metabolism of *Streptococcus pyogenes*. *J Bacteriol* 182:448–455
- Higuchi M, Yamamoto Y, Poole LB, Shimada M, Sato Y, Takahashi N, Kamio Y (1999) Functions of two types of NADH oxidases in energy metabolism and oxidative stress of *Streptococcus mutans*. *J Bacteriol* 181:5940–5947
- Hoskins J, Alborn WE Jr, Arnold J, Blaszcak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu DJ, Fuller W, Geringer C, Gilmour R, Glass JS, Khoja H, Kraft AR, Lagace RE, LeBlanc DJ, Lee LN, Lefkowitz EJ, Lu J, Matsushima P, McAhren SM, McHenney M, McLeaster K, Mundy CW, Nicas TI, Norris FH, O'Gara M, Peery RB, Robertson GT, Rockey P, Sun PM, Winkler ME, Yang Y, Young-Bellido M, Zhao G, Zook CA, Baltz RH, Jaskunas SR, Rosteck PR Jr, Skatrud PL, Glass JI (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* 183:5709–5717
- Jakubovics N, Smith A, Jenkinson H (2002) Oxidative stress tolerance is manganese (Mn²⁺) regulated in *Streptococcus gordonii*. *Microbiology* 148:3255–3263
- Kang WK, Icho T, Isono S, Kitakawa M, Isono K (1989) Characterization of the gene *rimK* responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in *Escherichia coli* K12. *Mol Genet* 217:281–288
- King KY, Horenstein JA, Caparon MG (2000) Aerotolerance and peroxide resistance in peroxidase and PerR mutants of *Streptococcus pyogenes*. *J Bacteriol* 182:5290–5299
- Koonin EV, Aravind L, Galperin YA (2000) A comparative-genomic view of the microbial stress response. In: Storz G, Henge-Aronis R (eds) *Bacterial stress response*. American Society of Microbiology, Washington, D.C.
- Larsen B, Wills NM, Nelson C, Atkins JF, Gesteland RF (2000) Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc Natl Acad Sci USA* 97:1683–1688
- Leenhouts K (1995) Integration strategies and vectors. *Dev Biol Stand* 85:523–530
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC(T)} method. *Methods* 25:402–408
- Maguin E, Prevost H, Ehrlich SD, Gruss A (1996) Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J Bacteriol* 178:931–935
- Niven DF, Ekins A (2001) Iron content of *Streptococcus suis* and evidence for a *dpr* homologue. *Can J Microbiol* 47:412–416
- Nygaard P (1993) Purine and pyrimidine salvage pathways. In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
- Pericone CD, Park S, Imlay JA, Weiser JN (2003) Factors contributing to hydrogen peroxide resistance in *Streptococcus pneumoniae* include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the fenton reaction. *J Bacteriol* 185:6815–6825
- Poyart C, Pellegrini E, Gaillot O, Boumaila C, Baptista M, Trieu-Cuot P (2001) Contribution of Mn-cofactored superoxide dismutase (SodA) to the virulence of *Streptococcus agalactiae*. *Infect Immun* 69:5098–5106
- Rallu F, Gruss A, Ehrlich SD, Maguin E (2000) Acid- and multi-stress-resistant mutants of *Lactococcus lactis*: identification of intracellular stress signals. *Mol Microbiol* 35:517–528

