



**HAL**  
open science

# Le gène *cse*, de création récente, code une hydrolase du peptidoglycane impliquée dans la séparation des cellules de *Streptococcus thermophilus*

Séverine Layec

## ► To cite this version:

Séverine Layec. Le gène *cse*, de création récente, code une hydrolase du peptidoglycane impliquée dans la séparation des cellules de *Streptococcus thermophilus*. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université Henri Poincaré - Nancy 1, 2008. Français. NNT : 2008NAN10066 . tel-01748447

**HAL Id: tel-01748447**

**<https://hal.univ-lorraine.fr/tel-01748447>**

Submitted on 29 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## 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](mailto: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.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

**Séverine LAYEC**

**Le gène *cse*, de création récente, code une hydrolase du peptidoglycane impliquée dans la séparation des cellules de *Streptococcus thermophilus*.**

Soutenance le 7 Novembre 2008

**Membres du jury :**

*Président du jury :*

**Mme LETT M.C.**

Professeur, Université Louis Pasteur, Strasbourg

*Rapporteurs :*

**M. KULAKAUSKAS S.**

Chargé de Recherche, INRA, Jouy-en-Josas

**M. MESNAGE S.**

Chargé de Recherche, INSERM, Paris

*Examineurs :*

**M. LEBLOND P.**

Professeur, Nancy-Université

**Mme RUL F.**

Chargée de Recherche, INRA, Jouy-en-Josas (Invitée)

**M. DECARIS B.**

Professeur, Nancy-Université

**Mme LEBLOND-BOURGET N.**

Maître de conférence, Nancy-Université







Thèse présentée pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ, NANCY 1**

en Génétique Moléculaire

par

**Séverine LAYEC**

**Le gène *cse*, de création récente, code une hydrolase du peptidoglycane impliquée dans la séparation des cellules de *Streptococcus thermophilus*.**

Soutenance le 7 Novembre 2008

**Membres du jury :**

*Président du jury :*

**Mme LETT M.C.**

Professeur, Université Louis Pasteur, Strasbourg

*Rapporteurs :*

**M. KULAKAUSKAS S.**

Chargé de Recherche, INRA, Jouy-en-Josas

**M. MESNAGE S.**

Chargé de Recherche, INSERM, Paris

*Examineurs :*

**M. LEBLOND P.**

Professeur, Nancy-Université

**Mme RUL F.**

Chargée de Recherche, INRA, Jouy-en-Josas (Invitée)

**M. DECARIS B.**

Professeur, Nancy-Université

**Mme LEBLOND-BOURGET N.**

Maître de conférence, Nancy-Université





## **Remerciements**

---

Je tiens à remercier le Professeur *Bernard Decaris* et le Professeur *Pierre Leblond* de m'avoir permis de réaliser ces travaux de recherche au sein du laboratoire de Génétique et Microbiologie – UMR INRA/UHP 1128 IFR110 de Nancy Université. Merci au professeur *Bernard Decaris*, mon directeur de thèse, de m'avoir proposé de travailler sur ce sujet de thèse et de m'avoir fait confiance.

Merci à Madame le Docteur *Nathalie Leblond-Bourget*, co-encadrant de ma thèse de m'avoir accordé sa confiance tout au long de cette aventure. Je la remercie également pour l'intérêt qu'elle a accordé avec dynamisme à ce travail et pour son aide précieuse lors de la rédaction de ce manuscrit.

Merci à Monsieur le Docteur *Saulius Kulakauskas* et à Monsieur le Docteur *Stéphane Mesnage* d'avoir accepté de rapporter ce travail. Merci à Madame le Professeur *Marie-Claire Lett* et à Madame le Docteur *Françoise Rul* d'avoir accepté de l'examiner.

Ce projet a bénéficié d'un financement par l'Institut Fédératif de Recherche (IFR 110) dirigé par le professeur *Jean-Pierre Jacquot* pour mener à bien ce travail. Par ailleurs, ce travail n'aurait pas abouti sans les collaborations qui ont été entreprises au cours de ma thèse. Ainsi, je tiens à remercier chacune des personnes qui se sont investies dans ce projet :

- Merci à *Christine Delorme*, du laboratoire de Génétique microbienne de l'INRA au domaine de Vilvert à Jouy-en-Josas d'avoir participé à ce travail par l'étude de la recherche du gène *cse* chez les espèces *Streptococcus salivarius* et *Streptococcus vestibularis*.

- Merci à *Marie-Pierre Chapot-Chartier* et à *Pascal Courtin* du laboratoire de Biochimie Bactérienne de l'INRA au domaine de Vilvert à Jouy-en-Josas d'avoir accepté la recherche du site de clivage du domaine CHAP de la protéine Cse par l'analyse des mucopeptides chez *Bacillus subtilis* et chez *Streptococcus thermophilus*.

- Merci à *Valérie Legué* et à *Joëlle Gérard* du laboratoire de Génomique, Ecophysiologie et Ecologie Fonctionnelles, « Interaction Arbres/Micro-organismes », de l'UMR INRA/UHP 1136, IFR110, pour les observations au microscope confocal à balayage laser et au microscope électronique à transmission.

Je suis reconnaissante envers *Frédéric Borges* qui a initié ce projet et porté un intérêt constant à mes travaux.

Un grand merci à toutes les personnes, qui de loin ou de près ont contribué à la réalisation de ce travail dans les meilleures conditions possibles.



A black and white photograph of a string of light-colored, oval-shaped beads or pearls. The beads are arranged in a slightly curved line, starting from the bottom left and moving towards the top right. The background is a dark, textured surface, possibly a fabric or a wall with a pebbled texture. The lighting is soft, highlighting the smooth, rounded surfaces of the beads.

*Sommaire*



# Sommaire

---

|   |             |
|---|-------------|
| <b>Introduction .....</b>   | <b>-5-</b>  |
| <b>I. Evolution génomique de l'espèce <i>Streptococcus thermophilus</i> .....</b>   | <b>-5-</b>  |
| 1. Le genre <i>Streptococcus</i> .....  | -5-         |
| 2. Adaptation de <i>S. thermophilus</i> dans le milieu lait : impact sur le génome.....   | -6-         |
| 2.1 <i>S. thermophilus</i> , une espèce bactérienne clonale.....  | -7-         |
| 2.2 <i>S. thermophilus</i> a évolué par perte de fonctions.....   | -8-         |
| a. Perte des voies métaboliques au cours de l'évolution .....   | -8-         |
| b. Perte des gènes de virulence.....  | -9-         |
| 2.3 Acquisition de nouvelles fonctions par transfert horizontal.....  | -11-        |
| a. Les éléments génétiques contribuant à l'acquisition de nouveaux gènes dans le génome de <i>S. thermophilus</i> . .....       | -11-        |
| b. Acquisition de nouveaux gènes par transfert horizontal .....   | -12-        |
| 3. Conclusion.....  | -13-        |
| <b>II. La duplication/divergence et le réassortiment de modules, deux mécanismes majeurs de la création de gènes. ....</b>      | <b>-14-</b> |
| 1. Création de nouveaux gènes par duplication et divergence .....   | -14-        |
| 1.1 Duplication de gènes .....  | -14-        |
| 1.2 Devenir des gènes dupliqués .....   | -15-        |
| 2. Création de nouveaux gènes par réassortiment de modules .....  | -16-        |
| 2.1 Création de gènes par recombinaison intramoléculaire.....   | -16-        |
| 2.2 Création de gènes par recombinaison intermoléculaire.....   | -17-        |
| 2.3 Les séquences de faible complexité pourraient favoriser la création de nouveaux gènes par le réassortiment de modules. .... | -18-        |
| 3. La duplication/divergence et le réassortiment de modules, deux processus non exclusifs .                                     | -20-        |
| 4. Le transfert horizontal contribue indirectement à la création de nouveaux gènes.....   | -21-        |
| 4.1 Duplication et transfert horizontal.....  | -21-        |
| 4.2 Réassortiment de modules et transfert horizontal.....   | -22-        |
| <b>III. Impact de l'acquisition et/ou de la création de gènes nouveaux dans les génomes bactériens. ....</b>                    | <b>-23-</b> |
| 1. Impact du transfert horizontal sur l'émergence de nouvelles fonctions .....  | -23-        |
| 2. Impact du mécanisme de duplication/divergence sur l'émergence de nouvelles fonctions   | -24-        |
| 3. Impact du réassortiment de modules sur l'émergence de nouvelles fonctions.....   | -24-        |
| 3.1 Le réassortiment de modules n'amène pas forcément à de nouvelles spécificités. ....   | -25-        |
| 3.2 Emergence de nouvelles spécificités par réassortiment de modules .....  | -26-        |
| 4. Impact de ces nouvelles fonctions sur l'adaptation des bactéries.....  | -27-        |
| <b>IV. Découverte d'un gène impliqué dans la séparation des cellules de <i>S. thermophilus</i>.....</b>                         | <b>-28-</b> |

**Résultats.....-31-****I. Mécanismes évolutifs conduisant à la création du gène *cse* chez *Streptococcus thermophilus* ..... -31-**

1. Le gène *cse* est impliqué dans la séparation des cellules de *S. thermophilus*..... -31-
2. Le gène *cse* est un gène chimérique. .... -32-
3. Le gène *cse* est un gène polymorphe. .... -33-
4. Importance de cette région de faible complexité. .... -35-
5. Mécanismes hypothétiques de création du gène *cse*. .... -36-

**II. Ce nouveau gène confère-t-il une fonction nouvelle à *S. thermophilus* ? ..... -39-**

1. *cse*, est-il un gène de création récente ? ..... -39-
  - 1.1 Recherche *in silico* du gène *cse* dans les génomes de *S. salivarius*. .... -40-
  - 1.2 Recherche par PCR du gène *cse* dans le groupe *salivarius*. .... -41-
2. Les protéines Cse et PcsB, ont-elles des fonctions cellulaires identiques ? ..... -46-
  - 2.1 Analyse du locus *pcsB* de *S. thermophilus* LMG18311. .... 46-
  - 2.2 *cse* et *pcsB*, deux gènes de fonction identique ou différente ? ..... -47-
    - a. Inactivation du gène *pcsB* chez *S. thermophilus*. .... -47-
    - b. Implication du gène *pcsB* dans l'organisation des septa chez *S. thermophilus*. .... -50-

**III. Le gène *cse* code une enzyme de séparation cellulaire chez *Streptococcus thermophilus*. . -53-**

1. Cse est une protéine extracellulaire. .... -54-
2. Les domaines LysM et CHAP sont essentiels à l'activité de la protéine Cse. .... -55-
3. La région centrale n'est pas essentielle à l'activité de Cse. .... -56-
4. Cse est une protéine de surface. .... -57-
5. Cse a une activité d'hydrolyse du peptidoglycane. .... -59-
  - 5.1 Mise en évidence de l'activité d'hydrolyse du peptidoglycane de Cse ..... -59-
  - 5.2 Spécificité de clivage du peptidoglycane du domaine CHAP ..... -61-

**Discussion / Perspectives ..... -85-****I. La « Superfamille des hydrolases du peptidoglycane CHAP » : les domaines CHAP et NLPC/P60..... -86-**

1. CHAP et NLPC/P60, deux domaines distincts..... -87-
2. La nature des liens peptidiques hydrolysés par la superfamille CHAP est très variée. .... -88-
3. La superfamille CHAP est associée à divers processus cellulaires. .... -89-
4. Distribution hétérogène des domaines CHAP et NLPC/P60 au sein des génomes des *Firmicutes*..... -90-

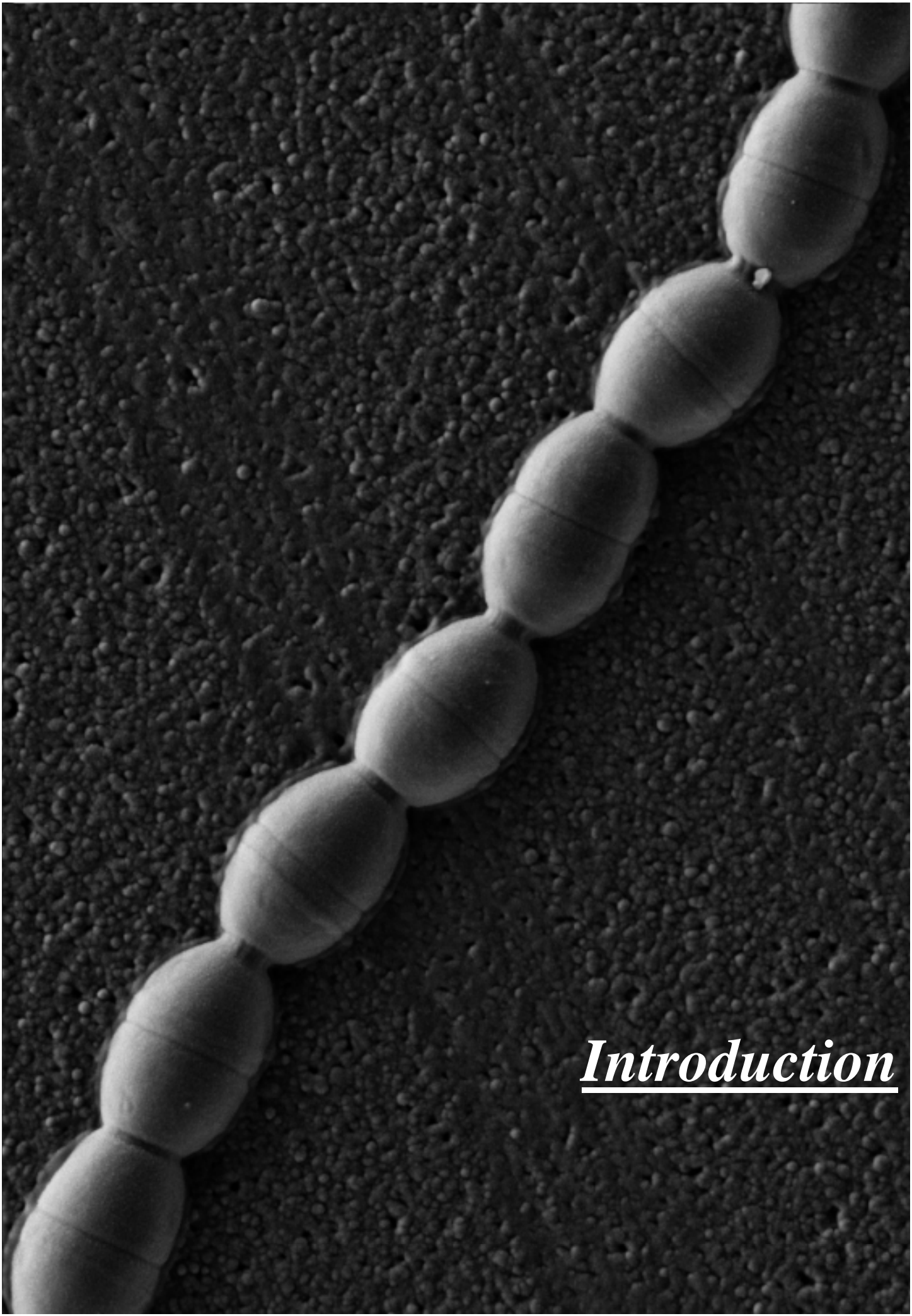
**II. La séparation cellulaire, un processus postérieur ou ultime de la division cellulaire ? ..... -92-**

1. La division cellulaire chez les streptocoques. .... -92-
2. La séparation cellulaire implique l'action d'hydrolases du peptidoglycane très diverses... -94-
  - 2.1 La séparation des cellules, un processus actif dépendant d'hydrolases du peptidoglycane..... -94-
  - 2.2 Diversité des enzymes de séparation cellulaire. .... -95-
  - 2.3 Les enzymes de séparation cellulaires seraient spécifiques d'espèces. .... -97-
  - 2.4 La séparation cellulaire, un processus postérieur à la division cellulaire ? ..... -97-
3. La séparation cellulaire nécessite une localisation spécifique des hydrolases du peptidoglycane. .... -98-

|   |   |              |
|---|---|--------------|
| 3.1   | Les enzymes de séparation cellulaires sont localisées dans des régions subcellulaires spécifiques : aux pôles et/ou aux septa des cellules. ....                                    | -98-         |
| 3.2   | Quel est le mécanisme conduisant à la localisation spécifique de Cse aux septa matures des cellules de <i>S. thermophilus</i> ?.....  | -98-         |
| a.  | Le peptide signal permet l'adressage des protéines de surface. ....   | -99-         |
| b.  | Le domaine LysM est impliqué dans l'attachement des protéines à la paroi. ....  | -100-        |
| c.  | La localisation spécifique de Cse peut-elle être influencée par les régions « Var » et « CHAP » ?.....  | -101-        |
| d.  | Les composants de la paroi pourraient être responsables de la localisation spécifique des hydrolases du peptidoglycane. ....  | -104-        |
| <b>III. Mécanisme de création de la diversité des hydrolases du peptidoglycane : implication du réassortiment de modules et des régions de faible complexité.....</b> |   |              |
| 1.  | L'architecture modulaire des hydrolases du peptidoglycane contribue à leur diversité....  | -106-        |
| 2.  | Le réassortiment de modules est à l'origine de la diversité des hydrolases du peptidoglycane .....  | -109-        |
| 3.  | La création de nouvelles hydrolases du peptidoglycane pourrait être favorisée par la présence de régions de faible complexité.....  | -110-        |
| 4.  | Conséquence fonctionnelle du réassortiment de module à l'origine de la création de <i>cse</i> .....   | -111-        |
| <b>IV. Impact de la longueur des chaînes sur l'adaptation de <i>Streptococcus thermophilus</i> à son environnement. ....</b>  |   |              |
| 1.  | Facteurs majeurs influençant la croissance des cellules en chaîne.....  | -113-        |
| 1.1   | Les facteurs génétiques déterminant le caractère chaînes de cellules. ....  | -113-        |
| a.  | La croissance des cellules en chaîne serait spécifique d'espèces. ....  | -113-        |
| b.  | La croissance des cellules en chaîne résulterait d'une déficience en hydrolase du peptidoglycane.....   | -114-        |
| c.  | Lien entre la division cellulaire et la croissance des cellules en chaîne. ....   | -115-        |
| 1.2   | Les facteurs environnementaux influent sur la longueur des chaînes de cellules. ....  | -115-        |
| 2.  | Influence de la longueur des chaînes de cellules sur l'adaptation des streptocoques à leur environnement.....   | -116-        |
| 2.1   | La longueur des chaînes de cellules module la virulence des streptocoques pathogènes.. ..   | -116-        |
| 2.2   | Causes et conséquences de la longueur des chaînes de cellules sur la croissance de <i>S. thermophilus</i> .....   | -117-        |
| a.  | Le polymorphisme intraspécifique de la longueur des chaînes de cellules pourrait être corrélé à un polymorphisme génétique.....   | -117-        |
| b.  | La longueur des chaînes de cellules de <i>S. thermophilus</i> pourrait être une réponse adaptative aux stress.....  | -117-        |
| c.  | La croissance des cellules en chaîne de <i>S. thermophilus</i> pourrait faciliter les relations intracellulaires avec ses partenaires dans la production des produits laitiers..... | -118-        |
| d.  | Lien entre la croissance des cellules en chaîne et le caractère probiotique de <i>S. thermophilus</i> .....   | -118-        |
| <b>Conclusion générale.....</b>   |   | <b>-119-</b> |
| <b>Références bibliographiques.....</b>   |   | <b>-121-</b> |

**Annexe .....-135-**  
*Articles .....-135-*  
*Actes de congrès .....-136-*





*Introduction*



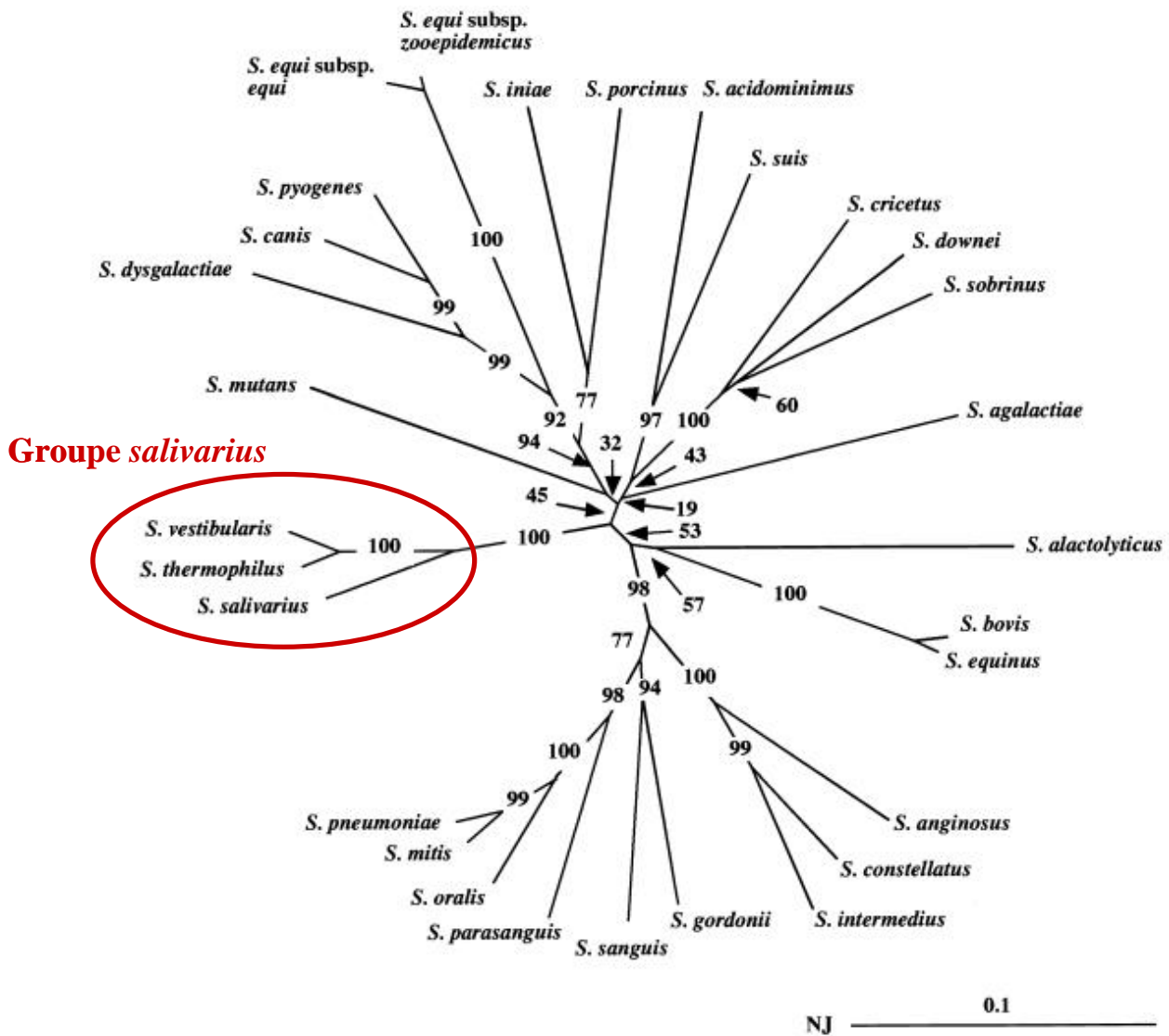
## **I. Evolution génomique de l'espèce *Streptococcus thermophilus***

*Streptococcus thermophilus* est une espèce bactérienne d'une grande importance économique, puisqu'elle est largement employée dans la fabrication des produits laitiers, et plus précisément dans la production des yaourts et de divers fromages. Le rôle de *S. thermophilus* dans la fermentation du lait consiste en la conversion rapide du lactose en acide lactique, entraînant une diminution rapide du pH. La conséquence de cette acidification est la précipitation des caséines, protéines du lait entraînant la coagulation de celui-ci. Pour la confection de fromages et de laits fermentés, *S. thermophilus* est utilisée seule ou en association avec plusieurs lactobacilles, « starters » mésophiles ou encore bifidobactéries, mais pour les yaourts, elle est toujours employée en co-culture avec *Lactobacillus delbrueckii* subsp. *bulgaricus* (*Lb. bulgaricus*) (Rasic et Kurmann, 1978 ; Tamime et Deeth, 1980).

### **1. Le genre *Streptococcus***

Le genre *Streptococcus* appartient au phylum des *Firmicutes* et regroupe des bactéries à coloration Gram positive, dépourvues d'activité catalase, homofermentaires et à bas pourcentage en bases G+C (Farrow et Collins, 1984). L'espèce *Streptococcus thermophilus* est phylogénétiquement apparentée aux espèces streptocoques de la division Viridans selon la classification établie par Sherman en 1937. La classe « streptocoques viridans » est subdivisée en 5 groupes, (i) le groupe *mutans* ; (ii) le groupe *anginosus* ; (iii) le groupe *sanguinis* ; (iv) le groupe *mitis* et (v) le groupe *salivarius* (Facklam, 2002). Toutes les espèces de la classe « streptocoques viridans », à l'exception d'une, *S. thermophilus*, sont commensales et rencontrées dans les cavités buccales, gastro-intestinales, et dans les tractus génitaux des mammifères (Facklam, 2002). Le dernier groupe *salivarius*, comprend trois espèces, *Streptococcus salivarius*, *Streptococcus vestibularis* et *S. thermophilus*. *S. salivarius* et *S. vestibularis* ont été identifiées dans des cas sporadiques d'endocardites et de méningites (Partridge, 2000 ; Idigoras et al., 2001), tandis que *S. thermophilus*, n'a été impliquée jusqu'à ce jour, dans aucune maladie infectieuse. Cependant, ces trois espèces sont étroitement apparentées génétiquement (Figure 1). Ceci a été démontré par l'analyse comparative des ARN ribosomiques 16S et par l'étude du polymorphisme du gène *sodA* de ces espèces (Kawamura et al., 1995 ; Poyart et al., 1998). La très forte valeur du « Bootstrap » (100%, Figure 1) reliant le groupe *salivarius* au reste de l'arbre et l'enracinement de l'arbre des streptocoques proposé par Onyenwoke en 2004 suggère que ce groupe est monophylétique et que les espèces qui le composent sont étroitement apparentées (Onyenwoke et al., 2004).

Ainsi, *S. thermophilus* est une espèce phylogénétiquement proche des streptocoques pathogènes.



**Figure 1 :** Relations phylogénétiques existantes entre les espèces du genre *Streptococcus* établies par l'analyse du polymorphisme du gène *sodA* (Poyart *et al.*, 1998).

Les distances ont été calculées à l'aide de la méthode du plus proche voisin (Neighbor-joining ou NJ) et représentées par des branches de longueur variable. La valeur indiquée sur la branche correspond à la valeur du « Bootstrap » indiquant la fiabilité des liens. La barre d'échelle représente 10% des différences dans les séquences nucléotidiques.

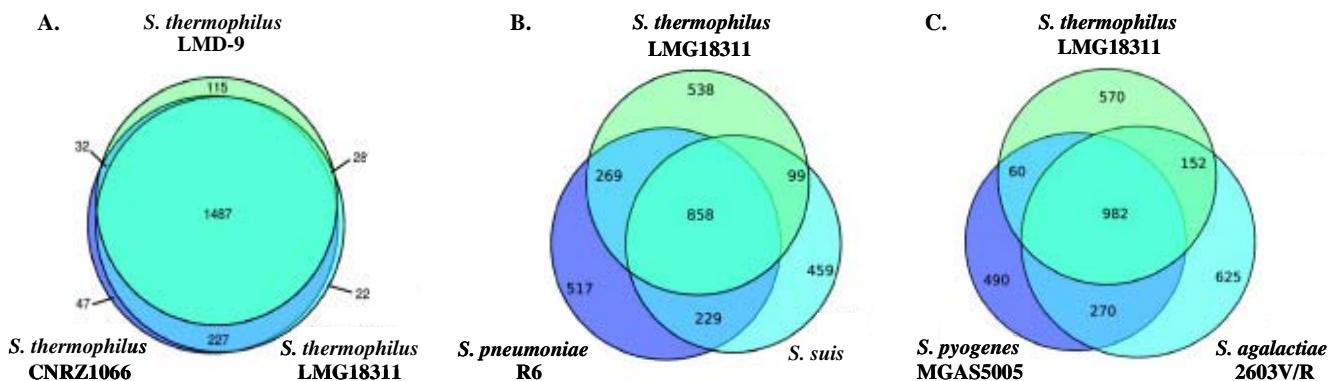
## 2. Adaptation de *S. thermophilus* dans le milieu lait : impact sur le génome

*S. thermophilus* est le seul streptocoque, connu à ce jour, à s'être adapté à un environnement particulier, le lait. La génomique comparative suggère que *S. thermophilus* a émergé récemment, et évolué par l'association de perte et/ou d'acquisition de fonctions via des événements de transfert horizontal (Bolotin *et al.*, 2004).

## 2.1 *S. thermophilus*, une espèce bactérienne clonale

A l'heure actuelle, la séquence génomique de trois souches de *S. thermophilus* isolées du yaourt : CNRZ1066, LMG18311 et LMD-9, est disponible dans les banques de données EMBL/GenBank (Bolotin et al., 2004 ; Makarova et al., 2006). Le génome de ces trois souches est circulaire d'une taille comprise entre 1,8 Mb et 1,86 Mb dont 76% à 83% est codant. Il est caractérisé par un faible pourcentage en base G+C de 39%.

L'analyse comparative des génomes des trois souches de *S. thermophilus* a révélé une importante conservation des gènes. En effet, 1487 séquences codantes sont homologues entre ces trois souches, soit 85% de leurs génomes (Figure 2A) (Lefébure et Stanhope, 2007). Les principales différences entre ces souches concernent les gènes impliqués dans la biosynthèse des exopolysaccharides, la synthèse de bactériocines, les résistances aux phages et les traces de prophages (Bolotin et al., 2004 ; Horvath et al., 2008 ; Deveau et al., 2008). De plus, la divergence de séquences au sein de ces trois souches est très faible, avec un taux de polymorphisme de 0,19%. Ce faible polymorphisme suggère que *S. thermophilus* représente une population clonale qui a récemment émergé (Delorme et al., 2007). L'émergence de *S. thermophilus* est estimée à près de 7000 ans, lorsque l'activité humaine laitière a commencé (Bolotin et al., 2004 ; Fox, 1993).



**Figure 2 : Représentation schématique de l'analyse comparative des génomes de *Streptococcus* (Lefébure et Stanhope, 2007).**

(A) Répartition des gènes pour les 3 génomes complètement séquencés de *S. thermophilus* LMD-9, CNRZ1066 et LMG18311. (B) Répartition des gènes pour les 3 génomes complètement séquencés de *S. thermophilus*, *S. pneumoniae* R6 et *S. suis*. (C) Répartition des gènes pour les 3 génomes complètement séquencés de *S. thermophilus* LMG18311, *S. pyogenes* MGAS5005 et *S. agalactiae* 2603V/R.

Par ailleurs, la comparaison du génome de *S. thermophilus* LMG18311 avec ceux de quatre espèces pathogènes complètement séquencés du genre *Streptococcus* : *Streptococcus pneumoniae*, *Streptococcus suis*, *Streptococcus pyogenes* et *Streptococcus agalactiae* (Figures 2B et 2C) a confirmé le lien de parenté entre *S. thermophilus* et ces espèces pathogènes. En effet, les figures 2B et 2C indiquent que plus de 54% des gènes de *S. thermophilus* sont orthologues avec les gènes des autres streptocoques, partageant ainsi une partie substantielle de leur métabolisme et physiologie (Hols et al., 2005). De plus, le faible polymorphisme intraspécifique de *S. thermophilus* (Figure 2A) renforce son statut, comme étant bien une espèce clonale qui a émergé récemment.

## 2.2 *S. thermophilus* a évolué par perte de fonctions

L'analyse comparative des génomes séquencés a révélé que le génome de *S. thermophilus* a perdu de nombreux gènes au cours de l'évolution. La plupart des gènes perdus ont été inactivés par des mutations soit dans la séquence codante, soit dans les signaux de régulation et d'expression tels que les promoteurs, les RBS (pour Ribosome Binding Sites) et les sites de fixations des régulateurs. L'inactivation de gènes par des mutations telles que les mutations non-sens, les décalages du cadre de lecture, les recodages ou encore les délétions lors des réarrangements génomiques, a conduit à l'apparition de pseudogènes dans le génome de *S. thermophilus* (Bolotin et al., 2004). Le nombre de gènes inactivés a été estimé à 181 pseudogènes, représentant ~ 10% de la totalité des séquences codantes du génome de *S. thermophilus* (Bolotin et al., 2004). Ce taux de pseudogènes est relativement élevé, par rapport à celui des streptocoques pathogènes qui a été estimé seulement à 4,2%, 0,7% et 1,6% respectivement, dans les génomes de *S. pneumoniae*, *S. pyogenes* et *S. agalactiae* (Bolotin et al., 2004).

### a. Perte des voies métaboliques au cours de l'évolution

L'analyse génomique de *S. thermophilus* a révélé la présence de pseudogènes dans différentes voies métaboliques, notamment dans le métabolisme énergétique (Tableau 1).

**Tableau 1 : Répartition des pseudogènes dans les différentes fonctions cellulaires de *S. thermophilus* CNRZ1066 et LMG18311 (Bolotin et al., 2004).**

|  | CNRZ1066           |                 |             | LMG18311           |                 |             |
|--|--------------------|-----------------|-------------|--------------------|-----------------|-------------|
|  | Total <sup>2</sup> | Nb <sup>3</sup> | %           | Total <sup>2</sup> | Nb <sup>3</sup> | %           |
| <b>Catégories fonctionnelles</b>                                   |                    |                 |             |                    |                 |             |
| Biosynthèse des acides aminés                                      | 75                 | 4               | 5,3         | 74                 | 4               | 5,4         |
| Biosynthèse des cofacteurs, groupes prosthétiques et transporteurs | 45                 | 8               | 17,8        | 45                 | 6               | 13,3        |
| Enveloppe cellulaire   | 93                 | 18              | 19,4        | 93                 | 16              | 17,2        |
| Processus cellulaires  | 82                 | 7               | 8,5         | 86                 | 7               | 8,1         |
| Métabolisme central intermédiaire                                  | 25                 | 3               | 12          | 20                 | 1               | 5           |
| <b>Métabolisme énergétique</b>                                     | <b>115</b>         | <b>34</b>       | <b>29,6</b> | <b>119</b>         | <b>39</b>       | <b>32,8</b> |
| Métabolisme des acides gras et phospholipides                      | 33                 | 2               | 6,1         | 34                 | 2               | 5,9         |
| Purines, pyrimidines, nucléotides et nucléosides                   | 72                 | 12              | 16,7        | 70                 | 12              | 17,1        |
| Fonctions de régulations   | 101                | 23              | 22,8        | 100                | 23              | 23          |
| Métabolisme ADN, Réplication                                       | 93                 | 15              | 16,1        | 93                 | 13              | 14          |
| Métabolisme ARN, Transcription                                     | 37                 | 0               | 0           | 37                 | 0               | 0           |
| Métabolisme des protéines, Traduction                              | 146                | 5               | 3,4         | 144                | 5               | 3,5         |
| <b>Protéines de liaisons et de transport</b>                       | <b>253</b>         | <b>76</b>       | <b>30</b>   | <b>246</b>         | <b>77</b>       | <b>31,3</b> |
| Autres fonctions   | 147                | 90              | 61,2        | 138                | 82              | 59,4        |

<sup>1</sup>CDS signifie les séquences codantes.

<sup>2</sup>indique le nombre total de CDS

<sup>3</sup>correspond au nombre de CDSs tronquées et retrouvées dans les génomes de *S. thermophilus*

Les différentes catégories fonctionnelles indiquées dans le tableau 1 ne comportent pas le même taux de pseudogènes. En effet, pas ou peu de pseudogènes sont retrouvés dans le métabolisme des ARN et des protéines. Ces voies métaboliques étant essentielles à la cellule bactérienne, n'acceptent peu ou pas de mutations. Par ailleurs, peu de pseudogènes sont retrouvés dans le métabolisme des acides aminés et des acides gras (tableau 1). Ces voies métaboliques très conservées, soulignent l'importance et surtout, la nécessité des acides aminés et des acides gras pour la croissance de *S. thermophilus* dans son environnement.

En revanche, l'analyse du génome de *S. thermophilus* a révélé de nombreux pseudogènes dans les loci dédiés à l'absorption et à la dégradation des sucres (~ 30%, tableau 1 « métabolisme énergétique et protéines de liaison et de transport »), affectant principalement les gènes codant les transporteurs de type PTS, transporteurs actifs des sucres (il s'agit d'une phosphotransférase couplée à l'hydrolyse du phosphoénolpyruvate) (Bolotin et al., 2004). C'est le cas des gènes *ptsG*, *fruA*, *bglP* et *treP* qui codent respectivement les systèmes PTS du glucose, du fructose, de la  $\beta$ -glucoside et du tréhalose. Mais aussi, les gènes codant les transporteurs de type ABC sont ciblés, notamment le gène *malE* de l'opéron maltose (Hols et al., 2005). *S. thermophilus* est incapable également de métaboliser le galactose, du fait que le gène *galK* codant l'enzyme nécessaire pour initier sa dégradation, la galactokinase, est inactivé (Van den Bogaard et al., 2004). Ceci peut s'expliquer par le fait que ces différents sucres (le glucose, le fructose et le galactose) ne sont pas présents dans l'habitat laitier de *S. thermophilus*, et par conséquent, ces différentes voies métaboliques ne sont d'aucune utilité pour la croissance de *S. thermophilus*.

Ainsi, l'adaptation de *S. thermophilus* dans son environnement laitier aurait contribué à la simplification des voies métaboliques des sucres qui a amené à se spécialiser uniquement dans la dégradation du lactose, source de carbone du lait (Bolotin et al., 2004). Cependant, cette spécialisation ne permet pas à *S. thermophilus* de coloniser de nouvelles niches écologiques, contrairement aux streptocoques pathogènes qui révèlent une haute flexibilité d'utilisation des sucres.

### **b. Perte des gènes de virulence**

La disponibilité de la séquence génomique de *S. thermophilus*, a permis de rechercher systématiquement dans le génome, les déterminants génétiques potentiels de virulence tels que les toxines, les protéines d'adhérences, les facteurs permettant la dissémination et la colonisation des bactéries pathogènes (Bolotin et al., 2004 ; Tettelin et al., 2001).

L'analyse comparative du génome de *S. thermophilus* CNRZ1066 avec ceux de deux streptocoques pathogènes, *S. pyogenes* M1 et *S. pneumoniae* TIGR4 (Tableau 2) a révélé que 25% des gènes liés à la virulence « VRGs (pour Virulence Related Genes) » étaient absents du génome de *S. thermophilus*, alors qu'ils sont présents à la fois dans les génomes de *S. pyogenes* et *S. pneumoniae*, et que 40% d'entre eux sont retrouvés dans des régions synthéniques entre ces deux génomes (Bolotin et al., 2004). Par ailleurs, les gènes de virulence présents dans le génome de *S. thermophilus* sont soit sous une forme non fonctionnelle à cause de mutations qui les inactivent (il s'agit de pseudogènes), soit des orthologues qui codent des protéines impliquées dans des fonctions cellulaires basiques de *S. thermophilus* (Bolotin et al., 2004) comme par exemple, les gènes *eps/cps* impliqués dans la biosynthèse des exopolysaccharides.