- Rawlinson EL, Nes IF, Skaugen M (2002) LasX, a transcriptional regulator of the lactocin S biosynthetic genes in *Lactobacillus sakei* L45, acts both as an activator and a repressor. *Biochimie* 84:559–567
- Ricci S, Januleczyk R, Bjorck L (2002) The regulator PerR is involved in oxidative stress response and iron homeostasis and is necessary for full virulence of *Streptococcus pyogenes*. *Infect Immun* 70:4968–4976
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Schell MA (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* 47:597–626
- Smoot LM, Smoot JC, Graham MR, Somerville GA, Sturdevant DE, Migliaccio CA, Sylva GL, Musser JM (2001) Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. *Proc Natl Acad Sci USA* 98:10416–10421
- Sonnhammer ELL, Von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow J, Littlejohn T, Major F, Sankoff D, R L, Sensen C (eds) *Proceedings of the sixth international conference on intelligent systems for molecular microbiology*. AAAI, Menlo Park, pp 175–182
- Spatafora G, Van Hoesen N, Wagner K, Fives-Taylor P (2002) Evidence that ORF3 at the *Streptococcus parasanguis fimA* locus encodes a thiol-specific antioxidant. *Microbiology* 148:755–762
- Storck T, Schulte S, Hofmann K, Stoffel W (1992) Structure, expression, and functional analysis of a Na⁺-dependent glutamate/aspartate transporter from rat brain. *Proc Natl Acad Sci USA* 89:10955–10959
- Sugano Y, Nakano R, Sasaki K, Shoda M (2000) Efficient heterologous expression in *Aspergillus oryzae* of a unique dye-decolorizing peroxidase, DyP, of *Geotrichum candidum* Dec 1. *Appl Environ Microbiol* 66:1754–1758
- Sulavik MC, Clewell DB (1996) Rgg is a positive transcriptional regulator of the *Streptococcus gordonii* *gtfG* gene. *J Bacteriol* 178:5826–5830
- Terzaghi B, Sandine W (1975) Improved medium for lactic streptococci and their bacteriophages. *Appl Environ Microbiol* 29:807–813
- Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald LA, Feldblyum TV, Angiuoli S, Dickinson T, Hickey EK, Holt IE, Loftus BJ, Yang F, Smith HO, Venter JC, Dougherty BA, Morrison DA, Hollingshead SK, Fraser CM (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 293:498–506
- 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
- Thibessard A, Borges F, Fernandez A, Gintz B, Decaris B, Leblond-Bourget N (2004) Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. *Appl Environ Microbiol* 70:2220–2229
- Tusnady GE, Simon I (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* 283:489–506
- Varcamonti M, Graziano MR, Pezzopane R, Naclerio G, Arsenijevic S, De Felice M (2003) Impaired temperature stress response of a *Streptococcus thermophilus deoD* mutant. *Appl Environ Microbiol* 69:1287–1289
- Voyich JM, Sturdevant DE, Braughton KR, Kobayashi SD, Lei B, Virtaneva K, Dorward DW, Musser JM, DeLeo FR (2003) Genome-wide protective response used by group A *Streptococcus* to evade destruction by human polymorphonuclear leukocytes. *Proc Natl Acad Sci USA* 100:1996–2001
- Yamamoto Y, Higuchi M, Poole LB, Kamio Y (2000) Role of the *dpr* product in oxygen tolerance in *Streptococcus mutans*. *J Bacteriol* 182:3740–3747
- Yesilkaya H, Kadioglu A, Gingles N, Alexander JE, Mitchell TJ, Andrew PW (2000) Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*. *Infect Immun* 68:2819–2826
- Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G (2001) DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J Bacteriol* 183:4562–4570

Characterisation of *Streptococcus thermophilus* CNRZ368 oxidative stress-resistant mutants: involvement of a potential Rgg-like transcriptional regulator

Annabelle FERNANDEZ, Frédéric BORGES, Annabelle THIBESSARD,
Brigitte GINTZ, Bernard DECARIS, Nathalie LEBLOND-BOURGET*

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences
et Techniques de l'Université Henri Poincaré, Nancy 1, BP239, 54506 Vandœuvre-lès-Nancy, France

Published online 12 December 2003

Abstract – Eight mutants of *Streptococcus thermophilus* CNRZ368 presenting a menadione-resistant phenotype were selected and the locus mutated in each mutant was identified. Among these clones, 5 were disrupted in a gene of unknown function, 2 were impaired in genes involved in cellular metabolism and the last one (the mutant 15H7) was disrupted in *rgg* encoding a putative transcriptional regulator. To determine if *rgg* could be a key regulator of the superoxide defence of *S. thermophilus*, the 15H7 mutant was further characterised. The results from our work indicate that the *rgg* gene, mutated in the 15H7 genome, is a pseudogene composed of 2 ORFs (*rggA* and *rggB*) that are potentially translated in two different frames. Moreover, transcription analysis demonstrated the existence of a transcript containing *rggA*, *rggB* and *orf2*. Construction of Δ *rggA*, Δ *rggB* and *orf2* mutants and their phenotypic analysis confirmed the involvement of *rggA* and *rggB* in the oxidative stress response. The question of the role of the pseudogene *rgg* is still open.

***Streptococcus thermophilus* / oxidative stress / Rgg-like transcriptional regulator**

Résumé – Caractérisation de mutants résistants au stress oxydant chez *Streptococcus thermophilus* CNRZ368 : implication d'un régulateur transcriptionnel putatif de type Rgg. Huit mutants de *Streptococcus thermophilus* CNRZ368, présentant un phénotype de résistance à la ménadione ont été sélectionnés et le locus muté a été identifié pour chacun d'eux. Parmi ces clones, 5 sont interrompus dans un gène de fonction inconnue, 2 sont mutés dans des gènes impliqués dans le métabolisme cellulaire et un autre (15H7) est interrompu dans le gène *rgg* codant potentiellement un régulateur transcriptionnel. Afin de déterminer si *rgg* est un régulateur majeur de la défense contre le stress superoxyde le mutant 15H7 a été plus amplement étudié. Les résultats de notre travail indiquent que le gène *rgg*, muté dans le génome de 15H7, est un pseudogène composé de deux ORF (*rggA* and *rggB*) potentiellement traduites dans deux phases de lecture différentes. De plus, l'analyse transcriptionnelle démontre l'existence d'un transcrit englobant *rggA*, *rggB* et *orf2*. La construction des mutants Δ *rggA*, Δ *rggB* et *orf2* ainsi que l'analyse de leur phénotype confirme l'implication de *rggA* et *rggB* dans la réponse contre le stress oxydant. La question concernant le rôle du pseudogène *rgg* est toujours ouverte.