Tableau 2 : Identification des « VGRs » putatifs dans le génome de *S. thermophilus* CNRZ1066 par comparaison avec ceux des génomes de *S. pyogenes* M1 et *S. pneumoniae* TIGR4 (Bolotin et al., 2004)

|                              | VRGs | VRGs dans<br><i>S. thermophilus</i> | VRGs inactivés <sup>1</sup> dans<br><i>S. thermophilus</i> | % de perte des VRGs dans<br><i>S. thermophilus</i> |
|------------------------------|------|-------------------------------------|--|--|
| <i>S. pyogenes</i>           |      |                                     |  |  |
| •VRGs démontrés <sup>2</sup> | 21   | 1                                   | 0  | 95,2   |
| •VRGs putatifs <sup>3</sup>  | 29   | 7                                   | 1  | 75,9   |
| <b>Total</b>                 | 50   | 8                                   | 1  | 84,0   |
| <i>S. pneumoniae</i>         |      |                                     |  |  |
| •VRGs démontrés <sup>2</sup> |      |                                     |  |  |
| Adhérence                    | 9    | 4                                   | 0  | 55,6   |
| Défenses de l'hôte           | 23   | 5                                   | 1  | 78,3   |
| Métabolisme                  | 35   | 20                                  | 2  | 42,9   |
| Transport/Protéases          | 29   | 19                                  | 3  | 34,5   |
| Autres                       | 22   | 10                                  | 1  | 45,5   |
| <b>Total</b>                 | 118  | 58                                  | 7  | 54,6   |

<sup>1</sup>Inactivation des gènes par un ou plusieurs décalages du cadre de lecture, par la présence d'un codon stop ou par l'extrémité 3' tronquée.

<sup>2</sup>Démonstration expérimentale

<sup>3</sup>Gènes identifiés lors de l'analyse génomique.

Parmi les 50 « VRGs » retrouvés dans le génome de *S. pyogenes* M1 (Ferretti et al., 2001), 84% sont absents du génome de *S. thermophilus* CNRZ1066 (Tableau 2). Il s'agit des gènes de virulence codant les protéines d'adhérence telles que des protéines d'adhésion à la matrice extracellulaire, des protéines de liaison au collagène, à la fibronectine ou encore des adhésines Lral, et ceux codant les toxines comme par exemple, les toxines érythrogéniques C et B (Bolotin et al., 2004).

Un important appauvrissement en gènes de virulence a également été confirmé par comparaison du génome de *S. thermophilus* avec celui de *S. pneumoniae* TIGR4 (~ 55%, tableau 2). Les gènes absents sont principalement les gènes qui codent les protéines de surface telles que des protéines de virulence PspA, PspC, des adhésines ou encore des protéases. Par ailleurs, parmi les 64 « VRGs » présents dans le génome de *S. thermophilus*, 12,5% de ces gènes sont des pseudogènes. Cependant, les « VRGs » actifs restant ne sont pas forcément des gènes de virulence chez *S. thermophilus*. En effet, si le métabolisme du lactose est un facteur de virulence avéré chez *S. pneumoniae* (catabolisme d'une source atypique du lactose lors d'une infection) (Tettelin et al., 2001), dans le cas de *S. thermophilus*, le lactose est la seule source de carbone dans le lait. De même que si *S. pneumoniae* dépourvu de capsule (*eps/csp*) n'est plus virulent, l'absence des gènes *eps* chez *S. thermophilus* confère des propriétés différentes aux produits laitiers.

De plus, *S. thermophilus* est dépourvu de gènes de résistances aux antibiotiques, autre facette importante du pouvoir pathogène (Bolotin et al., 2004)



Au cours de l'évolution, le génome de *S. thermophilus* a perdu de nombreux gènes de virulence. Si ces gènes peuvent être perdus sans pour autant affecter la survie de *S. thermophilus*, cela suggère que ces gènes sont non essentiels pour la survie de *S. thermophilus* en milieu lait.

### **2.3 Acquisition de nouvelles fonctions par transfert horizontal**

Bien que l'évolution du génome de *S. thermophilus* se soit traduite par la perte d'un grand nombre de gènes, le transfert horizontal a contribué à l'élaboration du génome de ce streptocoque laitier par acquisition de nouveaux gènes.

Le transfert horizontal joue un rôle majeur dans l'évolution des génomes bactériens. Il conduit notamment à un polymorphisme intraspécifique extrêmement important chez les bactéries (Lawrence et Roth, 1999 ; Welch et al., 2002). Le transfert horizontal constitue le principal mécanisme d'acquisition de nouveaux gènes. Par conséquent, il est impliqué dans l'acquisition d'un grand nombre de nouvelles fonctions, pouvant être le facteur clé d'adaptation des bactéries à de nouvelles niches écologiques. Ainsi, certains gènes présents dans le génome de *S. thermophilus* sont absents de ceux des streptocoques pathogènes. D'ailleurs, les ORFs apparentées à ces gènes sont fréquemment retrouvées dans le génome d'autres espèces ayant pour niche écologique, le lait. Il a été observé de récents événements de transfert horizontal entre *S. thermophilus* et des espèces relativement éloignées phylogénétiquement, mais de proximité écologique étroite (Guédon et al., 1995 ; Delcour et al., 2000).

Le nombre de gènes de *S. thermophilus* acquis par transfert horizontal est estimé à 226 gènes dont 61% d'entre eux présentent un pourcentage en base G+C significativement différent de celui du génome traduisant une origine différente de celles des streptocoques.

#### ***a. Les éléments génétiques contribuant à l'acquisition de nouveaux gènes dans le génome de S. thermophilus.***

L'étude du génome de *S. thermophilus* a révélé divers éléments mobiles jouant un rôle important dans l'acquisition de nouveaux gènes par les événements de transfert horizontal. Notamment, des séquences d'insertion (ISS1, IS981, IS911) qui sont homologues aux séquences appartenant aux lactocoques. Inversement, des séquences d'insertion (IS1191) de *S. thermophilus* sont retrouvées dans le génome de certaines souches de *L. lactis* (Guédon et al., 1995). Cependant, les séquences d'insertion ne sont pas les seuls éléments mobiles contribuant vraisemblablement à l'évolution des génomes de *S. thermophilus*, d'autres éléments mobiles ont été identifiés. Il s'agit des éléments intégratifs conjugatifs (ICES1, ICES3, pour Integrative Conjugative Elements) spécifiques de *S. thermophilus*. Ces deux îlots génomiques sont capables de s'exciser et de s'intégrer par recombinaison site-spécifique dans l'extrémité 3' du gène *fda* codant la fructose-1,6-diphosphate aldolase, enzyme intervenant dans la glycolyse (Burrus et al., 2000, Pavlovic et al., 2004). De même, des CIMEs (pour cis mobilizable elements), d'autres îlots génomiques dérivant des éléments intégratifs conjugatifs, pourraient permettre l'acquisition de nouveaux éléments génétiques par formation d'un tandem avec les ICEs (Pavlovic et al., 2004). Depuis peu, il a été démontré que ces deux éléments intégratifs conjugatifs ICES1 et ICES3 se transféraient par conjugaison vers d'autres souches de *S. thermophilus*, mais également vers d'autres bactéries lactiques, notamment les bactéries

pouvant être utilisées en co-culture avec *S. thermophilus* telles que *L. lactis*, mais aussi, vers des bactéries pathogènes opportunistes telles que *S. pyogenes* et *Enterococcus faecalis* (Bellanger X., Communication orale CBL2007, Rennes) suggérant que ces îlots génomiques pourraient jouer un rôle majeur dans les échanges de gènes entre *S. thermophilus* et les autres espèces bactériennes présentes en milieu laitier.

### b. Acquisition de nouveaux gènes par transfert horizontal

Des gènes acquis par transfert horizontal ont été identifiés dans le génome de *S. thermophilus*. Les gènes codent notamment des protéines impliquées dans la synthèse des exopolysaccharides (Bourgoin et al., 1999 ; Rallu et al., 2002 ; Broadbent et al., 2003), dans le catabolisme des protéines et des sucres du lait (Vaughan et al., 2001 ; Fernandez-Espla et al., 2000 ; Garault et al., 2000), dans la synthèse de bactériocines : les loci *blp* et *lab* (Mathot et al., 2003 ; Siezen et al., 1996), dans la tolérance à l'oxygène (Hols et al., 2005), dans les systèmes de restriction-modification ou encore des systèmes de résistance abortive (Pavlovic et al., 2004), deux mécanismes de résistance aux phages (Burrus et al., 2001 ; Solow et Somkuti, 2000). Il peut s'agir également des gènes codant des régulateurs transcriptionnels comme par exemple les gènes *rgg* (N. Leblond-Bourget, communication personnelle).

Un exemple d'acquisition de gènes par transfert horizontal, est l'îlot génomique de 17 Kb localisé au sein du gène *pepD*. Cet îlot est caractérisé par sa structure mosaïque qui possède trois séquences d'insertion de type *IS1191* à proximité desquelles se trouvent des fragments d'ADN ayant plus de 90% d'identité avec l'ADN génomique de trois autres bactéries laitières : *Lb. bulgaricus*, *L. lactis* subsp. *lactis* et *L. lactis* subsp. *cremoris* (Figure 3) (Bolotin et al., 2004).

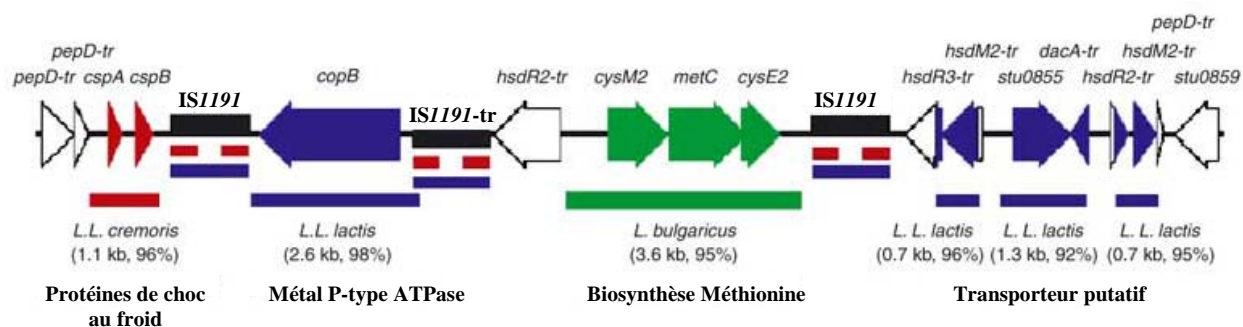


Figure 3 : Structure mosaïque de la région de 17 Kb acquis par transfert horizontal dans le génome de *S. thermophilus* (Bolotin et al., 2004).

Les gènes *cspA* et *cspB* codent des protéines de choc au froid ; le gène *copB* code un transporteur de cuivre et de potassium couplé à une ATPase ; les gènes *cys* et *met* sont impliqués dans la biosynthèse de la cystéine et méthionine, respectivement. Les gènes *hsdR*, *hsdM*, *dacA* et *pepD* codent respectivement, une endonucléase de restriction, une méthylase, une carboxypeptidase et une endopeptidase. Les gènes inactivés par troncature ou par frameshifts sont indiqués par « tr ».

Grâce à cet îlot génomique, *S. thermophilus* a acquis notamment les gènes de la biosynthèse de la méthionine de *Lb. bulgaricus*. La méthionine est un acide aminé rare et nécessaire à la croissance de *S. thermophilus* dans le lait (Tammine et Deeth, 1980).

La majorité des gènes acquis par transfert horizontal proviennent d'autres espèces laitières. *S. thermophilus* les aurait probablement acquis au cours des co-cultures lors de la fabrication des

produits laitiers (Bolotin *et al.*, 2004 ; Hols *et al.*, 2005; Guédon *et al.*, 1995). Ainsi, il semblerait que les espèces bactériennes proches phylogéniquement et appartenant à la même niche écologique puissent s'échanger fréquemment des informations génétiques par transfert horizontal. D'ailleurs, l'acquisition de ces gènes a une importance pour leur capacité d'adaptation dans cet environnement particulier, le lait (Hols *et al.*, 2005 ; Makarova *et al.*, 2006).

Par ailleurs, une quatrième souche de *S. thermophilus* a été depuis peu séquencée, il s'agit de la souche JIM8232 isolée du lait fermenté. La particularité de cette souche, est sa pigmentation jaune. La séquence génomique de la souche JIM8232 a révélé une région spécifique de 170 Kb d'ADN, porteur de VRGs. Il s'agit d'un élément mobile acquis récemment par transfert horizontal, présentant des homologies avec les VRGs de *Streptococcus suis* et ceux de *S. agalactiae* dont le gène conférant la pigmentation jaune. Ainsi, des échanges entre espèces phylogéniquement proches mais de proximité écologique éloigné sont également observés. Néanmoins, ce phénomène est plutôt rare, puisque ces bactéries ne partagent pas la même niche écologique (P. Renault, Communication orale CBL2007, Rennes ; C. Delorme, Communication orale Journée d'animation scientifique MICALIS, Paris).

### **3. Conclusion**

La comparaison des génomes des streptocoques a montré la perte de gènes ancestraux chez *S. thermophilus*. Ces gènes contribuant parfois fortement à la valeur adaptative chez les streptocoques pathogènes (« VRGs ») semblent avoir pas ou peu d'impact sur la valeur adaptative dans l'environnement laitier, et ont pu ainsi être perdu sans contre sélection chez *S. thermophilus*. Par ailleurs, cette comparaison de génomes a également montré l'émergence de nouveaux gènes acquis par transfert horizontal dont certains ont dans le milieu lait, un fort impact sur la valeur adaptative de *S. thermophilus*.

Ainsi, ces deux mécanismes ont tous deux contribué à l'évolution du génome de *S. thermophilus* et par conséquent à son adaptation à une niche écologique particulière, le lait.

## **II. La duplication/divergence et le réassortiment de modules, deux mécanismes majeurs de la création de gènes.**

Toutefois, bien que le transfert horizontal soit un mécanisme majeur de l'acquisition de nouveaux gènes dans le génome de *S. thermophilus*, on ne peut pas exclure l'apparition de nouveaux gènes par d'autres mécanismes. En effet, le transfert horizontal n'est pas le seul processus contribuant à l'acquisition de gènes dans les génomes bactériens. Deux autres mécanismes fondamentaux, la duplication de gènes suivie de leur divergence et le réassortiment de modules amènent à la création de nouveaux gènes au sein des génomes bactériens. Si le transfert horizontal n'est pas à l'origine de la création de gènes, ce processus permet néanmoins leur dissémination au travers d'échanges intra- et interspécifiques.

### **1. Création de nouveaux gènes par duplication et divergence**

La duplication de gènes est l'un des principaux mécanismes d'évolution qui a conduit à la formation de nouveaux gènes dans les génomes bactériens (Ohno, 1970 ; Koonin, 2005). La duplication de gènes est un moyen particulièrement efficace, puisque l'un des gènes dupliqués peut évoluer vers une nouvelle fonction, tandis que son homologue continue à diriger sa fonction d'origine (Koonin, 2005).

#### **1.1 Duplication de gènes**

L'analyse comparative des génomes séquencés a révélé une proportion importante de gènes dupliqués dans les génomes bactériens. En effet, dans le cas de *S. pneumoniae* par exemple, il a été estimé que 14% des gènes résultaient d'un événement de duplication (Tettelin et al., 2001). Un nombre équivalent de gènes dupliqués a également été retrouvé dans le génome de *L. lactis* (16%) et le génome de *Helicobacter pylori* (17%) (Bolotin et al., 2001 ; Tomb et al., 1997). Un cas extrême est le génome de *Mycoplasma pneumoniae* qui présenterait 44% de gènes dupliqués (Himmelreich et al., 1996). Ainsi, le nombre de gènes dupliqués est variable d'une espèce à l'autre.

Ce type de remaniements engendre une copie surnuméraire d'un gène dans le génome bactérien. Parfois, les deux copies d'un gène peuvent être conservées dans les génomes. C'est le cas par exemple des gènes *tufA* et *tufB* d'*Escherichia coli* qui codent des facteurs traductionnels EF-Tu (Filer et Furano., 1981). Filer et Furano, ont constaté que la duplication des gènes *tuf* était très répandue au sein des génomes des bactéries à coloration Gram négative, contrairement, aux bactéries à coloration Gram positive (à bas %G+C et à haut %G+C) qui ne possède qu'un seul gène *tuf* (Filer et Furano, 1981 ; Sela et al., 1989). Depuis près de 10 millions d'années, les gènes *tufA* et *tufB* d'*E. coli* ont très peu divergé, puisqu'ils présentent 99,9% d'identité en séquence nucléotidique (Filer et Furano,

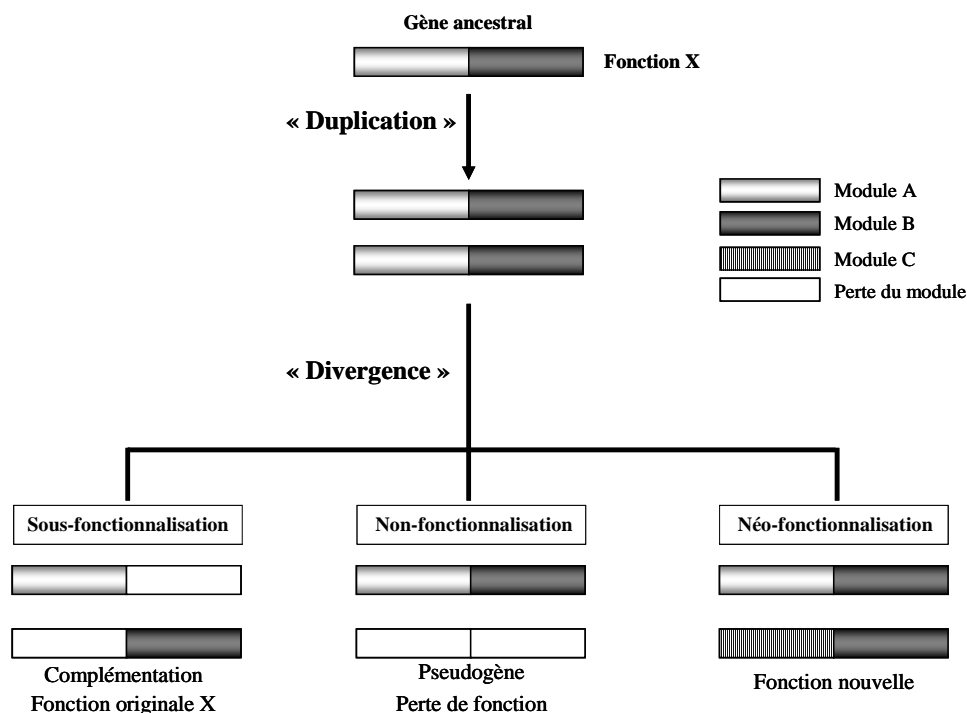
1981 ; Yakota *et al.*, 1980). Les auteurs suggèrent que ces deux gènes confèrent des fonctions essentielles chez *E. coli* (Filer et Furano, 1981 ; Sela *et al.*, 1989).

Toutefois, si une copie d'un gène dupliqué est suffisante pour la survie de l'organisme, alors il est fort probable que la seconde copie va rapidement diverger. La conservation de gènes dupliqués codant la même fonction est un phénomène très rare chez les procaryotes (Riley et Anilionis, 1978). En effet, pour la majorité des gènes dupliqués, l'une des deux copies peut accumuler des mutations, tandis que l'autre copie assure toujours la fonction essentielle. Par conséquent, l'événement mutationnel d'une des deux copies ne sera pas contre sélectionné. Cependant, les gènes en question peuvent se mettre à diverger fonctionnellement, ce qui peut présenter un avantage adaptatif potentiel pour la bactérie (Ohno, 1970).

## 1.2 Devenir des gènes dupliqués

La théorie d'évolution des gènes dupliqués établie par Force *et al.* en 1999, prévoit plusieurs devenirs possibles pour ces gènes (Figure 4) :

- (i) la sous-fonctionnalisation ;
- (ii) la non-fonctionnalisation ou la pseudogénéisation ;
- (iii) la néo-fonctionnalisation.



**Figure 4 : Mécanismes d'évolution des gènes dupliqués (Force *et al.*, 1999).**

Il s'agit du modèle « duplication-complémentation-dégénérescence » dans lequel les gènes dupliqués vont subir 3 modes des fonctionnalisations : la non-fonctionnalisation ; la sous-fonctionnalisation et la néo-fonctionnalisation. Dans ce modèle, le gène dupliqué est constitué de 2 modules (A et B) conférant à la protéine son activité. Après divergence, les modules peuvent soit évoluer par perte de fonction ou par acquisition d'une nouvelle fonction (module C).

Le processus de sous-fonctionnalisation peut avoir lieu lors d'une duplication d'un gène ayant au moins deux modules conférant à la protéine son activité. Dans l'hypothèse de la sous-fonctionnalisation, l'activité de la protéine est due à la complémentation fonctionnelle des deux copies du gène ayant perdu des modules différents par accumulation de mutations (Force et al., 1999). Contrairement à la sous-fonctionnalisation, la non-fonctionnalisation aboutit à la perte d'une copie du gène par accumulation de mutations conduisant à l'apparition de pseudogènes dans le génome bactérien. En revanche, la néo-fonctionnalisation va conduire à l'apparition d'une fonction nouvelle pour l'une des copies du gène qui subit une divergence (Force et al., 1999), tandis que l'autre assure toujours la fonction première.

Ainsi, le mécanisme de duplication/divergence conduit à la création d'une nouvelle fonction seulement dans le cas d'une néo-fonctionnalisation, puisque les deux autres processus, la sous-fonctionnalisation et la non-fonctionnalisation amènent à la fonction d'origine. Néanmoins, la néo-fonctionnalisation est un processus évolutif lent puisqu'il nécessite l'accumulation d'événements mutationnels de type délétion, substitution et/ou insertion dans le gène dupliqué. Toutefois, ces mutations ponctuelles conduiront plus facilement à la perte de la fonction qu'à l'apparition d'une fonction nouvelle. Par conséquent, l'apparition de pseudogènes est un phénomène plus fréquemment retrouvé dans les génomes bactériens que la création de nouveaux gènes par la néo-fonctionnalisation (Force et al., 1999).

## **2. Création de nouveaux gènes par réassortiment de modules**

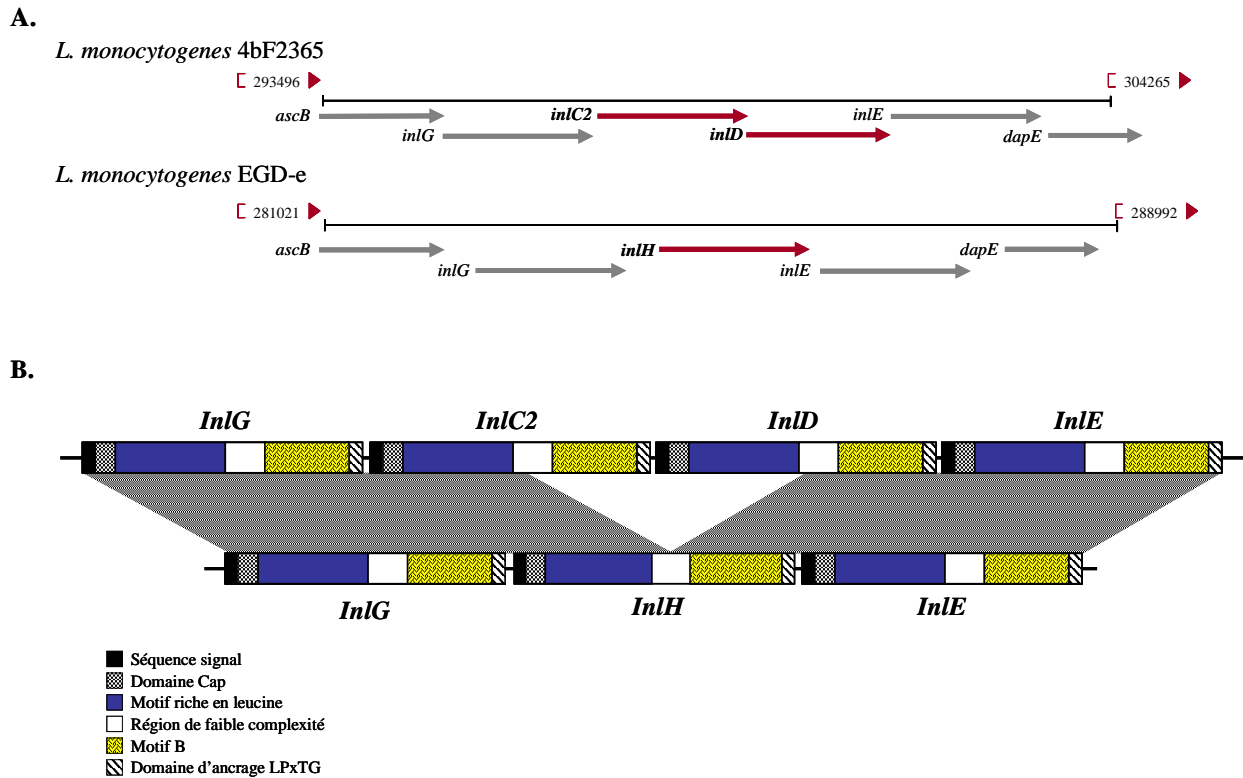
La plupart des gènes bactériens présentent des modules ou « domaines » pouvant assurer une fonction indépendante. Il s'agit d'unités d'évolution à part entière qui ont une structure. Le réassortiment de modules est un mécanisme qui permet de générer de nouveaux gènes à partir de gènes préexistants par assemblage différentiel de leurs modules/domaines. Contrairement au processus de duplication/divergence, le réassortiment de modules est un mécanisme d'évolution rapide qui nécessite un seul événement mutationnel pour générer de nouveaux gènes dans les génomes bactériens. Les événements mutationnels contribuant au réassortiment de modules peuvent être des recombinaisons intra- ou intergéniques. La conséquence de ces événements de réassortiment de modules est l'apparition de gènes « chimériques ou mosaïques » codant des protéines multi-domaines au cours de l'évolution des génomes bactériens.

### **2.1 Création de gènes par recombinaison intramoléculaire**

Chez *Listeria monocytogenes* 4b F2365, l'opéron *inlGC2DE* codent des protéines de surface appartenant à la famille des internalines, impliquées dans l'adhésion et l'invasion des cellules hôtes (Vazquez-Boland et al., 2001 ; Tsai et al., 2006). Cet opéron est aussi retrouvé au sein de la souche *L. monocytogenes* EGD-e, il est nommé *inlGHE* (Raffelsbauer et al., 1998 ; Dramsi et al., 1997).

Toutefois, la comparaison des séquences des gènes *inl* des deux souches de *L. monocytogenes* révèle la conservation uniquement de deux gènes, le gène *inlG* et le gène *inlE*, alors que les gènes *inlC2*, *inlD* et *inlH* sont spécifiques aux souches de *L. monocytogenes* (Figure 5A)

(Vazquez-Boland *et al.*, 2001). Par contre, le gène *inlH* présente deux régions homologues à celles des deux gènes *inlC2* et *inlD* de *L. monocytogenes* 4b F2365 (Figure 5B). Ce gène est vraisemblablement le résultat d'une fusion entre les gènes *inlC2* et *inlD* par recombinaison intramoléculaire ayant conduit à une délétion de l'extrémité 3' du gène *inlC2* et de l'extrémité 5' du gène *inlD* chez la souche *L. monocytogenes* EDG-e (Raffelsbauer *et al.*, 1998 ; Tsai *et al.*, 2006).



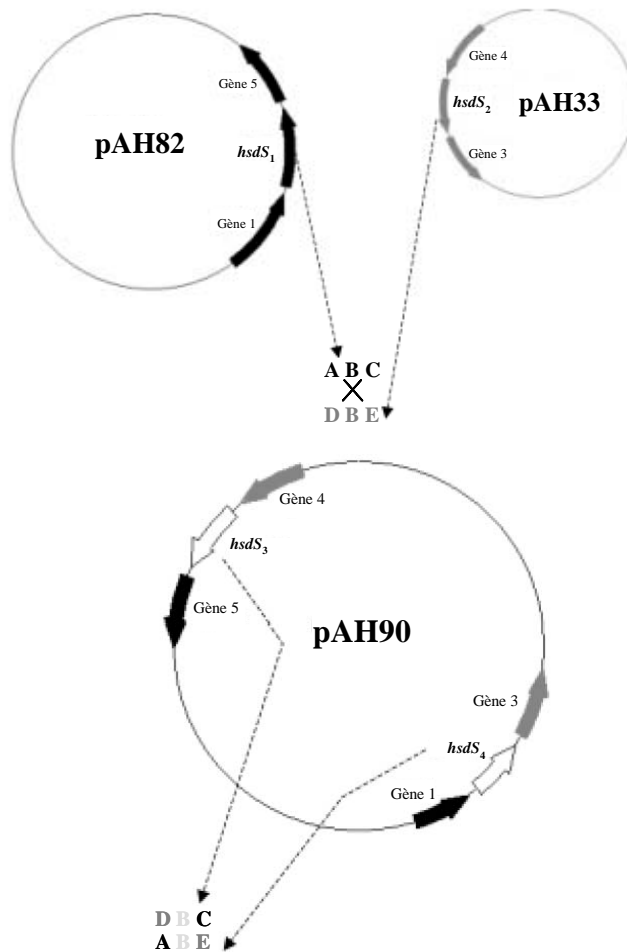
**Figure 5 :** Le gène *inlH* de *L. monocytogenes* EGD-e résulterait d'un événement de réassortiment de module entre deux gènes adjacents (Vazquez-Boland *et al.*, 2001).

A. Illustration de l'environnement génique de l'opéron *inlGC2DE* de *L. monocytogenes* 4b F2365 et de l'opéron *inlGHE* de *L. monocytogenes* EGD-e. B. Représentation schématique de l'architecture des internalines *InlC2*, *InlD* et *InlH*. La protéine *InlH* est une protéine chimérique constituée de l'extrémité N-terminale de la protéine *InlC2* et de l'extrémité C-terminale de la protéine *InlD*. Les régions grisées correspondent au début et à la fin des homologies observées.

Ainsi, le gène *inlH* est un gène chimérique créé par réassortiment de modules entre deux gènes adjacents. Ce gène résulte probablement d'un événement de recombinaison intramoléculaire.

## 2.2 Création de gènes par recombinaison intermoléculaire

Des systèmes naturels de création de gènes mosaïques par recombinaison intermoléculaire ont été identifiés chez *Lactococcus lactis* (Dybvig *et al.*, 1998 ; O'Sullivan *et al.*, 2000). Cet événement est illustré par la recombinaison homologue entre deux gènes *hsdS* portés par les plasmides pAH82 et pAH33 de *L. lactis* (Figure 6) (O'Sullivan *et al.*, 2000).



**Figure 6 : Recombinaison homologue entre les gènes *hsdS* portés par les plasmides pAH82 et pAH33 de *L. lactis* (O'Sullivan *et al.*, 2000).**

Les plasmides pAH82 et pAH33 sont tous deux porteurs du gène *hsdS*. Ces deux gènes possèdent une architecture modulaire différente qui est illustrée ici par les lettres ABC et DBE pour les gènes *hsdS1* et *hsdS2*, respectivement. Chaque gène est constitué d'une région centrale conservée (la région B), siège d'événements de recombinaison homologue conduisant à l'échange des domaines variables (A, B, D et E). La conséquence de cet événement est la création de deux nouveaux gènes *hsdS3* et *hsdS4* dont l'architecture modulaire est respectivement DBC et ABE.

Cette recombinaison homologue s'accompagne également d'un événement de co-intégration des plasmides pAH82 et pAH33 formant ainsi le plasmide pAH90 pour lequel l'ordre des gènes est bouleversé au voisinage des gènes recombinés.

Les plasmides pAH82 et pAH33 de *L. lactis* représentés dans la figure 6, portent chacun un gène *hsdS*, qui contient une région centrale conservée (notée « B ») et deux modules variables (notés A-C pour le gène *hsdS1* et D-E pour le gène *hsdS2*) codant des domaines TDR (pour Target Recognition Domain). La recombinaison homologue entre les régions conservées B des gènes *hsdS* a conduit à un échange entre les modules variables des gènes *hsdS1* et *hsdS2* (Figure 6). La conséquence de ce réassortiment de modules est la création de deux nouveaux gènes mosaïques *hsdS3* et *hsdS4* dont l'architecture modulaire est différente de celle des gènes parentaux.

### 2.3 Les séquences de faible complexité pourraient favoriser la création de nouveaux gènes par le réassortiment de modules.

La recombinaison entre fragments de gènes préexistants par un unique événement mutationnel contribue à la création de nouveaux gènes par réassortiment de modules. Toutefois, pour être efficace, les réarrangements doivent respecter la phase de lecture et la localisation des sites de recombinaison ne doit pas induire d'inactivation de la fonction propre à chaque module ou domaine. La fonctionnalité d'un module ou domaine aura notamment plus de chance de rester intact si le site de fusion n'est pas localisé à l'intérieur des modules.

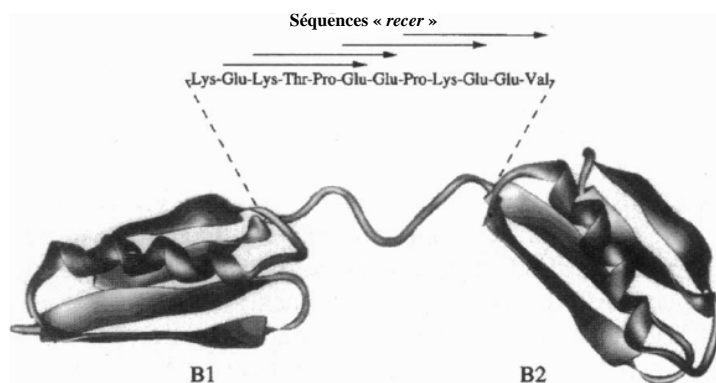
La création de gènes fonctionnels par réassortiment de modules dans les génomes bactériens peut s'expliquer par la présence de séquences de faible complexité. Ce sont des séquences



répétées de un à plusieurs nucléotides qui sont très fréquentes dans les génomes bactériens (Rocha, 2003a). Les séquences répétées les plus abondantes dans les génomes bactériens sont par exemple, les motifs de type SSRs (pour Simple Sequence Repeats) qui sont des séquences répétées en tandem d'un motif de 1 à 5 nucléotides ; les motifs de type TRs (pour Tandem Repeats) qui sont des répétitions en tandem d'une unité de taille supérieure à 5 nucléotides ; ou encore des séquences dégénérées de SSRs et TRs (Rocha., 2003a). Ces séquences répétées très diverses et très abondantes sont la cible des événements de recombinaison (Achaz et al., 2002 ; Rocha, 2003a), et par conséquent si elles sont retrouvées à l'intérieur même des séquences codantes, ces séquences répétées seront la cible des réassortiments de modules (De Château et Björck, 1994).

La famille des gènes mosaïques *pab* chez *Peptostreptococcus magnus* est un exemple de gènes ayant évolué par réassortiment de modules par recombinaison entre séquences de faible complexité de type SDRs (pour Short Directed Repeats) (De Château et Björck, 1994). La protéine PAB (pour Peptostreptococcal Albumin-Binding) qui en résulte, possède un domaine de liaison HSA (pour Human Serum Albumin) qui est homologue à celui des protéines G des streptocoques (Reis et al., 1984 ; Björck et al., 1987). Par ailleurs, cette protéine présente des séquences interdomaines particulières nommées « *recer* » (pour *r*ecombinaison *s*pac*e*r) qui sont des séquences de faible complexité. De même que la protéine PAB, des séquences de faible complexité sont retrouvées dans les protéines G des streptocoques (Achari et al., 1992). Par conséquent, la structure mosaïque du gène *pab* résulterait d'une recombinaison entre gènes de différentes espèces au niveau de ces séquences « *recer* » (De château et Björck, 1996). D'autres protéines possèdent également des séquences « *recer* » telles que la protéine de surface M de *S. pyogenes* (Mouw et al., 1988). Les événements de recombinaison impliquant ces séquences « *recer* » pourraient être à l'origine de l'évolution rapide des protéines de surface de ces bactéries pathogènes (De Château et Björck, 1996 ; Rocha, 2002 ; Rocha et al., 2003a et 2003b).

Par conséquent, les séquences de faible complexité sont des éléments génétiques bénéfiques pour le réassortiment de modules chez les bactéries. Par ailleurs, le réassortiment de modules via les séquences de faible complexité aurait permis l'émergence de protéines avec des domaines protéiques distincts séparés par un linker flexible de faible complexité structurale. La protéine L de *P. magnus* est un exemple représentatif de cette architecture modulaire (Figure 7) (De château et Björck, 1996).



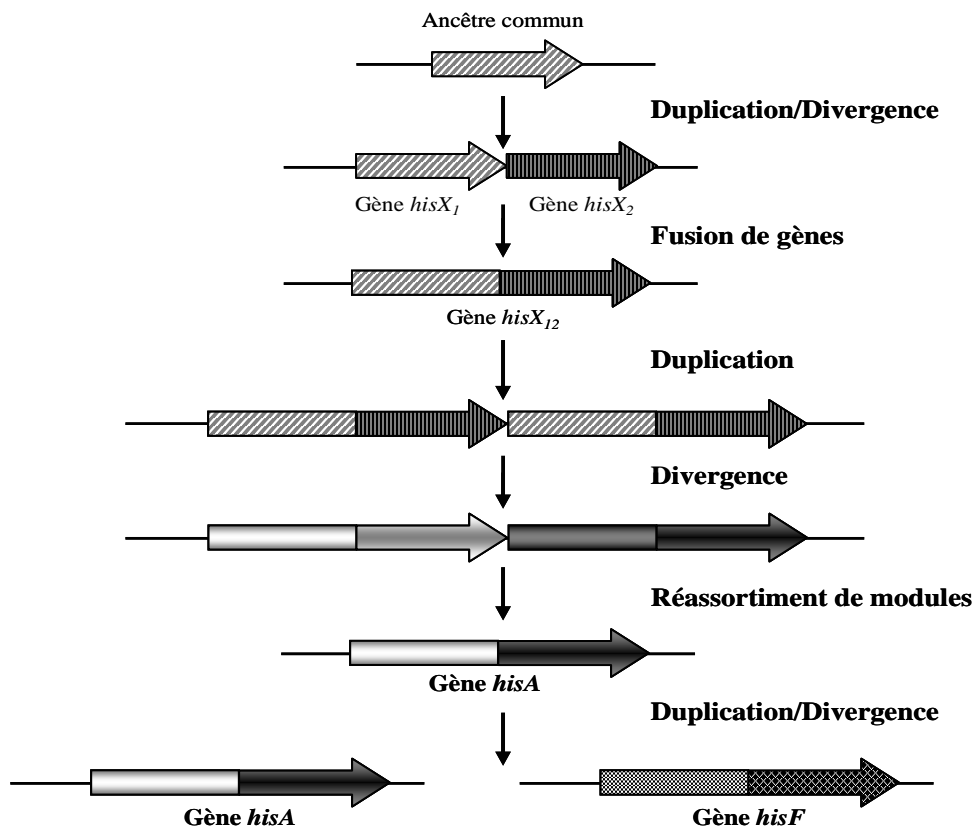
**Figure 7 : Architecture modulaire de la protéine L de *P. magnus* (De château et Björck, 1996).**

Les domaines B1 et B2 de la protéine L sont des domaines de liaison des chaînes légères des immunoglobulines. Ces deux domaines sont liés entre eux par un linker (ou une charnière) flexible dans lequel se trouve l'emplacement de quatre séquences « *recer* » qui se chevauchent.

### 3. La duplication/divergence et le réassortiment de modules, deux processus non exclusifs

De part leur homologie, les copies de gènes dupliqués, si elles ont peu divergé sont des sites préférentiels de réarrangements géniques contribuant à l'évolution des génomes bactériens. Ainsi, ces deux processus, duplication/divergence et réassortiment de modules ne sont pas exclusifs, et contribueraient ensemble à l'émergence de nouveaux gènes

Plusieurs opérons codant des enzymes de diverses voies de biosynthèse des acides aminés sont apparus dans le génome bactérien par des événements successifs de duplication et de fusion de gènes. C'est le cas par exemple de l'opéron *his* qui code les enzymes de la voie de biosynthèse de l'histidine (Fani et al., 1995). La voie de biosynthèse de l'histidine nécessite l'action de 8 enzymes codées par les gènes *hisG*, *hisD*, *hisC*, *hisB*, *hisH*, *hisA*, *hisF* et *hisIE* chez *E. coli* (Carlomagno et al., 1988 ; Winkler, 1978). Parmi ces gènes, deux résultent à la fois d'événements successifs de duplication et de fusion, il s'agit des gènes *hisA* et *hisF* qui codent tous deux, des protéines modulaires assurant respectivement la fonction d'imidazole carboxamide isomérase et la fonction d'imidazole glycérol phosphate synthase. La figure 8 présente les différents événements ayant conduit à l'émergence des gènes *hisA* et *hisF* (Fani et al., 2007).