***Streptococcus thermophilus* / stress oxydant / régulateur transcriptionnel de type Rgg**

* Corresponding author: bourget@nancy.inra.fr

1. INTRODUCTION

S. thermophilus is a Gram-positive bacterium with low G+C content that belongs to the group of lactic acid bacteria (LAB). Due to their ability to metabolise sugar in lactic acid, these microorganisms are commonly used in dairy factories as starters of fermentation of yoghurt and are involved in the process of ripening of cheeses.

S. thermophilus CNRZ368 is considered to be an anaerobic aerotolerant microorganism. Although this bacterium grows more slowly in the presence of oxygen (growth rate = 40 min) than in its absence (growth rate = 27 min), it can survive in the presence of oxygen [20]. The ability of *S. thermophilus* to survive in the presence of oxygen presumably results from the existence of a defence mechanism that helps cells to eliminate reactive oxygen species (ROS) and to repair damage [21]. Data concerning the ability of lactic acid bacteria to survive in the presence of oxidative stress are scarce. In the literature, the existence of antioxidant enzymes is reported in *Streptococcus* genera [1, 7, 9, 14, 25]. Moreover, the existence of PerR, an oxidative stress-responsive repressor is reported in *Streptococcus pyogenes* [9, 15]; nevertheless, the global network of oxidative stress regulation in streptococci remains to be elucidated.

On the contrary, the oxidative stress response of aerobic organisms such as *E. coli* is well documented. This response necessitates antioxidant enzymes involved in the detoxification of cells and repair of enzymes necessary to eliminate damage. In *E. coli*, two networks of proteins can be induced depending on the reactive oxygen species: superoxide radicals induce the SoxR/S regulon, whereas the OxyR/S regulon is induced in the presence of hydrogen peroxide (for a review see [17]). These two regulons are partially overlapping. Although the presence of genes homologous to *oxyR* and *soxR* is a characteristic shared by many bacteria, other regulatory proteins have also emerged through evolution. Among them are PerR [15], sigma(B) and OhrR [8] and σ^R /RsrA that are

involved in the response against H₂O₂ and superoxide radicals, respectively [12, 13].

To identify the genes involved in the oxidative stress response of *S. thermophilus* CNRZ368, a collection of mutants was generated and screened for the menadione-resistant phenotype. This work allowed the identification of eight loci potentially involved in the mechanism of superoxide tolerance: among them is the locus *rgg*, described in the literature as a transcriptional regulator [2, 3, 18, 19].

2. MATERIALS AND METHODS

2.1. Bacterial strains and growth conditions

S. thermophilus CNRZ368 and *E. coli* TG1 were purchased from the INRA-CNRZ strain collection (Jouy-en-Josas, France) and from Stratagene (Amsterdam, The Netherlands), respectively. The thermosensitive plasmids pGhost9 and pGhost9:ISS1 were used [11]. They carry an erythromycin resistance gene and a thermosensitive origin of replication. Depending on the experiment, *S. thermophilus* CNRZ368 was cultivated in TPTY or M17 media. For selection of menadione-resistant clones, plates containing TPTY medium supplemented with menadione (concentration going from 1 to 120 $\mu\text{g}\cdot\text{mL}^{-1}$) were used. *E. coli* TG1 cultures were grown on Luria-Bertani medium.

2.2. Cloning of the fragments flanking the chromosomal insertion site

Cloning of the fragments flanking the pGhost9:ISS1 insertion locus was done as previously described by Thibessard et al. [23] and allowed the identification of the mutagenised site of each mutant.

2.3. General DNA techniques, DNA sequencing and analysis

Isolation of genomic or plasmidic DNA, agarose gel electrophoresis, restriction enzyme digests, Southern hybridisation,

PCR, DNA ligation and *E. coli* transformation were carried out as described in [16]. Sequencing was done as previously described in [22].

2.4. Construction of *S. thermophilus* mutants

S. thermophilus mutants were constructed via allelic replacement. The $\Delta rggA$ and $\Delta rggB$ mutants presented a chromosomal deletion of the residues 14 to 41 of RggA and of the residues 106 to 207 of RggB, respectively. The mutant *orf2* presented 2 stop mutations affecting the first two residues of the protein.

2.5. RNA preparation, RT-PCR and Northern analysis

Total RNA was extracted from *S. thermophilus* with a single guanidinium thiocyanate step and phenol-chloroform extraction by using Tri-Reagent (Sigma, Saint-Quentin-Fallavier, France). In the RNA preparation, possible contaminating DNA was eliminated by the action of DNaseI. cDNA was generated from 2.5 μ g of RNA reverse transcribed with MMLV-reverse transcriptase and hexamer oligonucleotide random primers (Amersham Pharmacia Biotech, Freiburg, Germany). For Northern hybridisation analysis, template RNA samples (30 μ g) were electrophoresed into a denaturing (i.e. formaldehyde) agarose gel and transferred to a HybondN⁺ (Amersham Pharmacia Biotech) according to the manufacturer's instructions. Digoxigenin-labelled transcripts complementary to *rggB* were used as a probe. These *rggB* transcripts were synthesised in vitro using the T3 promoter RNA polymerase system carried by the pBKS-*rggB* vector. The latter was constructed by cloning a 533 bp internal fragment of *rggB* in the pBKS vector. Hybridisation and detection procedures were done according to the manufacturer's instructions (Roche Molecular Biochemicals, Meylan, France).