**Figure 8 : Mécanisme d'évolution des gènes *hisA* et *hisF* (d'après Fani et al., 2007)**

Illustration schématique des événements ayant conduit à la création des gènes *hisA* et *hisF* dans la voie de biosynthèse de l'histidine chez *E. coli*. L'émergence de ces deux gènes résulterait d'une succession d'événements de duplication/divergence et de réassortiment de modules.

A partir du gène ancestral, un événement de duplication/divergence a conduit à l'apparition des gènes *hisX<sub>1</sub>* et *hisX<sub>2</sub>* adjacents. Puis leur fusion a conduit à la création du gène *hisX<sub>12</sub>*. Ce gène a ensuite été dupliqué et les deux copies qui en résultent ont divergé. Un événement de réassortiment de modules entre ces deux copies a conduit à la création d'un gène nouveau, le gène *hisA*. Enfin, le gène *hisF* est apparu par duplication/divergence à partir du gène *hisA* dans la voie de biosynthèse de l'histidine (Fani et al., 1995).

Ainsi, bien que la duplication de gènes soit un des mécanismes majeur à l'origine de l'apparition des voies métaboliques dans les génomes bactériens, d'autres événements importants ont amené à leur émergence, notamment le remaniement des gènes dupliqués (Fani et al., 2007).

#### **4. Le transfert horizontal contribue indirectement à la création de nouveaux gènes.**

Au cours de l'évolution, le transfert horizontal est un processus primordial dans l'acquisition de nouveaux gènes chez les bactéries. Cependant, ces gènes nouvellement acquis peuvent constituer une importante source de matériel génétique et servir de matrice pour les mécanismes de duplication/divergence et de réassortiment de modules. Par conséquent, il est fort probable que les gènes acquis par transfert horizontal aient été impliqués pour une grande partie dans la création de nouveaux gènes dans les génomes bactériens.

##### **4.1 Duplication et transfert horizontal**

Au cours de l'évolution, des événements de duplication/divergence et de transfert horizontal ont contribué conjointement à l'évolution des génomes bactériens. En effet, un gène nouvellement acquis par transfert horizontal peut évoluer par un mécanisme de duplication/divergence. Ce gène peut correspondre à une seconde copie divergente d'un gène déjà présent dans le génome bactérien. Par conséquent, les événements de transfert horizontal et de duplication/divergence ne pourront pas être différenciés. Dans ce cas, le devenir d'une des deux copies du gène dupliqué contribuera soit à la formation d'un pseudogène ou à la création d'un gène nouveau dans le génome bactérien.

Un exemple de gène qui est apparu au cours de l'évolution dans une population bactérienne par des événements de duplication/divergence et de transfert horizontal est le gène *fhuD1* de *Staphylococcus aureus* (Sebulsky et al., 2004). *S. aureus* utilise les hydroxymates ferriques comme source de fer pour sa croissance et ceci grâce au système Fhu (pour Ferric hydroxymate uptake). Ce système est constitué de 5 gènes : *fhuD1*, *fhuD2*, *fhuC*, *fhuB* et *fhuG*. Alors que les gènes *fhuB*, *fhuC* et *fhuG* codent des sous-unités de l'ATPase, enzyme essentielle pour l'utilisation des sidérophores dans *S. aureus* (Sebulsky et al., 2000), les gènes *fhuD1* et *fhuD2* codent tous deux, une lipoprotéine fixant des sidérophores ferriques. L'analyse comparative des génomes de sept souches de *S. aureus* a montré que les gènes *fhuD2*, *fhuB*, *fhuC* et *fhuG* étaient ubiquitaires. Tandis que le gène *fhuD1* n'est pas présent dans toutes les souches. La répartition des gènes *fhuD1* et *fhuD2* dans les génomes séquencés indique que ces deux gènes sont issus d'événements de duplication/divergence à partir d'un ancêtre commun. Par ailleurs, l'analyse des séquences a suggéré que le gène *fhuD1* était sujet à des transferts

horizontaux au sein de la population de *S. aureus*, puisque ce gène se trouvait à proximité d'une intégrase sur un îlot génomique.

## 4.2 Réassortiment de modules et transfert horizontal

Le réassortiment de modules associé aux événements de transfert horizontal est important pour la création de nouveaux gènes dans les génomes bactériens. En effet, l'analyse comparative des génomes a révélé la présence de gènes mosaïques issus de recombinaison interspécifique. C'est l'exemple du gène *pbp2x* de *Streptococcus pneumoniae* codant la protéine PBP2x essentielle dans la synthèse du peptidoglycane septal (Morlot et al., 2003). Ce gène présente des régions homologues à celles des gènes issus des espèces de *Streptococcus mitis* et de *Streptococcus oralis* (Dowson et al., 1993). Récemment, un mécanisme d'évolution des gènes mosaïques *pbp2x* chez *S. pneumoniae* a été proposé par Chi et al. (Figure 9) (Chi et al., 2007). L'émergence des gènes *pbp2x* conférant à *S. pneumoniae*, une résistance aux antibiotiques résulterait du remplacement par recombinaison d'un fragment de gène d'une souche sensible de *S. pneumoniae* par un fragment de gène d'une souche résistante de *S. mitis*, acquis par transfert horizontal. Par ailleurs, la diversification des gènes *pbp2x* serait le résultat d'autres événements de transfert horizontal et de recombinaison avec d'autres espèces résistantes comme *S. oralis*.

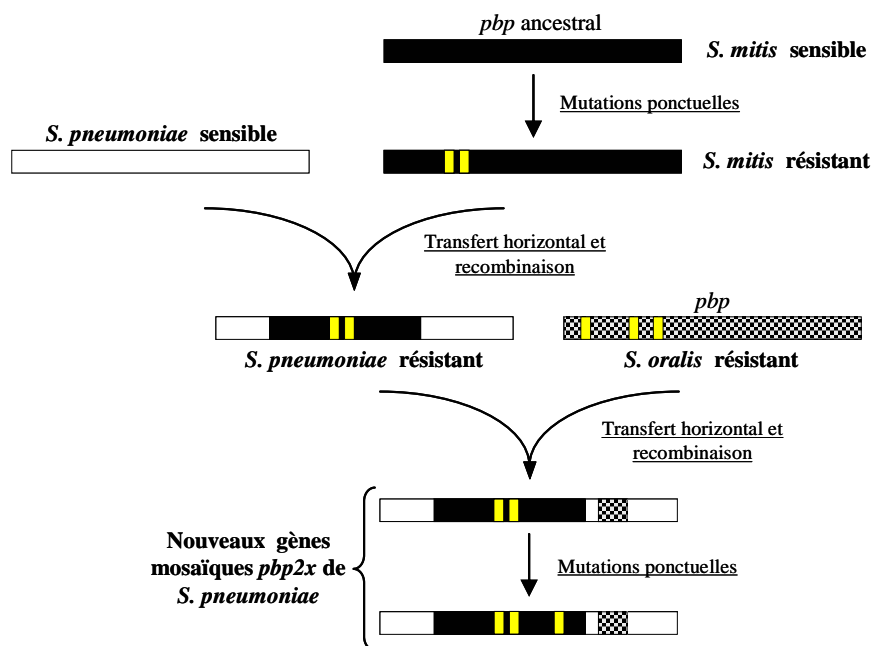


Figure 9 : Evolution des gènes mosaïques *pbp2x* chez *S. pneumoniae* (d'après chi et al., 2007)

*S. mitis* et *S. oralis* sont des streptocoques commensaux qui ont développé des résistances aux antibiotiques par acquisition de mutations ponctuelles (rectangle jaune) dans leurs gènes *pbp*. Des fragments de ces gènes sont acquis par la souche sensible de *S. pneumoniae* devenant résistante. Des mutations ponctuelles contribuent en association avec les événements de transfert horizontal et de recombinaison, à la diversification des gènes *pbp2x*.

### **III. Impact de l'acquisition et/ou de la création de gènes nouveaux dans les génomes bactériens.**

Le nombre important de génomes bactériens séquencés permet de constater que ces événements : transfert horizontal, duplication/divergence et réassortiment de modules, ont contribué de manière significative à l'acquisition et/ou à la création de nouveaux gènes dans les génomes bactériens. Cependant, les gènes nouvellement acquis et/ou créés dans un génome bactérien ne participent pas tous de la même manière à la valeur adaptative de la bactérie. En effet, les gènes fonctionnels conférant un avantage sélectif et par conséquent, contribuant à une augmentation de cette valeur adaptative seront sélectionnés positivement dans la cellule bactérienne. En revanche, les gènes présentant un désavantage, c'est-à-dire ayant un effet négatif seront quant eux, contre sélectionnés par délétion ou mutation au cours de l'évolution. De même que les gènes n'apportant aucun avantage, c'est-à-dire ayant un effet neutre, seront également contre sélectionnés. Par conséquent, ces mutations sont à l'origine de l'apparition de nombreux pseudogènes dans le génome bactérien.

Par ailleurs, les gènes nouvellement acquis par transfert horizontal et/ou créés par les mécanismes de duplication/divergence et de réassortiment de modules à partir de gènes préexistants posent la question de savoir si ces gènes confèrent de nouvelles fonctions dans les génomes bactériens. Et si oui, est-ce que ces fonctions nouvelles confèrent ou non un avantage adaptatif à la bactérie ? Le terme de « fonction nouvelle » désigne une fonction qui n'était pas présente initialement chez la bactérie et qui est différente de la fonction parentale. Ainsi, une modification de spécificité de substrats, de réaction catalytique, des sites de fixation ou encore des interactions moléculaires constituent des exemples de nouvelles fonctions chez les bactéries puisqu'elles diffèrent de la fonction initiale.

#### **1. Impact du transfert horizontal sur l'émergence de nouvelles fonctions**

Le transfert horizontal joue un rôle clé dans l'acquisition de nouveaux gènes dans les génomes bactériens et par conséquent dans l'acquisition de nouvelles fonctions chez les bactéries. Toutefois, le gène nouvellement acquis peut ne pas être maintenu, si ce gène ne confère aucun avantage à la cellule bactérienne receveuse. Dans ce cas, il dérivera en pseudogène par accumulation de mutations ponctuelles. Par ailleurs, si le gène nouvellement acquis est déjà présent dans le génome bactérien, cela équivaut à un événement de duplication/divergence. Dès lors, seule la voie de néofonctionnalisation pourrait amener à une nouvelle fonction dans le génome bactérien.

Dans la plupart des cas, les gènes acquis par transfert horizontal sont inactivés par des mutations ponctuelles ou perdus par délétions au cours de l'évolution. Par exemple, dans le génome de

*S. thermophilus*, 21% des gènes acquis par transfert horizontal semblent être tronqués et/ou présumés inactifs (Hols *et al.*, 2005).

Néanmoins, de nombreux exemples de transferts horizontaux conférant un avantage sélectif à la bactérie ayant acquis de nouveaux gènes sont connus. Ainsi, des événements de transfert horizontal sont responsables par exemple de la dissémination des résistances à la tétracycline et/ou à l'érythromycine chez les bactéries (Clewell *et al.*, 1995 ; Swartley *et al.*, 1993), des îlots de pathogénicité, notamment de l'îlot HPI (pour High Pathogenicity Island) de *Yersinia pseudotuberculosis* (Lesic et Carniel, 2005) ou encore des voies cataboliques du saccharose chez *E. coli* (Jahreis *et al.*, 2002).

## **2. Impact du mécanisme de duplication/divergence sur l'émergence de nouvelles fonctions**

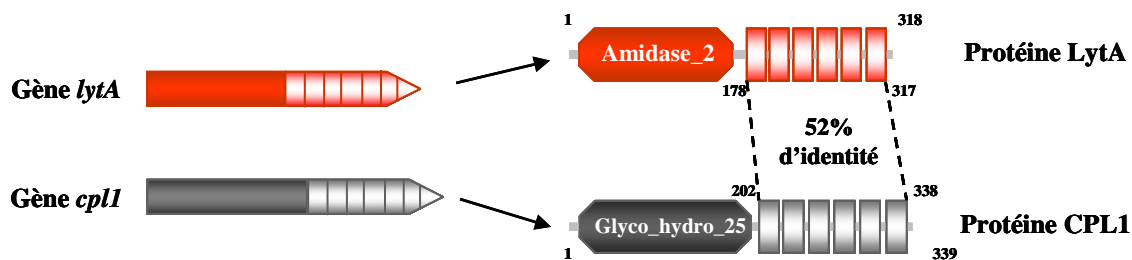
Chez les bactéries, l'acquisition de nouvelles fonctions par néo-fonctionnalisation est très peu connue dans la littérature, laissant suggérer que ce phénomène est extrêmement rare ou très peu étudié. Toutefois, des gènes responsables de nouvelles fonctions biologiques ont été récemment découverts dans les génomes bactériens grâce à l'analyse comparative des génomes et à la génomique environnementale. C'est le cas par exemple de la famille des nitrilases de l'opéron *nitC*, contenant 7 gènes qui codent des protéines impliquées dans la détoxification des antibiotiques et des nitriles (Pace et Brenner, 2001). L'analyse comparative des génomes a révélé que ni les gènes de l'opéron *nitC*, ni leurs fonctions n'étaient conservées entre les différentes espèces bactériennes. En effet, l'hypothèse de modifications de l'opéron *nitC* par des événements de duplication/divergence et de transfert horizontal a été émise pour expliquer ces divergences (Podar *et al.*, 2005). Par ailleurs, la divergence des gènes de l'opéron *nitC* s'est accompagnée par des modifications des propriétés biochimiques des nitrilases amenant de nouvelles fonctions telles que de nouvelles spécificités de substrat ou des propriétés enzymatiques. Ainsi, la néo-fonctionnalisation contribuerait positivement à l'émergence de nouvelles fonctions dans les génomes bactériens (Podar *et al.*, 2005).

## **3. Impact du réassortiment de modules sur l'émergence de nouvelles fonctions**

Le fait que le réassortiment de module contribue largement à la création de nouveaux gènes dans les génomes bactériens, nous amène à nous poser la question de savoir si ces gènes créés à partir de fragments d'ADN préexistants confèrent ou non une nouvelle fonction. Pour y répondre, la création de protéines artificielles par échange de modules a été réalisée par génie génétique sur la base de leur relation structurelle et fonctionnelle.

### 3.1 Le réassortiment de modules n'amène pas forcément à de nouvelles spécificités.

Diaz *et al.*, (1990) ont utilisé comme modèle d'étude deux hydrolases du peptidoglycane : la protéine LytA et la protéine CPL1. Ces deux enzymes présentent dans leur extrémité C-terminale, un domaine apparenté, le domaine ChBD (pour Choline Binding Domain). C'est un domaine de liaison se fixant aux résidus cholines que l'on retrouve dans les acides téichoïques et les acides lipotéichoïques de la paroi bactérienne. Et dans leur extrémité N-terminale, un domaine catalytique de spécificité de substrat qui diffère tant du point de vue de leur séquence que du point de vue de leur fonction (Figure 10).

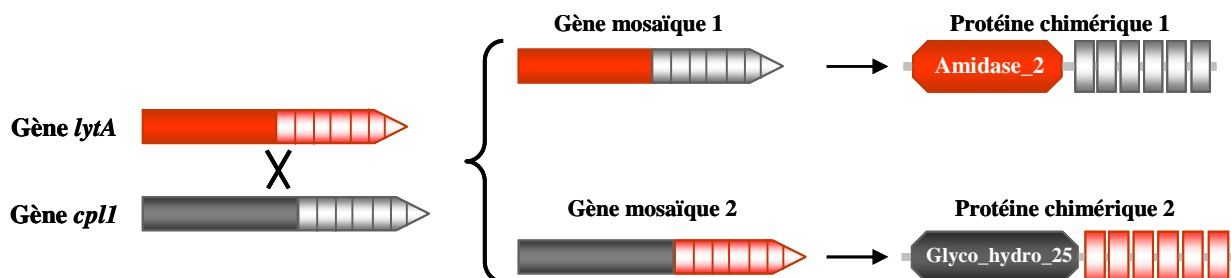


**Figure 10 :** Identité partielle des protéines modulaires LytA et CPL1 codée respectivement par les gènes *lytA* et *cplI*.

Représentation schématique des hydrolases du peptidoglycane LytA de *S. pneumoniae* et CPL1 du bactériophage Cp-1. La protéine LytA présente dans son extrémité N-terminale, le domaine catalytique amidase\_2 et dans son extrémité C-terminale, 6 exemplaires du domaine ChBD. De même pour la protéine CPL1 sauf que le domaine catalytique de celle-ci est le domaine Glyco\_hydro\_25. Le pourcentage d'identité entre les régions délimitées par des pointillés est indiqué. Les positions sont indiquées en acides aminés.

Le fait que ces deux protéines exhibent un domaine en commun, les auteurs suggéraient que les domaines pouvaient être interchangeables et que leur recombinaison pourrait créer des enzymes chimériques actives avec de nouvelles spécificités (Diaz *et al.*, 1990).

Ainsi, la construction artificielle de gènes mosaïques par fusion du module codant le domaine catalytique du gène *lytA* et du module codant le domaine de liaison à la paroi du gène *cplI*, a été réalisée et vice-versa (Figure 11).



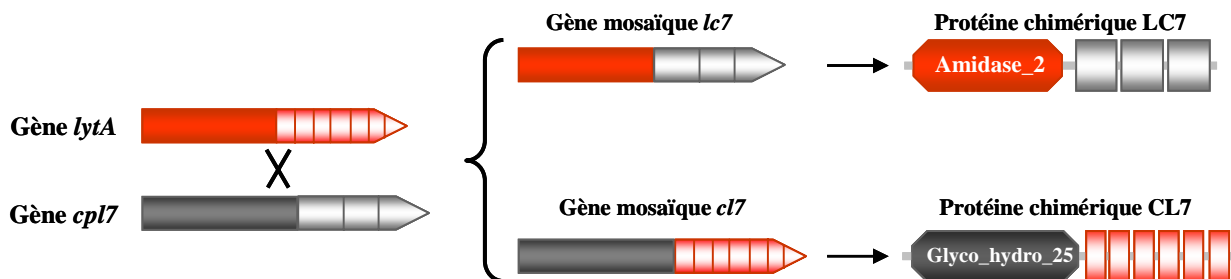
**Figure 11 :** Architecture modulaire des gènes mosaïques et des protéines chimériques résultant des échanges de modules entre le gène *lytA* de *S. pneumoniae* et le gène *cplI* du bactériophage Cp-1.

La protéine chimérique 1 possède dans son extrémité N-terminale, le domaine catalytique de la protéine LytA et dans son extrémité C-terminale, le domaine de liaison à la choline de la protéine CPL1. Inversement, la protéine chimérique 2 possède dans son extrémité N-terminale, le domaine catalytique de la protéine CPL1 et dans son extrémité C-terminale, le domaine de liaison à la choline de la protéine LytA.

Les protéines chimériques qui en résultent ont montré, d'une part que le site actif des enzymes LytA et CPL1 se situait dans le domaine N-terminal et que le domaine C-terminal était responsable de la reconnaissance à la choline. Et d'autre part, que les deux domaines étaient interchangeables et par conséquent indépendamment actifs. Toutefois, les protéines chimériques ont exprimé une activité lytique semblable à celle des protéines parentales. Ce qui signifie que cet échange n'a pas contribué à la modification ou à la création de nouvelles propriétés (Diaz *et al.*, 1990).

### 3.2 Emergence de nouvelles spécificités par réassortiment de modules

Les protéines LC7 et CL7 sont deux chimères qui ont été construites par recombinaison *in vitro*, à partir de gènes ne présentant aucune identité nucléotidique entre eux (Diaz *et al.*, 1991). Les gènes en question, *lytA* et *cpl7* codent tous deux, des enzymes capables d'hydrolyser la paroi cellulaire de *S. pneumoniae*. La protéine LytA a une activité *N*-acétylmuramoyl-L-alanine amidase, tandis que la protéine CPL7 du bactériophage Cp-7 a une activité *N*-acétylmuramidase. Les deux protéines LytA et CPL7 ont pour seul point commun, une structure modulaire. A leur extrémité, ces protéines présentent des domaines catalytiques différents, tant du point de vue de leur séquence que du point de vue de leur fonction. Leur extrémité C-terminale présente un domaine d'attachement à la paroi. Il s'agit du domaine ChBD pour la protéine LytA se fixant aux résidus cholines et du domaine Cpl-7 pour la protéine CPL7 qui ne se lie pas aux résidus cholines (Garcia *et al.*, 1990 ; Lopez *et al.*, 1997). Un remplacement réciproque des modules a été réalisé entre les gènes *lytA* et *cpl7*, conduisant à la création des gènes mosaïques *lc7* et *cl7* (Figure 12).



**Figure 12:** Création artificielle des chimères LC7 et CL7 par réassortiment de modules.

Les gènes mosaïques *lc7* et *cl7* ont été construits à partir des gènes *lytA* et *cpl7*, tous codant des protéines modulaires. Le gène *lytA* code la protéine LytA qui présente dans son extrémité N-terminale, le domaine catalytique amidase\_2 et dans son extrémité C-terminale, 6 exemplaires du domaine ChBD. Tandis que le gène *cpl7* code la protéine CPL7 qui possède un domaine catalytique de type Glyco\_hydro\_25 dans son extrémité N-terminale et 3 exemplaires du domaine Cpl-7 dans son extrémité C-terminale.

Ce réarrangement a conduit à un échange des domaines catalytiques et d'attachement.

Le gène *lc7* résulte de la fusion entre le module codant le domaine catalytique du gène *lytA* et le module codant le domaine de liaison à la paroi du gène *cpl7*. Inversement, le gène *cl7* présente quant à lui, dans son extrémité 5', le module codant le domaine catalytique du gène *cpl7* et dans son extrémité 3', le module codant le domaine de liaison dépendant de la choline du gène *lytA*. Les deux protéines chimériques qui en résultent présentent toutes les deux, une activité différente de celle des protéines d'origine. En effet, la protéine LC7 a une activité amidasique et ne se lie pas aux résidus cholines (Croux *et al.*, 1993b). En revanche, la protéine CL7 est une muramidase se fixant à la choline



(Diaz et al., 1991 ; Croux et al., 1993a). Ainsi, ces résultats montrent que de nouvelles fonctions peuvent être générées par réassortiment de modules.

#### **4. Impact de ces nouvelles fonctions sur l'adaptation des bactéries**

Les exemples décrits ci-dessus illustrent le fait que la création de nouveaux gènes ne confère pas forcément de nouvelles propriétés à la protéine. Toutefois, le fait que de nouvelles fonctions puissent être générées, nous amène à la question de savoir quel avantage adaptatif est conféré à la bactérie.

Dans la littérature, des exemples d'acquisition de nouvelles fonctions chez les bactéries sont décrits. Parmi eux, les plus étudiés sont les résistances aux antibiotiques, les facteurs de virulence (protéines de surface, toxines, îlots de pathogénicité...) dans le cas des bactéries pathogènes (Clewell et al., 1995 ; Swartley et al., 1993 ; Jenkinson et Lamont, 1997). La plupart de ces nouvelles fonctions permettent d'échapper aux systèmes immunitaires et contribuent aussi à la survie des bactéries pathogènes. C'est par exemple le cas des protéines de surface telles que la protéine PspA, les internalines ou encore les gènes de résistances aux  $\beta$ -lactamases (Hollingshead et al., 2000 ; Tsai et al., 2006 ; Thulin et al., 2006).

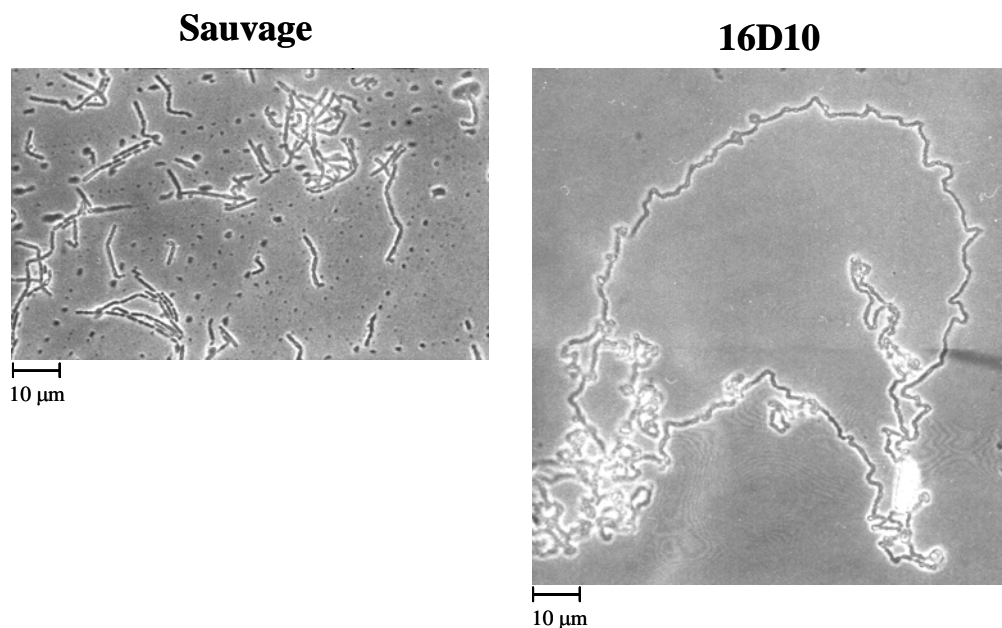
De la même façon, les bactéries non pathogènes peuvent avoir acquis de nouvelles fonctions. Il s'agit généralement de nouvelles voies métaboliques telles que la voie de biosynthèse de la méthionine chez *S. thermophilus* (Bolotin et al., 2004). Les protéines HsdS constituent un autre exemple de protéines conférant un avantage adaptatif aux bactéries qui les portent. Ces protéines appartiennent aux systèmes de restriction-modification de type I qui interviennent dans la reconnaissance d'ADN étrangers, notamment l'ADN phagique. Ces protéines confèrent un mécanisme de défense aux bactéries lors des infections phagiques (O'Sullivan et al., 2000).

Par conséquent, dans certains cas, l'émergence de nouvelles fonctions est avantageuse car elle permet la colonisation de nouvelles niches écologiques et/ou l'adaptation aux fluctuations des conditions environnementales.

#### IV. Découverte d'un gène impliqué dans la séparation des cellules de *S. thermophilus*.

Au laboratoire de Génétique et Microbiologie de Nancy-Université, l'organisme modèle d'étude est *S. thermophilus*, une bactérie à fort intérêt industriel puisqu'elle est utilisée comme ferment lactique dans la production de yaourts et de divers fromages. La thématique première de l'équipe de Nathalie Leblond-Bourget était d'étudier le mode d'adaptation de *S. thermophilus* en réponse au stress oxydatif.

Pour cette étude, une banque de 2200 mutants de la souche *S. thermophilus* CNRZ368 avait été obtenue par mutagénèse insertionnelle à l'aide du plasmide pG<sup>+</sup>host9 ::ISS1 (Thibessard *et al.*, 2002a). Parmi, ces mutants, le mutant 16D10 présentait un phénotype particulier à « longue chaîne de cellules » (Figure 13), suggérant que ce mutant était interrompu dans un gène impliqué dans le processus de la séparation des cellules filles.



**Figure 13 : Phénotype « longue chaîne de cellules » du mutant *S. thermophilus* 16D10.**

Les cellules de *S. thermophilus* ont été photographiées en microscopies à contraste de phase au grossissement 100, en phase en phase exponentielle (DO600nm = 0,6) en milieu TPPY.

L'analyse de la séquence du locus interrompu au sein du génome du mutant 16D10 avait révélé que le plasmide pG<sup>+</sup>host9 ::ISS1 interrompait un gène de fonction inconnue (Borges *et al.*, 2005). Ce gène a été nommé *cse* (pour cell segregation) pour son implication dans la séparation des cellules de *S. thermophilus*.

Bien que ces dernières années, de grands progrès aient été accomplis dans la compréhension des mécanismes de la division cellulaire, notamment chez *Escherichia coli*, *Bacillus subtilis*, *Staphylococcus aureus*, et depuis peu chez *Streptococcus pneumoniae* (Goehring *et*

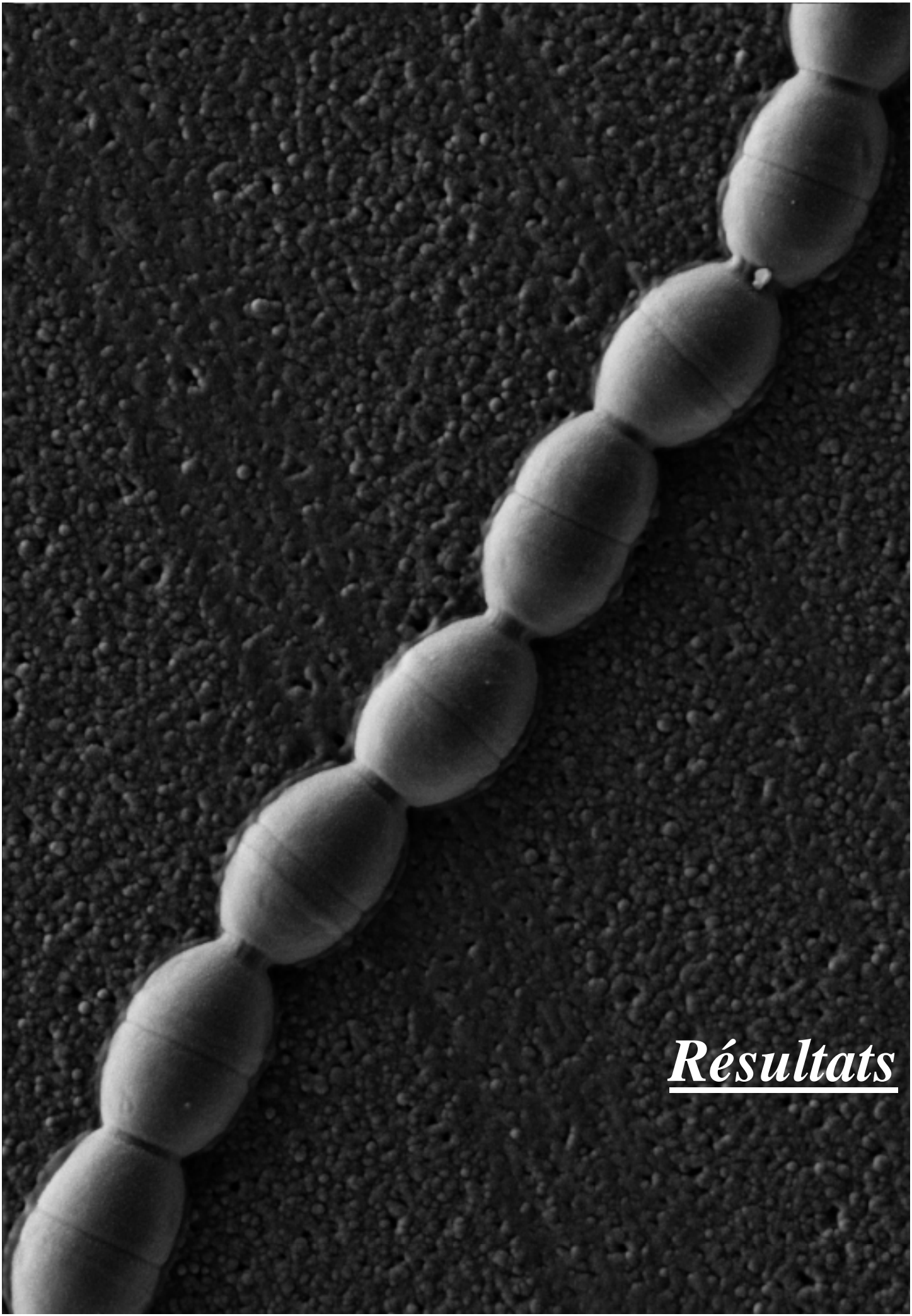
Beckwith, 2005 ; Pucci *et al.*, 1997 ; Giesbrecht *et al.*, 1998; Morlot *et al.*, 2004), le mécanisme de la séparation des cellules reste très peu connu, notamment chez *S. thermophilus*. Ceci nous a conduit ainsi à nous intéresser au gène *cse* de *S. thermophilus*.

L'analyse des données de sa séquence montre qu'il présente des caractéristiques intéressantes d'un point de vue évolutif, puisque les données préliminaires ont permis d'identifier une structure chimérique pour ce gène.

La structure chimérique du gène *cse* nous a amené à nous interroger sur le(s) mécanisme(s) moléculaire(s) ayant contribué à sa création et sur l'impact de la création de ce gène sur la valeur adaptative de *S. thermophilus*. Une recherche de la présence de ce gène chez les streptocoques de la division Viridans a aussi été menée et a contribué à « dater » l'émergence de ce gène.

Parallèlement, le phénotype « longue chaînes de cellules » du mutant 16D10, interrompu dans le gène *cse*, nous a conduit à étudier la fonction de la protéine Cse afin d'élucider le mécanisme de séparation des cellules de *S. thermophilus* et de le comparer à celui d'autres bactéries.





*Résultats*



## I. Mécanismes évolutifs conduisant à la création du gène *cse* chez *Streptococcus thermophilus*.

Publication n°1 :

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

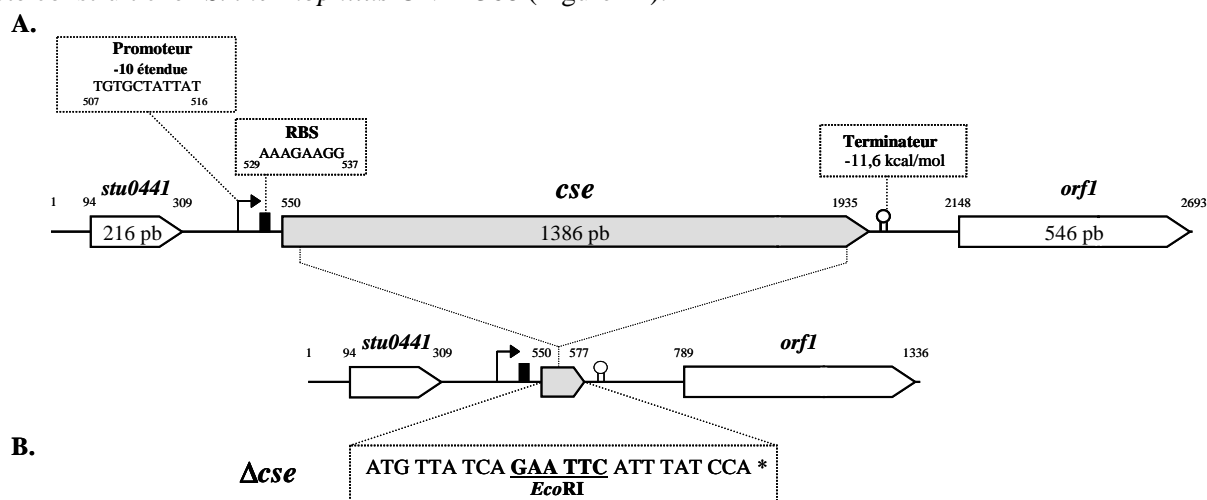
Frédéric Borges, Séverine Layec, Annabelle Thibessard, Annabelle Fernandez, Brigitte Gintz, Pascal Hols, Bernard Decaris et Nathalie Leblond-Bourget.

**Journal of Bacteriology.** Avril 2005, p. 2737-2746, Vol. 187, No. 8

L'analyse préliminaire de la séquence du gène *cse* indique que celui-ci présente des caractéristiques intéressantes d'un point de vue évolutif, puisque les premières données ont permis d'identifier une structure chimérique pour ce gène. L'ensemble de ces informations nous a donc amené à nous interroger sur le(s) mécanisme(s) moléculaire(s) ayant contribué à sa création dans le génome de *S. thermophilus*.

### 1. Le gène *cse* est impliqué dans la séparation des cellules de *S. thermophilus*.

Afin de confirmer l'implication du gène *cse* dans le processus de séparation des cellules de *S. thermophilus*, un mutant délété de la quasi-totalité de ce gène, en respectant la phase de lecture, a été construit chez *S. thermophilus* CNRZ368 (Figure 14).



**Figure 14 : Représentation schématique de la délétion en phase de la quasi-totalité du gène *cse*.**

A. L'environnement du gène *cse* de *S. thermophilus* CNRZ368 est indiqué. Le gène *cse* est précédé par un rectangle noir correspondant au site putatif de fixation des ribosomes (RBS) et d'un promoteur constitué d'une boîte -10 étendue de la position 507 à 516. Le codon d'initiation est un codon ATG. Un terminateur potentiel de transcription est situé en aval du gène *cse*. Le chiffre indique la position du gène *cse* et des ORFs adjacents. B. Délétion en phase du gène *cse*. Remplacement de la quasi-totalité de la séquence de *cse* par le site de restriction *EcoRI* dont la séquence est indiquée en gras et soulignée. Le codon stop est indiqué par une étoile.

Le mutant délété du gène *cse* de *S. thermophilus* présente un phénotype « longue chaîne de cellules » (Figure 15), argumentant en faveur de l'implication de ce gène dans le mécanisme de séparation des cellules. Le dénombrement du nombre de cellules par chaînes a été effectué afin de confirmer ces observations. Les résultats obtenus ont indiqué que les chaînes de cellules du mutant  $\Delta cse$  sont extrêmement longues, pouvant contenir plus de 3 000 cellules, alors que la souche sauvage *S. thermophilus* CNRZ368 n'en présente qu'une quarantaine en moyenne. Une complémentation avec une copie du gène *cse* porté par le plasmide pNST260<sup>+</sup> a été réalisée. Le mutant complémenté présente un phénotype « courte chaîne de cellules » semblable à celui de la souche sauvage (Figure 15), ce qui a été confirmé par comptage du nombre de cellules par chaînes (Résultats non montrés).

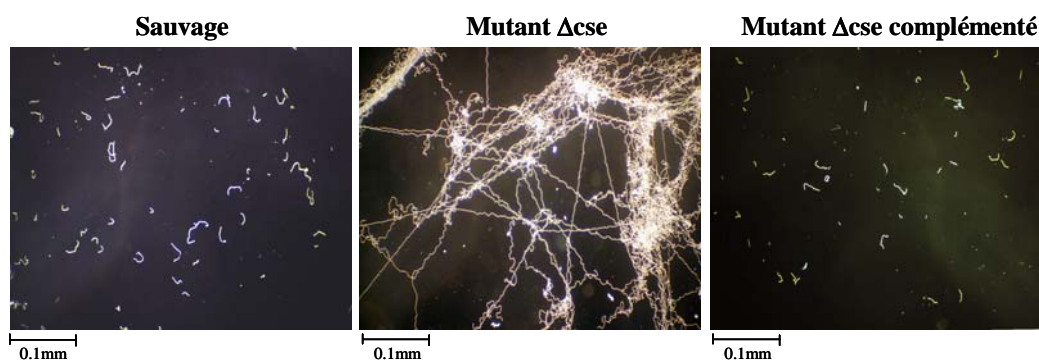


Figure 15 : Effet de la délétion du gène *cse* sur le phénotype de *S. thermophilus*.

Les cellules de *S. thermophilus* ont été photographiées en microscopies à contraste de phase au grossissement 100, en phase stationnaire de culture (20h de culture).