2.6. Sensitivity to oxidising agent

The sensitivity of CNRZ368 derivatives to oxidative stress was checked on exponentially growing cells ($OD_{600} = 0.6$) by the addition of menadione (1 to 20 mg·mL⁻¹). The cells were left in contact with menadione for 3 h. Aliquots of the culture were diluted in TPPY, and appropriate dilutions plated on TPPY agar plates to determine cell viability. The plates were stocked in jars containing a GENbox anaer system (bioMérieux, Marcy l'Étoile, France) providing hypoxic conditions, and colonies were counted after 20 h of incubation. The percentage of surviving CFU was determined by comparing viability with or without treatment. All experiments were performed in triplicate.

3. RESULTS

3.1. Identification of genes disrupted in *S. thermophilus* menadione-resistant mutants

To identify the genes implicated in the oxidative stress response of *S. thermophilus*, a collection of mutants, constructed by insertional mutagenesis using pGhost9:ISS1, was screened for the tolerance of the clones to menadione (a superoxide radical generating molecule). Among the 2112 clones, eight were selected for their increased resistance to menadione and were further characterised. For each mutant, the disrupted locus was cloned and sequenced. Sequence analysis allowed the grouping of these 8 mutants into 3 classes (Fig. 1). The first class assembled the mutants 7H3, 13B12, 14H11, 15H8 and 11C1, that were disrupted in genes of unknown function. For the last clone 11C1, the disrupted gene was specific to the genome of *S. thermophilus*. The second class brought together the mutants 6H7 and 16C10, that were mutagenised in genes involved in cellular metabolism. The mutant 6H7 was impaired in a gene encoding DeoB, a protein

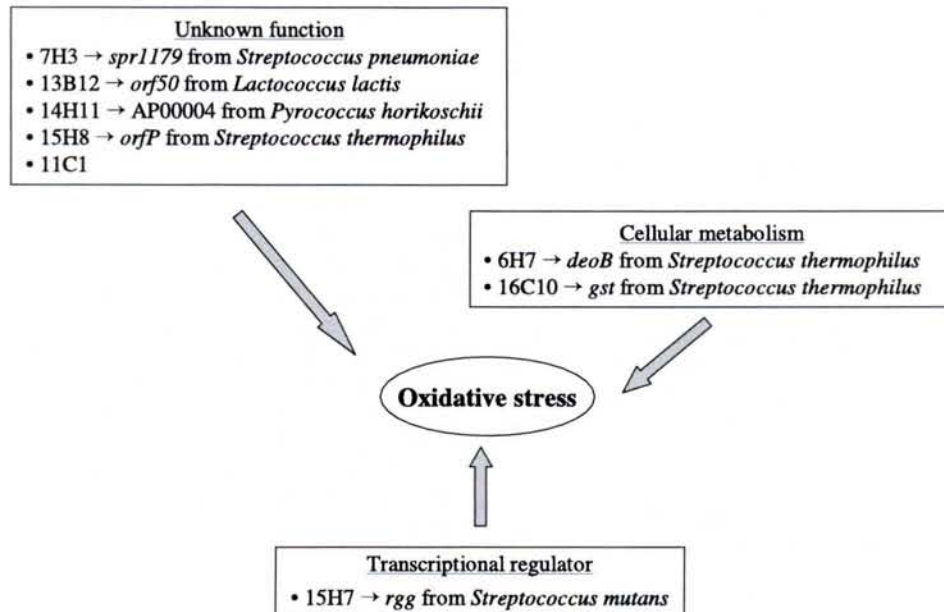


Figure 1. Classification of the menadione-resistant mutants according to their putative function. The best homologous gene for ORF disrupted in CNRZ368 mutants is indicated for each clone after the arrows.

involved in purine metabolism. The mutant 16C10 was affected in *gst*, that encodes a putative proton-glutamate symport protein.

The third class included the mutant 15H7, that was disrupted in a gene, homologous to a putative transcriptional regulator of *Streptococcus mutans*, encoding a protein belonging to the Rgg-like family. To determine the role of the *rgg* locus in menadione tolerance and to understand the mechanism of regulation of *S. thermophilus* oxidative stress response, we decided to focus our research on the mutant 15H7.

3.2. Characterisation of the locus disrupted in the mutant 15H7

3.2.1. Identification of the *rgg* locus

Figure 2A shows a scheme of the three ORFs, named *rggA*, *rggB* and *orf2*, identified within the nucleotide sequence flank-

ing the pGhost9:ISS1 insertion locus of the genome of 15H7. The *rggA* ORF potentially encoded a protein of 65 residues that showed 77% identity with the putative transcriptional regulator Rgg of *S. mutans*. The 232 amino acid residues potentially encoded by *rggB* presented 56% identity with the same putative transcriptional regulator Rgg of *S. mutans*. Thus, the homology of the Rgg protein from *S. mutans* began with the protein RggA and continued with the protein RggB of *S. thermophilus* in the frame +1. These data indicate that the *rgg* gene disrupted in the mutant 15H7 is a pseudogene. Finally, *orf2* potentially encoded a protein of 35 residues presenting no homology with sequences present in the databases.