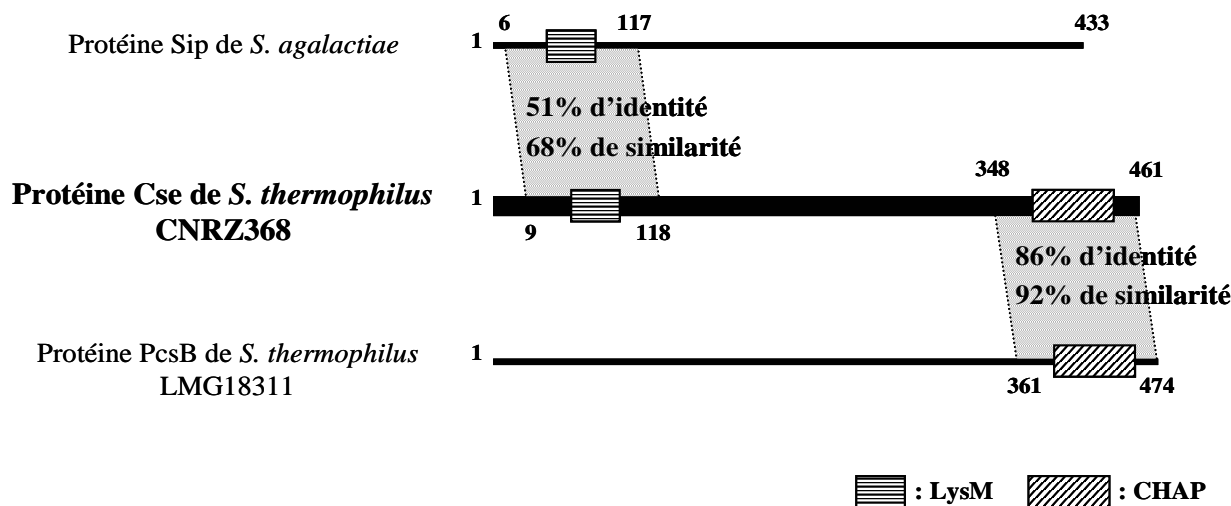
**Ainsi, l'ensemble de ces résultats démontre que le gène *cse* code une protéine impliquée dans la séparation des cellules de *S. thermophilus*. Par ailleurs, c'est le premier et unique gène de ce genre identifié à ce jour chez *S. thermophilus*.**

## 2. Le gène *cse* est un gène chimérique.

Afin de faire un bilan des connaissances concernant *cse*, une recherche des homologies en séquences par Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) de ce gène a été effectuée. L'interrogation des banques de données a révélé qu'il est absent de tous les génomes bactériens entièrement séquencés à ce jour. Ainsi, le gène *cse* est spécifique de l'espèce *S. thermophilus*.

Toutefois, l'analyse comparative de la séquence du gène *cse* de *S. thermophilus* CNRZ368 montre qu'il code une protéine dont l'extrémité N-terminale est homologue à celle de la protéine Sip (pour Surface immunogenic protein) de *S. agalactiae*, alors que son extrémité C-terminale est homologue à celle de la protéine PcsB (pour Protein for cell separation) de *S. thermophilus* (Figure 16). Les données bibliographiques indiquent que la protéine Sip est une protéine de surface de fonction inconnue (Brodeur et al., 2000 ; Rioux et al., 2001), alors que la protéine PcsB est essentielle à la croissance des streptocoques pathogènes, notamment de *S. agalactiae*, *S. mutans* et *S. pneumoniae* (Chia et al., 2001 ; Reinscheid et al., 2001 ; Reinscheid et al., 2003 ; Ng et al., 2004).





**Figure 16 : Caractère chimérique de la protéine Cse de *S. thermophilus*.**

Alignement de la séquence protéique de Cse de *S. thermophilus* CNRZ368 avec les protéines Sip de *S. agalactiae* et PcsB de *S. thermophilus* LMG18311.

La protéine Cse présente deux homologies partielles, l'une à l'extrémité N-terminale avec la protéine Sip et l'autre à l'extrémité C-terminale avec la protéine PcsB. Aucune homologie n'a été révélée dans la région centrale de Cse. Les pourcentages d'identité et de similarité en acides aminés de ces protéines avec la protéine Cse sont indiqués. Les chiffres en gras indiquent les positions protéiques du début et de la fin des homologies observées.

Par ailleurs, l'interrogation de la banque de données CDD (pour Conserved Domain Database) par le logiciel Blast disponible sur le site <http://www.ncbi.nih.gov/structure/cdd> a révélé la présence d'un domaine LysM à l'extrémité N-terminale des protéines Cse et Sip. De même qu'un domaine CHAP (pour Cysteine Histidine dependent Aminohydrolase/Peptidase) a été identifié à l'extrémité C-terminale, des protéines Cse et PcsB (Figure 16).

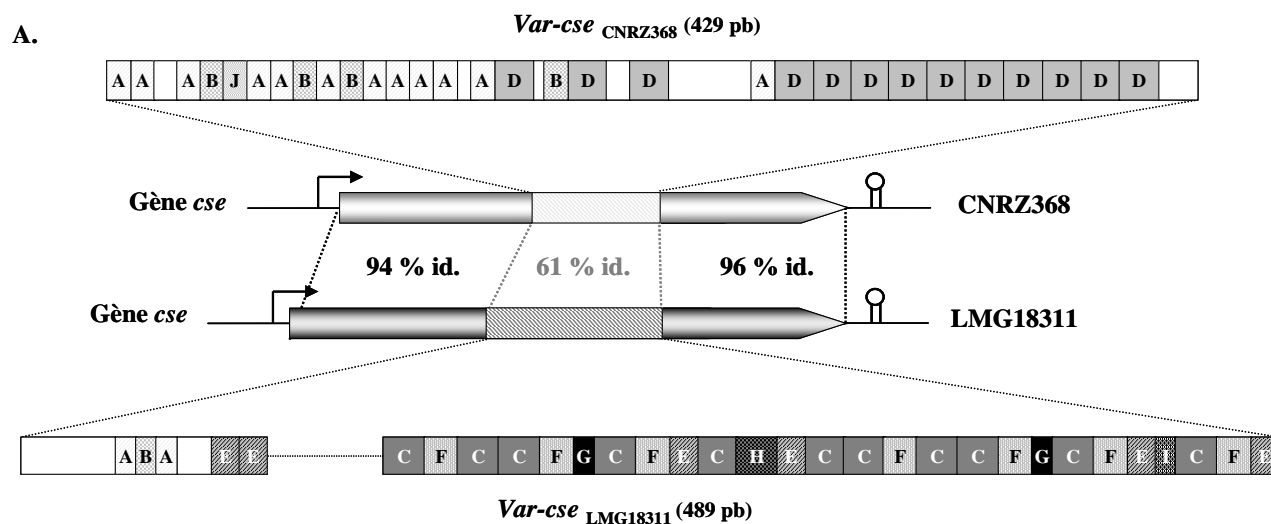
Le domaine LysM, un domaine d'attachement non covalent à la paroi cellulaire, est présent au sein de nombreuses protéines de surface (Steen *et al.*, 2003). Ce domaine est probablement responsable de la localisation préférentielle de la protéine Sip aux pôles cellulaires de *S. agalactiae* (Rioux *et al.*, 2001). Et le domaine CHAP présente une activité glutathionylspermidine amidase chez *E. coli* (Bollinger *et al.*, 1995). Ce domaine est notamment retrouvé au sein de protéines impliquées dans l'hydrolyse du peptidoglycane (Bateman et Rawlings, 2003).

**Ainsi, l'ensemble de ces résultats montre que le gène *cse* est un gène chimérique puisque son extrémité 5' est homologue à celle du gène *sip* de *S. agalactiae* et son extrémité 3' est homologue à celle du gène *pcsB* de *S. thermophilus*. Par ailleurs, ce gène serait spécifique de *S. thermophilus*.**

### 3. Le gène *cse* est un gène polymorphe.

L'analyse *in silico* de la séquence du gène *cse* des souches *S. thermophilus* CNRZ368 et LMG18311 a révélé l'existence d'une variabilité intraspécifique au niveau de la région centrale (61% d'identité) alors que les extrémités 5' et 3' sont beaucoup plus conservées avec respectivement 94% et 96% d'identité (Figure 17). La région centrale variable en séquence et en taille a été nommée « *var-cse* ».

L'analyse approfondie de la région centrale des deux allèles *cse* a révélé que les régions *var-cse* (*Var-cse*<sub>CNRZ368</sub> et *Var-cse*<sub>LMG18311</sub>) sont constituées exclusivement de motifs répétés. La nature des répétitions diffère entre les deux souches de *S. thermophilus* (Figure 17).



**B.**

| Nom      | Consensus des motifs nucléotidiques <sup>a</sup> | Consensus des motifs protéiques |
|----------|--|---------------------------------|
| <b>A</b> | GAA ACA <b>WCW</b>                               | E T (S/T)                       |
| <b>B</b> | GCA CAA GCT                                      | A Q A                           |
| <b>C</b> | TCA GAD <b>GYD CCW GMD RDM</b>                   | S (E/D) (A/V) P (A/E) (E/S/V)   |
| <b>D</b> | GAA <b>GYW CYA KYW</b> GAA                       | E (A/V) (P/L) (A/V/S) E         |
| <b>E</b> | TCA GAA GCT GYA                                  | S E A (A/V)                     |
| <b>F</b> | CAA <b>CYA GCA GCD RCA</b>                       | Q (L/P) A A (A/T)               |
| <b>G</b> | CCA GAG AGC                                      | P E S                           |
| <b>H</b> | GTA GYA CCT GTC GCA ACA                          | V (A/V) P V A T                 |
| <b>I</b> | CCR GCT <b>DYW</b>                               | P A (T/V/A/S)                   |

<sup>a</sup>W : A/T ; R : A/G ; Y : C/T ; M : A/C ; K : G/T ; D : A/G/T

**Figure 17 : Caractère polymorphique du gène *cse* chez *S. thermophilus*.**

A. Mise en évidence d'une hétérogénéité au niveau de la région centrale des gènes *cse* de souches *S. thermophilus* CNRZ368 et LMG18311. Ces régions centrales sont riches en motifs répétés. Les motifs répétés sont représentés par des rectangles contenant le nom de l'unité répétée. B. Le tableau répertorie la séquence consensus des différents types de motifs répétés.

Les motifs répétés sont tous des multiples de trois en séquence nucléotidique et sont positionnés dans la même phase de lecture les uns par rapport aux autres. Par conséquent, chaque motif nucléotidique répété correspond à un motif protéique répété (Figure 17).

Une analyse similaire de la région centrale du gène *cse* de 16 souches supplémentaires de *S. thermophilus* a été réalisée (Borges et al., 2006 ; Thèse de Frédéric Borges, 2005). Les résultats ont montré d'une part, une importante variabilité intraspécifique au niveau de cette région et d'autre part que la région *var-cse* est caractérisée par une structure en motifs répétés qui a évolué par perte d'une ou plusieurs répétitions et par duplication en tandem des motifs répétés. De plus, l'analyse des séquences de toutes les régions centrales a révélé que le polymorphisme de *var-cse* serait le résultat

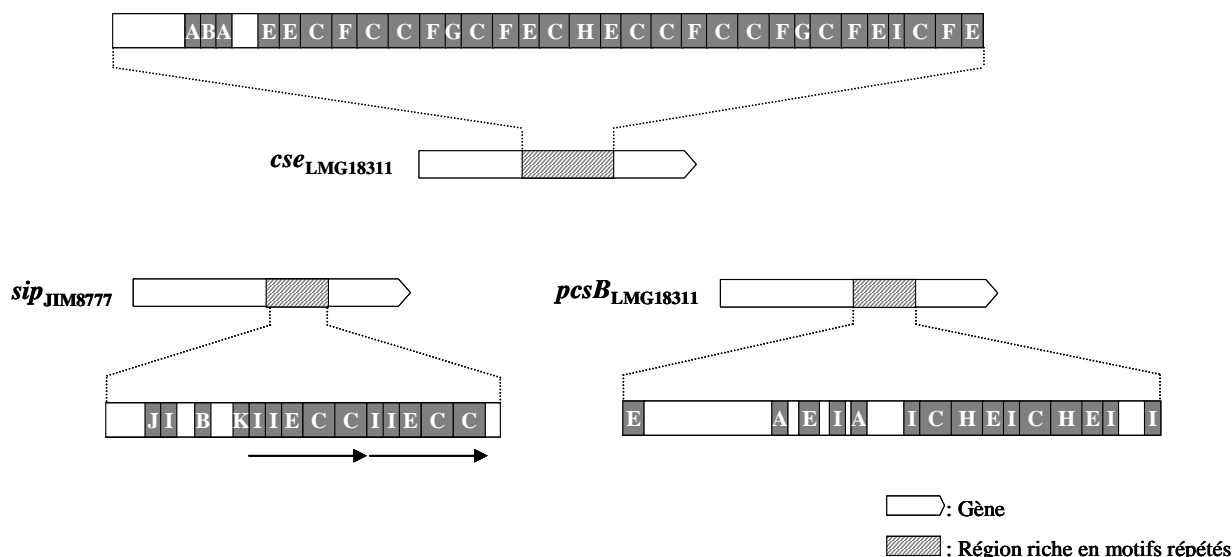
d'événements de transferts horizontaux. Ainsi, autre que son caractère chimérique, le gène *cse* est un gène mosaïque (Borges et al., 2006).

**En conclusion, outre son important polymorphisme intraspécifique, la région centrale du gène *cse* est caractérisée par une structure organisée en motifs répétés. Cette région est par conséquent une séquence de faible complexité puisque sa composition en nucléotides est biaisée par rapport à la moyenne globale.**

#### **4. Importance de cette région de faible complexité.**

Le fait que *cse* soit un gène chimérique spécifique de *S. thermophilus* et que sa région de faible complexité soit située entre les modules LysM et CHAP, suggère que cette région aurait pu favorisé le réassortiment de modules ayant contribué à l'émergence du gène *cse* chez *S. thermophilus*. Un argument en faveur de cette hypothèse est que la création de gènes nouveaux par recombinaison entre des séquences de faible complexité de gènes pré-existants, a conduit à l'émergence de protéines fonctionnelles dans les génomes bactériens (De Château et Björck, 1994 et 1996). Afin de confirmer cette hypothèse, les bornes de réassortiment de modules dans le gène *pcsB* de *S. thermophilus* LMG18311 et plus récemment dans le gène *sip* de *S. salivarius* JIM8777 dont le génome est en cours d'assemblage et d'annotation (C. Delorme, communication personnelle) ont été recherchées. *S. salivarius* est une espèce phylogénétiquement proche de *S. thermophilus* appartenant toutes deux au groupe *salivarius* (Kawamura et al., 1995 ; Poyart et al., 1998).

L'analyse de la séquence protéique de la protéine PcsB de *S. thermophilus* LMG18311 a révélé la présence d'une région centrale de faible complexité constituée de motifs répétés. Ces motifs répétés sont similaires à ceux identifiés dans la région *var-cse*. Il s'agit des motifs A, C, H, E et I (Figure 18). Par ailleurs, la séquence protéique de Sip (de 420 aa) de *S. salivarius* JIM8777 (fournie par C. Delorme, Laboratoire de Génétique microbienne de l'INRA de Jouy-en-Josas) présente également une région de faible complexité du résidu aminé 199 au résidu aminé 270. De même que les protéines Cse et PcsB, la région centrale de la protéine Sip est riche en motifs répétés commun avec ceux de la région centrale des protéines Cse et PcsB de *S. thermophilus* LMG18311. Il s'agit des motifs E, C et I (Figure 18).



**Figure 18 : Représentation schématique des régions riches en motifs répétés de *cse*, *sip* et *pcsB*.**

Les régions de faible complexité sont indiquées par des rectangles hachurés. Les régions riches en répétitions sont agrandies au-dessus et en dessous de la représentation schématique des gènes *cse*, *sip* et *pcsB*. Les motifs répétés sont représentés par des rectangles contenant le nom de l'unité répété. Les flèches indiquent une séquence constituée de 5 motifs répétés en tandem. Les séquences consensus correspondant à l'unité répétée ne sont pas indiquées.

Une particularité intéressante de la région centrale de la protéine Sip de *S. salivarius* JIM8777 est la présence d'une séquence de 5 motifs répétés en tandem qui est indiquée par des flèches sur la figure 18. La présence de cette séquence de faible complexité pourrait suggérer que cette séquence soit le point de recombinaison ayant permis le réassortiment de modules entre les gènes *pcsB* et *sip*.

Ainsi, ces résultats montrent que les gènes *sip* et *pcsB* présentent aussi une région centrale riches en motifs répétés. Par ailleurs, le fait que les régions de faible complexité se trouvent en aval du module LysM, commun entre les gènes *cse* et *sip*, et en amont du module CHAP, commun entre les gènes *cse* et *pcsB* suggèrent que ces régions soient le siège des bornes de réassortiment de modules ayant conduit à la création de *cse* chez *S. thermophilus*.

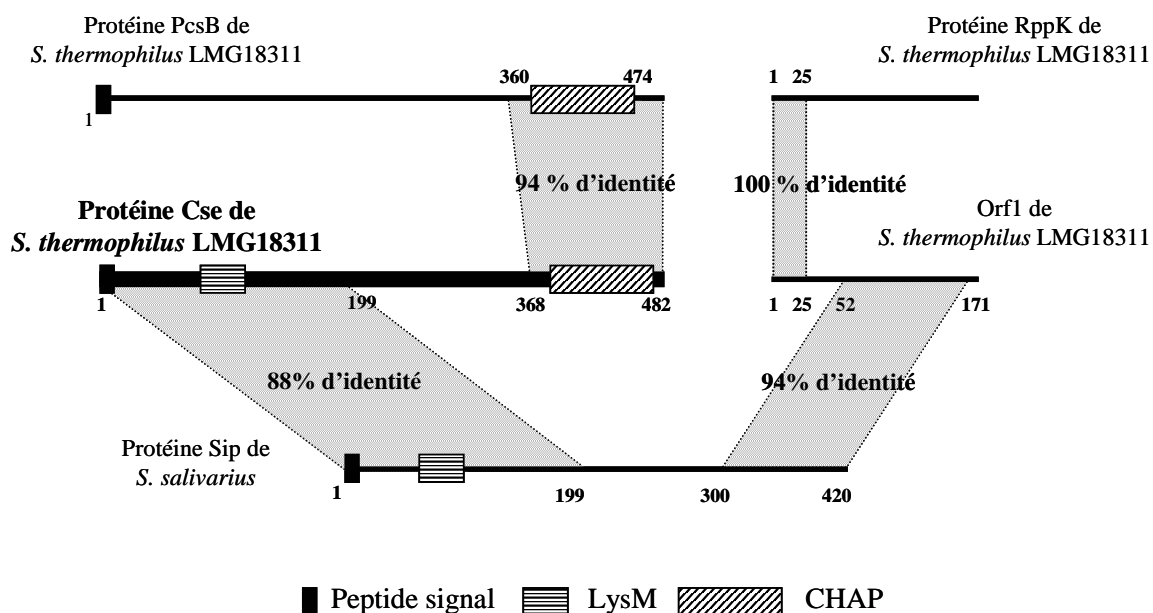
## 5. Mécanismes hypothétiques de création du gène *cse*.

Le concept d'organisation modulaire des gènes postule que de nouveaux gènes apparaissent dans les génomes bactériens par réassortiment de modules à partir de gènes pré-existants (Sheehan et al., 1997).

Le fait que Cse soit caractérisée par une structure chimérique avec le domaine LysM de la protéine Sip et le domaine CHAP de la protéine PcsB suggère que le gène *cse* ait été créé par réassortiment de modules. L'analyse de la séquence protéique de Sip de *S. salivarius* JIM8777 sur la séquence génomique de *S. thermophilus* LMG18311 a révélé deux zones d'homologies partielles. L'une correspondant à l'extrémité N-terminale de Cse avec 88% d'identité et l'autre à l'extrémité C-terminale de la protéine Orf1 avec 94% d'identité (Stu0446) dont le gène se situe en aval du gène *cse* (Figure 19). De plus, il est intéressant de noter d'importantes similitudes entre les extrémités N-terminales de la protéine Sip de *S. salivarius* JIM8777 et la protéine Cse de *S. thermophilus*

LMG18311. En effet, outre le fait qu'elles possèdent toutes les deux d'une part un domaine LysM qui est positionné au même endroit et présente 88,6% d'identité, elles possèdent d'autre part, un peptide signal dont la séquence entre ces deux protéines est 100% identique.