The search for transcriptional signals allowed the identification of a potential promoter composed of an extended -10 sequence located 61 pb upstream of the start codon of *rggA*. Moreover, a putative

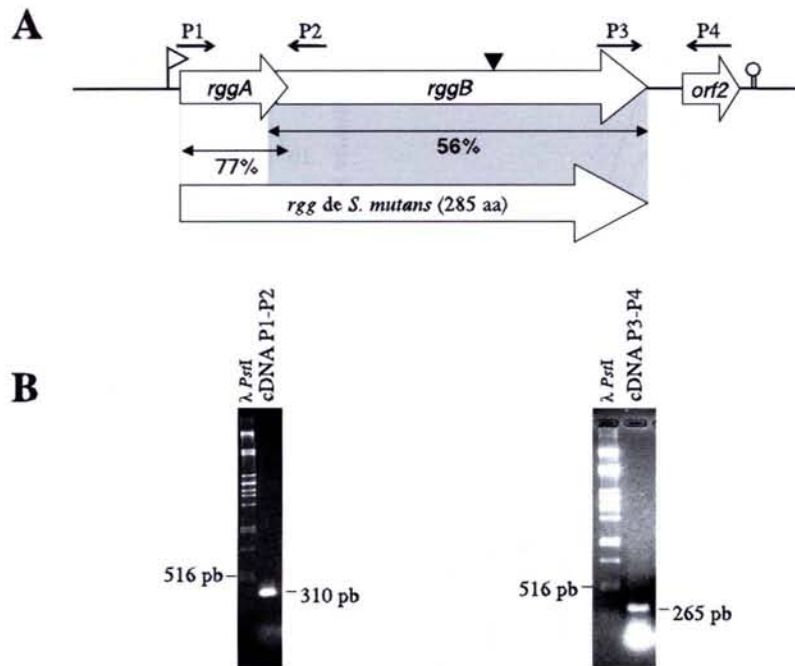


Figure 2. Organisation of the *rgg* locus of *S. thermophilus* CNRZ368. (A) The solid line symbolises the *S. thermophilus* genomic DNA. ORFs are indicated as open arrows. The black arrowhead indicates the insertion site of pGh9:ISS1 within the genome of the 15H7 mutant. The open flag indicates the putative promoter and the putative rho-independent terminator is represented as a hairpin loop. Grey boxes symbolise the homologous region between *rgg* from *S. thermophilus* and from *S. mutans*: the percentage of amino acid identity is indicated below the double arrows. Primers used for RT-PCR experiments are indicated. The expected size of the P1-P2 and P3-P4 PCR fragments is 310 pb and 265 pb, respectively. (B) Gel electrophoresis showing the RT-PCR products obtained using the primers P1-P2 and P3-P4.

rho-independent terminator consisting of a 10 pb stem and a 3 pb loop was seen downstream of the stop codon of *orf2*. Located 8 pb upstream of *rggA* was a potential Shine-Dalgarno sequence, suggesting that this gene is likely to be translated. On the contrary, no Shine-Dalgarno sequence preceded *rggB*, thus rendering improbable the synthesis of RggB alone.

3.2.2. Transcriptional organisation of the *rgg* locus

Northern blot analysis performed with a RNA probe complementary to *rggB* revealed a strong band of approximately

1200 nucleotides (data not shown). To determine which ORFs were included in this transcript, RT-PCR analyses were carried out. cDNA generated from *S. thermophilus* CNRZ368 RNA were used in PCR experiments (Fig. 2B). Primers P1 and P2 were designed to verify the presence of a transcript going from *rggA* to *rggB* and P3 and P4 to determine the existence of a transcript going from *rggB* to *orf2*. In both cases, a PCR product of the expected size (310 pb and 265 pb, respectively) was amplified. These data suggested the existence of one transcript including *rggA*, *rggB* and *orf2*. Thus, the *rgg* locus showed an operonic structure.

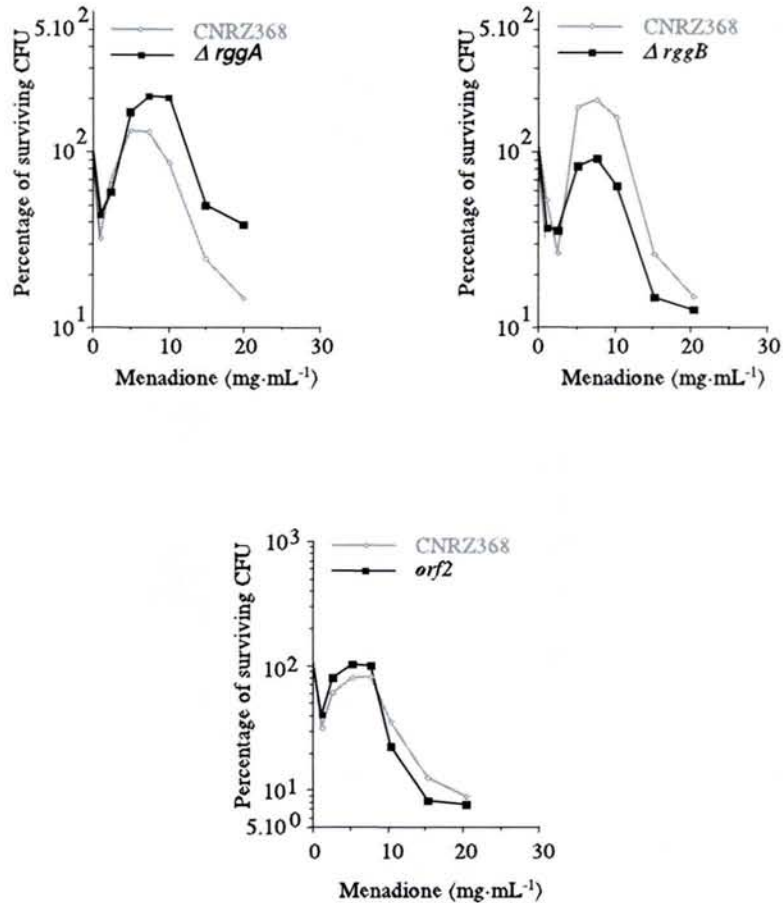


Figure 3. Survival of the $\Delta rggA$, $\Delta rggB$ and *orf2* mutants in the presence of increasing concentrations of menadione. Cells grown overnight in milk medium were diluted a hundred-fold into fresh TPPY medium. At $OD_{600} = 0.6$, samples of the culture were treated with menadione for 3 h. Cells were diluted and plated on TPPY solid medium and then incubated for 20 h in jars. Colony forming units (CFUs) were then counted and the percentages of surviving CFUs were estimated relative to non-treated samples. All these experiments were done in triplicate. Results of typical experiments are presented here.

3.2.3. Involvement of the *rgg* locus in *S. thermophilus* oxidative stress response

The results from our analysis indicated that impairment of a pseudogene, composed of *rggA* and *rggB*, resulted in an increased tolerance of *S. thermophilus* to menadione. To gain insight into the involve-

ment of the *rgg* locus in the oxidative stress defence, a series of mutants ($\Delta rggA$, $\Delta rggB$ and *orf2*) was constructed via allelic replacement. The $\Delta rggA$ and $\Delta rggB$ mutants presented a deletion of 43% and of 44% of the total protein, respectively. The mutant *orf2* carried 2 stop mutations.

The ability of these 3 mutants to survive in the presence of increasing concentrations

of menadione was tested and compared with that of the wild-type strain. As shown in Figure 3, the $\Delta rggA$ mutant showed an increased ability to survive in the presence of menadione compared with the wild-type whereas the ability of the $\Delta rggB$ clone to tolerate menadione was reduced. Moreover, no difference in behaviour was seen for the mutant *orf2* and the wild-type strain. These data suggest that *rggA* and *rggB* are both involved in *S. thermophilus* response to oxidative stress. However, their exact role in this mechanism remains to be solved. In contrast, *orf2* does not seem to play a role in the survival of *S. thermophilus* submitted to oxidative stress.

4. DISCUSSION

The dairy organism, *S. thermophilus*, is continuously exposed to stress conditions generated during industrial processes. To identify the mechanisms that confer tolerance to ROS, we isolated menadione-resistant strains by insertional mutagenesis. This study reports the characterisation of 8 genes involved in *S. thermophilus* CNRZ368 oxidative stress defence. According to their putative function, these genes were classified into 3 groups. The first class, which contains 5/8 genes, brings together genes of unknown function. Thus, the strategy used in this work allowed the identification of new genes involved in oxidative stress: a phenotype of superoxide resistance can now be assigned to their disruption. The second class groups genes implicated in cellular metabolism: *deoB* encodes a phosphopentomutase that is involved in the purine salvage pathway, whereas *gst* is a potential glutamate-proton transporter. Both pathways are known to be connected with stress responses. In *Lactococcus lactis*, the *deoB* mutation confers multiple stress resistance on the cell [5, 6]. In *S. mutans*, genes encoding enzymes responsible for glutamate synthesis are repressed in acid and thermal stress [4]. All these results reveal that cellular metabolic pathways are

intimately related to stress response and that the flux of particular metabolites are likely to be implicated in stress response in LAB. The last class contains a transcriptional regulator that belongs to the family of Rgg-like proteins. These are DNA-binding proteins including a helix-turn-helix (HTH) motif at their N-terminal end. Rgg-like proteins are only found in Gram-positive bacteria belonging to the *Streptococcus*, *Lactococcus*, *Lactobacillus* and *Listeria* genera. These regulators are reported to stimulate positively or negatively the transcription of numerous genes of different functions [3].

Our work describes for the first time the involvement of a Rgg-like protein in oxidative stress response. Moreover, it asks the question of how the pseudogene *rgg*, disrupted in the mutant 15H7, can influence the tolerance of ROS. To address the question of whether RggA or RggB or both were involved in stress response: both $\Delta rggA$ and $\Delta rggB$ mutants were constructed and their ability to respond to menadione was checked. The results of these analyses indicated that both RggA and RggB are implicated in oxidative stress defence. Taking into account the lack of visible RBS sites downstream of *rggB*, the independent synthesis of RggB seems improbable. An alternative hypothesis concerns the synthesis of a RggAB protein that would be generated by frameshifting. Sequence analysis of the region overlapping the *rggA* and *rggB* ORFs revealed the presence of a frameshift mutation within the sequence TTT TTT TTT. One T-deletion would abolish the frameshift mutation and restore the proper *rgg* ORF. In the literature, frameshift mutations (and particularly transcriptional frameshifting) are known to arise in sequences composed of multiple thymine or adenine [10, 24]. Such events can result either from DNA polymerase slippage during replication, or may require RNA polymerase slippage (transcriptional frameshift) or ribosome slippage (translational frameshift). It is now important to know if such a frameshifting

event could appear in *S. thermophilus*, and if so, if it could be stimulated by environmental conditions (for instance, by the presence of oxidative molecules). To fully understand the role of Rgg in the oxidative stress response, the characterisation of the genes regulated by this protein and implicated in stress defence would be essential.

Acknowledgements: This work was supported by "Le Ministère de la Recherche et de la Technologie" and by "l'Institut National de la Recherche Agronomique".

REFERENCES

- [1] Chang S.K., Hassan H.M., Characterization of superoxide dismutase in *Streptococcus thermophilus*, Appl. Environ. Microbiol. 63 (1997) 3732–3735.
- [2] Chaussee M.S., Watson R.O., Smoot J.C., Musser J.M., Identification of Rgg-regulated exoproteins of *Streptococcus pyogenes*, Infect. Immun. 69 (2001) 822–831.
- [3] Chaussee M.S., Sylva G.L., Sturdevant D.E., Smoot L.M., Graham M.R., Watson R.O., Musser J.M., Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*, Infect. Immun. 70 (2002) 762–770.
- [4] Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y., Identification of stress-responsive genes in *Streptococcus mutans* by differential display reverse transcription-PCR, Infect. Immun. 69 (2001) 2493–2501.
- [5] Duwat P., Ehrlich S.D., Gruss A., Effects of metabolic flux on stress response pathways in *Lactococcus lactis*, Mol. Microbiol. 31 (1999) 845–858.
- [6] Duwat P., Cesselin B., Sourice S., Gruss A., *Lactococcus lactis*, a bacterial model for stress responses and survival, Int. J. Food Microbiol. 55 (2000) 83–86.
- [7] Gibson C.M., Mallett T.C., Claiborne A., Caparon M.G., Contribution of NADH oxidase to aerobic metabolism of *Streptococcus pyogenes*, J. Bacteriol. 182 (2000) 448–455.
- [8] Helmann J.D., Wu M.F., Gaballa A., Kobel P.A., Morshedi M.M., Fawcett P., Paddon C., The global transcriptional response of *Bacillus subtilis* to peroxide stress is coordinated by three transcription factors, J. Bacteriol. 185 (2003) 243–253.
- [9] King K.Y., Horenstein J.A., Caparon M.G., Aerotolerance and peroxide resistance in peroxidase and PerR mutants of *Streptococcus pyogenes*, J. Bacteriol. 182 (2000) 5290–5299.
- [10] Larsen B., Wills N.M., Nelson C., Atkins J.F., Gesteland R.F., Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting, Proc. Natl. Acad. Sci. USA 97 (2000) 1683–1688.
- [11] Maguin E., Prevost H., Ehrlich S.D., Gruss A., Efficient insertional mutagenesis in lactococci and other Gram-positive bacteria, J. Bacteriol. 178 (1996) 931–935.
- [12] Paget M.S., Kang J.G., Roe J.H., Buttner M.J., sigmaR, an RNA polymerase sigma factor that modulates expression of the thio-redoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2), Embo J. 17 (1998) 5776–5782.
- [13] Paget M.S., Molle V., Cohen G., Aharonowitz Y., Buttner M.J., Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon, Mol. Microbiol. 42 (2001) 1007–1020.
- [14] Poole L.B., Higuchi M., Shimada M., Calzi M.L., Kamio Y., *Streptococcus mutans* H₂O₂-forming NADH oxidase is an alkyl hydroperoxide reductase protein, Free Radical Biol. Med. 28 (2000) 108–120.
- [15] Ricci S., Janulczyk R., Bjorck L., The regulator PerR is involved in oxidative stress response and iron homeostasis and is necessary for full virulence of *Streptococcus pyogenes*, Infect. Immun. 70 (2002) 4968–4976.
- [16] Sambrook J., Fritsch E.F., Maniatis T., Molecular cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, New York, USA, 1989.
- [17] Storz G., Zheng M., Oxidative stress, in: Storz G., Henge-Aronis R. (Eds.), Bacterial stress responses, ASM press, Washington D.C., USA, 2000, pp. 47–59.
- [18] Sulavik M.C., Clewell D.B., Rgg is a positive transcriptional regulator of the *Streptococcus gordonii* *gffG* gene, J. Bacteriol. 178 (1996) 5826–5830.
- [19] Sulavik M.C., Tardif G., Clewell D.B., Identification of a gene, *rgg*, which regulates expression of glucosyltransferase and influences the Spp phenotype of *Streptococcus gordonii* Challis, J. Bacteriol. 174 (1992) 3577–3586.
- [20] Thibessard A., Défense de *Streptococcus thermophilus* contre le stress oxydatif : existence d'un système de réponse, construction et sélection de mutants et identification de

- gènes impliqués, Ph.D. thesis, Université Henri Poincaré, Nancy, France, 2002.
- [21] Thibessard A., Leblond-Bourget N., Fernandez A., Gintz B., Decaris B., Response of *Streptococcus thermophilus* CNRZ368 and its colonial variants to oxidative stress: evidence for an inducible defence system, *Lait* 81 (2001) 311–316.
- [22] Thibessard A., Fernandez A., Gintz B., Decaris B., Leblond-Bourget N., The proteins RodA and PBP2b are implicated in the control of the ovoid-shape of *Streptococcus thermophilus* CNRZ368 and play a role in cells defence against superoxide radicals, *Sci. Aliments* 22 (2002) 75–85.
- [23] Thibessard A., Fernandez A., Gintz B., Decaris B., Leblond-Bourget N., Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368, *Can. J. Microbiol.* 48 (2002) 473–478.
- [24] Wagner L.A., Weiss R.B., Driscoll R., Dunn D.S., Gesteland R.F., Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*, *Nucl. Acids Res.* 18 (1990) 3529–3535.
- [25] Yamamoto Y., Higuchi M., Poole L.B., Kamio Y., Role of the *dpr* product in oxygen tolerance in *Streptococcus mutans*, *J. Bacteriol.* 182 (2000) 3740–3747.

Monsieur **BORGES Frédéric**

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY 1
en GENETIQUE MOLECULAIRE

VU, APPROUVÉ ET PERMIS D'IMPRIMER N° 1048

Nancy, le 3 mai 2005

Le Président de l'Université



Résumé

La génération de nouvelles fonctions dans le monde bactérien peut résulter de la formation de protéines par réassortiment de domaines. L'isolement d'un mutant de *Streptococcus thermophilus* CNRZ368 présentant un phénotype de longues chaînes de cellules a permis d'identifier le gène *cse* (pour cell segregation). Ce gène code une protéine extracellulaire dont la structure modulaire est superposable à sa constitution chimérique. Les données acquises suggèrent que Cse aurait été générée par réassortiment de domaines : elle résulterait en effet de la fusion du domaine LysM de la protéine SIP au domaine CHAP de la protéine PcsB. La protéine Cse est fonctionnelle puisqu'elle est impliquée dans la ségrégation cellulaire. En effet, l'absence du gène *cse* entraîne une augmentation du nombre de cellules par chaîne.

A l'inverse de la plupart des exemples décrits dans la littérature, le caractère chimérique est détectable à l'échelle nucléique. En effet, contrairement à la région 5' du gène *cse*, sa région 3' présente de fortes identités (93%) avec l'extrémité 3' du gène *pcsB* du même génome. Par conséquent, il a été possible d'émettre une hypothèse concernant la localisation des bornes du réarrangement à l'origine de la formation du gène *cse*. Il semble que le site de fusion des fragments de gènes soit localisé à l'intérieur d'une région de faible complexité et riche en répétitions, située dans la région centrale du gène *cse*. L'analyse de cette région centrale, appelée *var-cse*, chez différentes souches de *S. thermophilus*, montre qu'elle est variable. Dans le but de déterminer l'influence de la variabilité de cette région sur l'activité de ségrégation cellulaire de la protéine Cse, l'allèle *var-cse* d'une souche de *S. thermophilus* a été remplacé par celui d'une autre souche. Ce remplacement de séquences aurait peu ou pas d'influence sur l'activité de ségrégation cellulaire de la protéine Cse. Ainsi, le site de fusion des deux fragments de gènes formant *cse* serait localisé dans la région *var-cse*, dont la fonction tolère d'importantes modifications de sa séquence.

Nous proposons que la présence, dans les gènes, de régions variables et de faible complexité favorise la génération de nouvelles protéines actives par réassortiment de domaines. Leur présence augmenterait la probabilité pour que les points de fusion des fragments de gènes ne soient pas localisés à l'intérieur des modules, séquences codant les domaines protéiques. Leur existence augmenterait ainsi la probabilité de réassortir des domaines tout en respectant leur intégrité.

Mots Clés : *Streptococcus thermophilus*, ségrégation cellulaire, gène chimérique, création de gènes, variabilité génétique, répétitions