Ainsi, ces données suggèrent qu'au cours de l'évolution, un fragment d'ADN ait été inséré dans le gène ancestral *sip* de *S. thermophilus*.

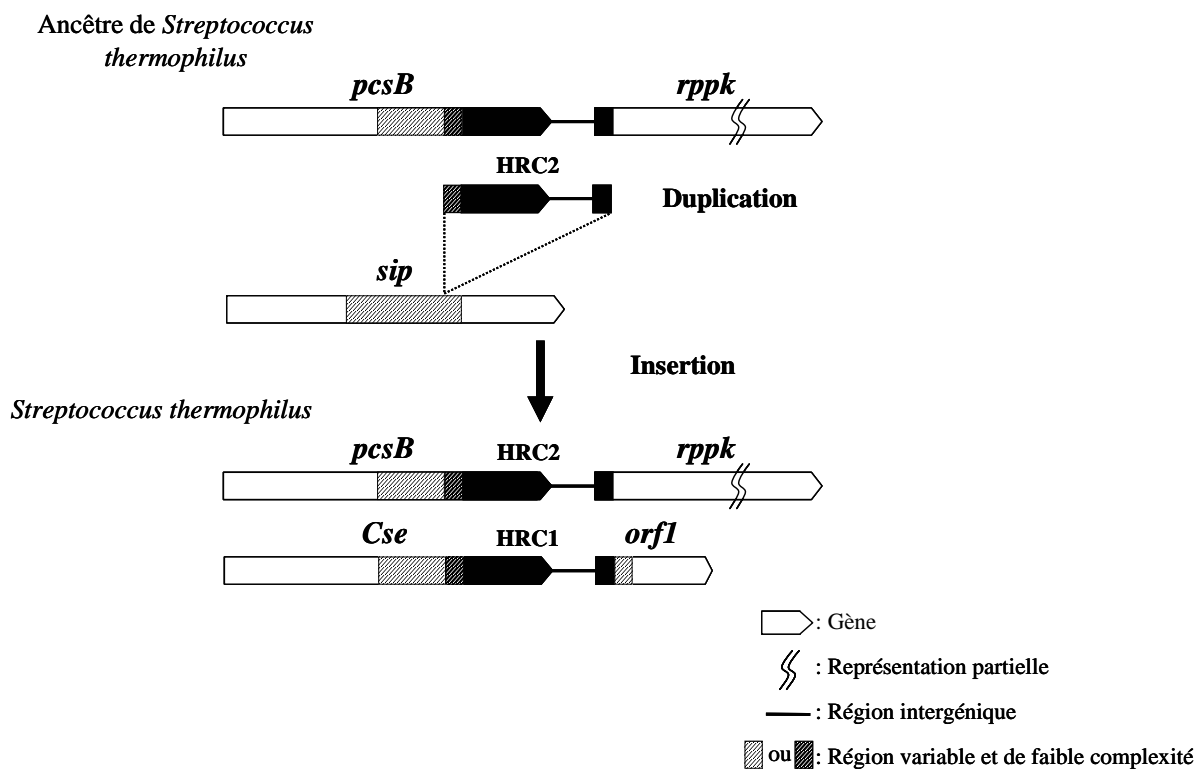


**Figure 19 :** Mise en évidence des homologies partielles entre les protéines Sip, Cse, Orf1, PcsB et Rppk. Représentation schématique de l'alignement de la séquence protéique de Sip avec celle de Cse et de Orf1. De même que les alignements de la séquence Cse avec PcsB et de la séquence Orf1 et RppK sont représentés.. Les pourcentages d'identité et de similarité en acides aminés entre ces protéines sont indiqués. Les chiffres en gras indiquent les positions protéiques du début et de la fin des homologies observées.

Une particularité intéressante de l'Orf1 de *S. thermophilus* LMG18311 est que cette protéine possède à son extrémité N-terminale, une région d'homologie partielle avec la protéine RppK (pour Ribose-phosphate pyrophosphoKinase) (Figure 19). La protéine RppK est une enzyme de biosynthèse du phosphoribosyldiphosphate, un précurseur des nucléotides puriques et pyrimidiques (Eriksen et al., 2002).

L'analyse de la séquence nucléotidique du gène *cse* avec ses régions flanquantes a révélé qu'une région de 619 pb, nommé HRC1 (pour Homologous Region Copy 1) présente 93% d'identité avec une autre région nommé HRC2 (pour Homologous Region Copy 2) du génome de *S. thermophilus* LMG18311. La région HRC1 comprend l'extrémité 3' du gène *cse*, suivie d'une région intergénique et de l'extrémité 5' du gène *orf1*, tandis que la région HRC2 est constituée de l'extrémité 3' du gène *pcsB*, de l'extrémité 5' du gène *rppk* (*prsA1*) qui le suit et de leur région intergénique. Ces résultats suggèrent que la région HRC2 a été le sujet d'un événement de duplication.

Par conséquent, l'ensemble de ces résultats nous a permis de proposer un modèle hypothétique sur la création des gènes chimériques *cse* et *orf1* qui est illustré dans la figure 20.



**Figure 20 : Modèle hypothétique sur la création des gènes chimériques *cse* et *orf1* de *S. thermophilus*.** Les régions HRC2 et HRC1 sont représentées en noir. La région HRC2 aurait été dupliquée puis insérée dans le gène *sip* ancestral par recombinaison entre les séquences de faible complexité pour créer les gènes chimériques : *cse* et *orf1*.

L'insertion du fragment d'ADN dupliqué comprenant l'extrémité 3' du gène *pcsB*, de la région intergénique et de l'extrémité 5' du gène *rppk* dans le gène *sip* aurait contribué à la scission de ce dernier pour créer deux nouveaux gènes, *cse* et *orf1* dans le génome de *S. thermophilus* (Figure 20).

Ces observations montrent que le gène *cse* est un gène chimérique qui a été créé naturellement par réassortiment de modules au cours de l'évolution de *S. thermophilus*. Ce réassortiment de modules aurait permis d'associer le module LysM de *sip* avec le module CHAP de *pcsB* par recombinaison intra- ou intermoléculaire entre les séquences de faible complexité du gène *pcsB* et du gène ancestral *sip* de *S. thermophilus*. Cette association a conduit à la formation d'un gène fonctionnel qui code la protéine majeure de la séparation des cellules de *S. thermophilus*. Par ailleurs, l'important polymorphisme intraspécifique de ce gène suggère que le transfert horizontal serait également intervenu dans sa création.

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

Frédéric Borges,<sup>1</sup> Séverine Layec,<sup>1</sup> Annabelle Thibessard,<sup>1</sup> Annabelle Fernandez,<sup>1</sup> Brigitte Gintz,<sup>1</sup> Pascal Hols,<sup>2</sup> Bernard Decaris,<sup>1</sup> and Nathalie Leblond-Bourget<sup>1\*</sup>

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,<sup>1</sup> and Unité de Génétique, Institut des Sciences de la Vie, Université catholique de Louvain, Louvain-La-Neuve, Belgium<sup>2</sup>

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  $\Delta cse$  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<sup>-1</sup>; Proteose peptone, 7 g · liter<sup>-1</sup>; yeast extract, 2 g · liter<sup>-1</sup>; lactose, 20 g · liter<sup>-1</sup>) (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  $\mu\text{g}$  of erythromycin ml<sup>-1</sup>. 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 $\alpha$  and selected on LB medium containing 100  $\mu\text{g}$  of ampicillin ml<sup>-1</sup>.

**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               |
|--|--|-----------------------------------|
| <b>Strains</b>   |  |                                   |
| <i>S. thermophilus</i><br>CNRZ368  | Wild-type strain   | INRA-CNRZ, strain collection      |
| 16D10  | Em <sup>r</sup> <i>S. thermophilus</i> CNRZ368 containing integrated pGh9:ISS1 into <i>cse</i>   | This work                         |
| CNRZ368- $\Delta$ <i>cse</i>   | <i>S. thermophilus</i> CNRZ368 with $\Delta$ <i>cse</i> mutation   | This work                         |
| CNRZ368- $\Delta$ <i>cse</i> -T  | Em <sup>r</sup> , CNRZ368- $\Delta$ <i>cse</i> with pNST260+ integrated into the chromosome  | This work                         |
| CNRZ368- $\Delta$ <i>cse</i> -C  | Em <sup>r</sup> , CNRZ368- $\Delta$ <i>cse</i> with pNST260+:: <i>cse</i> integrated into the chromosome   | This work                         |
| LMG18311   | Wild-type strain   | BCCM/LMG, strain collection       |
| LMG18311- $\Delta$ <i>cse</i>  | <i>S. thermophilus</i> LMG18311 with $\Delta$ <i>cse</i> mutation  | This work                         |
| LMG18311- <i>cse</i> <sub>varCNRZ368</sub>                               | <i>S. thermophilus</i> LMG18311 with the chimeric allele <i>cse</i> <sub>LMG18311varCNRZ368</sub>  | This work                         |
| <i>L. lactis</i> subsp. <i>cremoris</i><br>MG1363 <i>acmA</i> $\Delta$ 1 | Derivative of MG1363 wild-type strain containing 701-bp <i>SacI</i> - <i>SpeI</i> chromosomal deletion in <i>acmA</i> gene   | 6                                 |
| <i>E. coli</i><br>EC101  | <i>supE hsd-5 thi</i> $\Delta$ ( <i>lac-proAB</i> ) F' ( <i>traD6 proAB</i> <sup>+</sup> <i>lacI</i> <sup>q</sup> <i>lacZ</i> $\Delta$ M15) <i>repA</i> <sup>+</sup> , derivative of TG1 strain (47) | 23                                |
| DH5 $\alpha$   | <i>supE44</i> $\Delta$ <i>lacU169</i> ( $\phi$ 80 <i>lacZ</i> $\Delta$ M15) <i>hsdR17 endA1 gyrA96 thi-1 relA1</i>   | 47                                |
| BL21 (DE3)   | <i>hsdS gal</i> ( $\lambda$ <i>clis857 ind1 Sam7 nin5 lacUV5-T7 gene I</i> )   | 47                                |
| <b>Plasmids</b>  |  |                                   |
| pGh9   | Em <sup>r</sup> , thermosensitive replication origin from pVE6002  | 27                                |
| pGh9:ISS1  | pGh9 with ISS1   | 28                                |
| pGh9:: $\Delta$ <i>cse</i> <sub>CNRZ368</sub>                            | pGh9 with $\Delta$ <i>cse</i> and CNRZ368 surrounding regions  | This work                         |
| pGh9:: $\Delta$ <i>cse</i> <sub>LMG18311</sub>                           | pGh9 with $\Delta$ <i>cse</i> and LMG18311 surrounding regions   | This work                         |
| pGh9:: <i>cse</i> <sub>LMG18311varCNRZ368</sub>                          | pGh9 with chimeric allele <i>cse</i> <sub>LMG18311varCNRZ368</sub>   | This work                         |
| pNST260+   | pGh9 with <i>int</i> and <i>attI</i> from ICES <i>I</i>  | G. Guédon, personal communication |
| pNST260+:: <i>cse</i>  | pNST260+ with CNRZ368 <i>cse</i> ORF, its putative promoter and terminator   | This work                         |
| pFUN   | Em <sup>r</sup> and Amp <sup>r</sup> , derivative of pIL252, with $\Delta$ <sub>Sp</sub> <i>nuc</i> gene   | 38                                |
| pFUN:: <i>cse</i>  | pFUN with <i>cse</i> - $\Delta$ <sub>Sp</sub> <i>nuc</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> <sub>1-183</sub>                                     | pET15b with hexa-His-Cse <sub>1-183</sub> translational fusion   | This work                         |
| pET15b:: <i>cse</i> <sub>184-461</sub>                                   | pET15b with hexa-His-Cse <sub>184-461</sub> 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/gene/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  $\Delta$ <sub>Sp</sub>*nuc*.** 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*- $\Delta$ <sub>Sp</sub>*nuc* fusion. For pFUN:: $\Delta$ <sub>Sp</sub>*cse*, 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  $\Delta$ <sub>Sp</sub>*cse*- $\Delta$ <sub>Sp</sub>*nuc* fusion where the sequence encoding the putative signal peptide of Cse was replaced by a HindIII restriction site. Following selection in *E. coli* DH5 $\alpha$ , 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 <sup>a</sup> (5'-3')            |
|-----------------|--|
| delcse.E5'      | .....CCCCCCCCAAGCTTACAAGATCAGTTTGGAAC    |
| delcse.I5'      | .....CCCCCGAATTCTGATAACATGTATAATTCAT     |
| delcse.I3'      | .....CCCCCGAATTCATTTATCCATAATGAGTACA     |
| delcse.E3'      | .....CCCCCTGCAGCAGCCTGTGGTTGAAG          |
| fcompl          | .....TTGTTGGCGCGCTTTAGCCGATTTGGCTTTTGAAG |
| rcmpl           | .....TTGTTGGCGCGCCCATTTTCTCAGGATGAAT     |
| racse.E5'       | .....CCCCCAAGCTTTGAATGTTTGGCTAATATC      |
| racse.I5'       | .....CCCCCGATATCAATAGCGTCTGCAACT         |
| racse.I3'       | .....CCCCCGATATCAACTCCTAATACATATC        |
| racse.E3'       | .....CCCCCTGCAGTTATGGATAAATAATATAC       |
| ravar.5'        | .....CATGTTACTTCAGCAACAAC                |
| ravar.3'        | .....GCATATTCGTATGTAGCTGC                |
| fnuc.cse.5'     | .....CCCCCGAATTCTGATAGGATAGGGCTT         |
| fnuc.cse.15'    | .....CCCCCAAGCTTCATGTATAATTCATTCCTT      |
| fnuc.cse.13'    | .....CCCCCAAGCTTCATGTATAATTCATTCCTT      |
| fnuc.cse.3'     | .....CCCCCGGATCCCTGGATAAATAATATACAG      |
| sex.E5'         | .....CCCCCCCATATGTTATCAAAATCTAAAAC       |
| sex.I5'         | .....CCCCGGATCCCTAAGCTGGAGTATCTGATGC     |
| sex.I3'         | .....CCCCCCCATATGACGATACATACTGAACA       |
| sex.E3'         | .....CCCCGGATCCCTTATGGATAAATAATATAC      |

<sup>a</sup> 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)  **$\Delta cse$  complementation.** Complementation of the  $\Delta cse$  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 $\tau$ 1 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 $\tau$ 1 (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*<sub>varCNRZ368</sub> 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 $\alpha$ , 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  $\mu$ g of ampicillin ml<sup>-1</sup>. At an optical density at 600 nm (OD<sub>600</sub>) of 0.6, isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was added at a final concentration of 100  $\mu$ 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:ISSI plasmid (28). The locus disrupted by pGh9:ISSI in the 16D10 mutant was identified by cloning and sequencing the 2,697 bp flanking the disruption site. Sequence analysis revealed that pGh9:ISSI 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- $\Delta cse$  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- $\Delta cse$  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- $\Delta cse$  cells did (Fig. 1C). During exponential growth (OD<sub>600</sub> = 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- $\Delta cse$  cells formed chains containing more than 100 cells. These results show that the long-chain phenotype of the CNRZ368- $\Delta cse$  mutant is not growth phase dependent. No major modification of the cell shape or size was observed in the CNRZ368- $\Delta cse$  mutant. Moreover, mutant and wild-type strain growth curves showed no significant difference in either doubling time or in stationary phase entry OD<sub>600</sub> 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 $\tau$ 1 (G. Guédon, personal communication), a resident integrative element (7). The  $\Delta cse$  mutation was complemented in *trans* by chromosomal integration of the *cse* wild-type gene carried by the pNST260+::*cse* plasmid, resulting in the CNRZ368- $\Delta cse$ -C strain. The negative control consisted of the chromosomal insertion of empty pNST260+, resulting in the CNRZ368- $\Delta cse$ -T strain. The

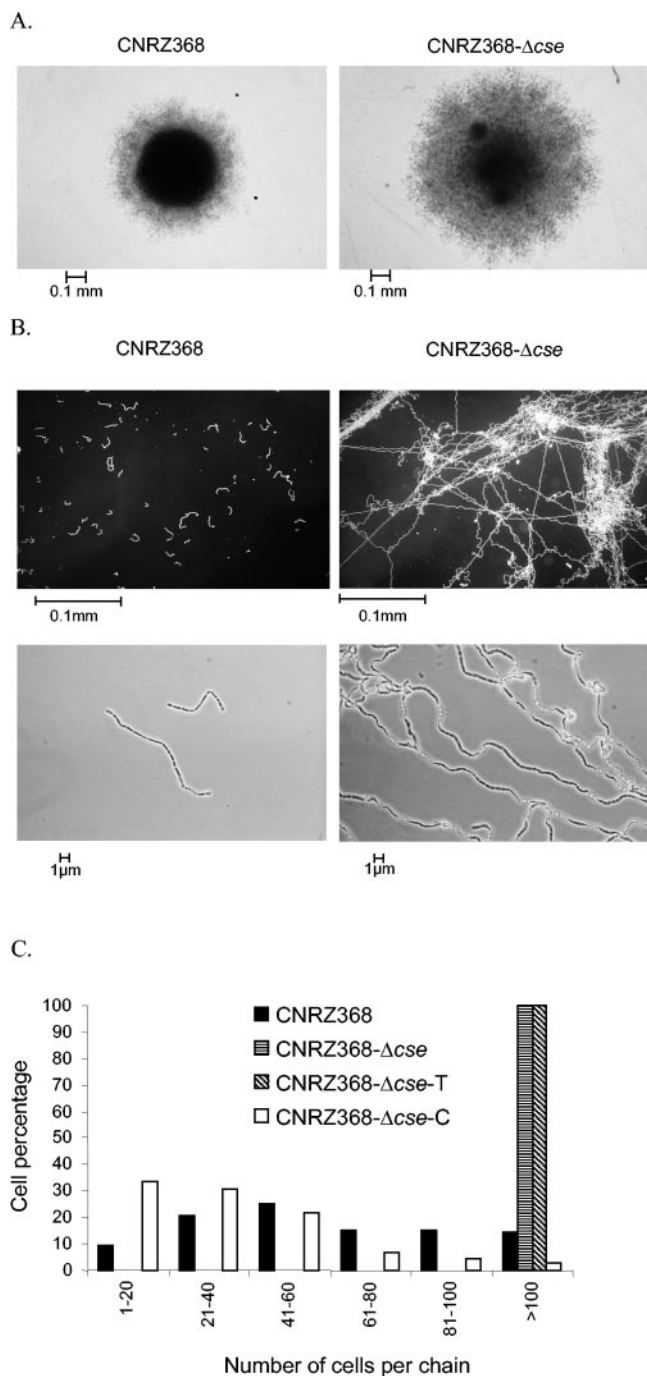


FIG. 1. Morphological consequences of *cse* deletion. (A) Colony morphology pictured after 20 h of growth; (B) cell chain morphology of *S. thermophilus* CNRZ368 strains; (C) results from counting number of cells per chain. A total of 1,000 cells were counted from three independent cultures. Cells were photographed and counted in stationary phase after 20 h of growth. After this time of growth,  $OD_{600}$  values reached  $0.66 \pm 0.074$ ,  $0.56 \pm 0.038$ ,  $0.50 \pm 0.029$ , and  $0.62 \pm 0.022$ , respectively, for strains CNRZ368, CNRZ368- $\Delta cse$ , CNRZ368- $\Delta cse$ -T, and CNRZ368- $\Delta cse$ -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*<sub>CNRZ368</sub> and *cse*<sub>LMG18311</sub>) 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*<sub>CNRZ368</sub> and 503 bp in *cse*<sub>LMG18311</sub>, 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*<sub>LMG18311</sub> is 60 bp longer than *var-cse*<sub>CNRZ368</sub>. 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<sub>CNRZ368</sub> (93%) region and the whole Var-Cse<sub>LMG18311</sub> region are predicted to adopt coil structures, suggesting that this region is a nonglobular domain. However, the confidence value of the prediction is low ( $3.56 \pm 1.79$  for Var-Cse<sub>CNRZ368</sub> and  $3.45 \pm 1.83$  for Var-Cse<sub>LMG18311</sub>, 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).

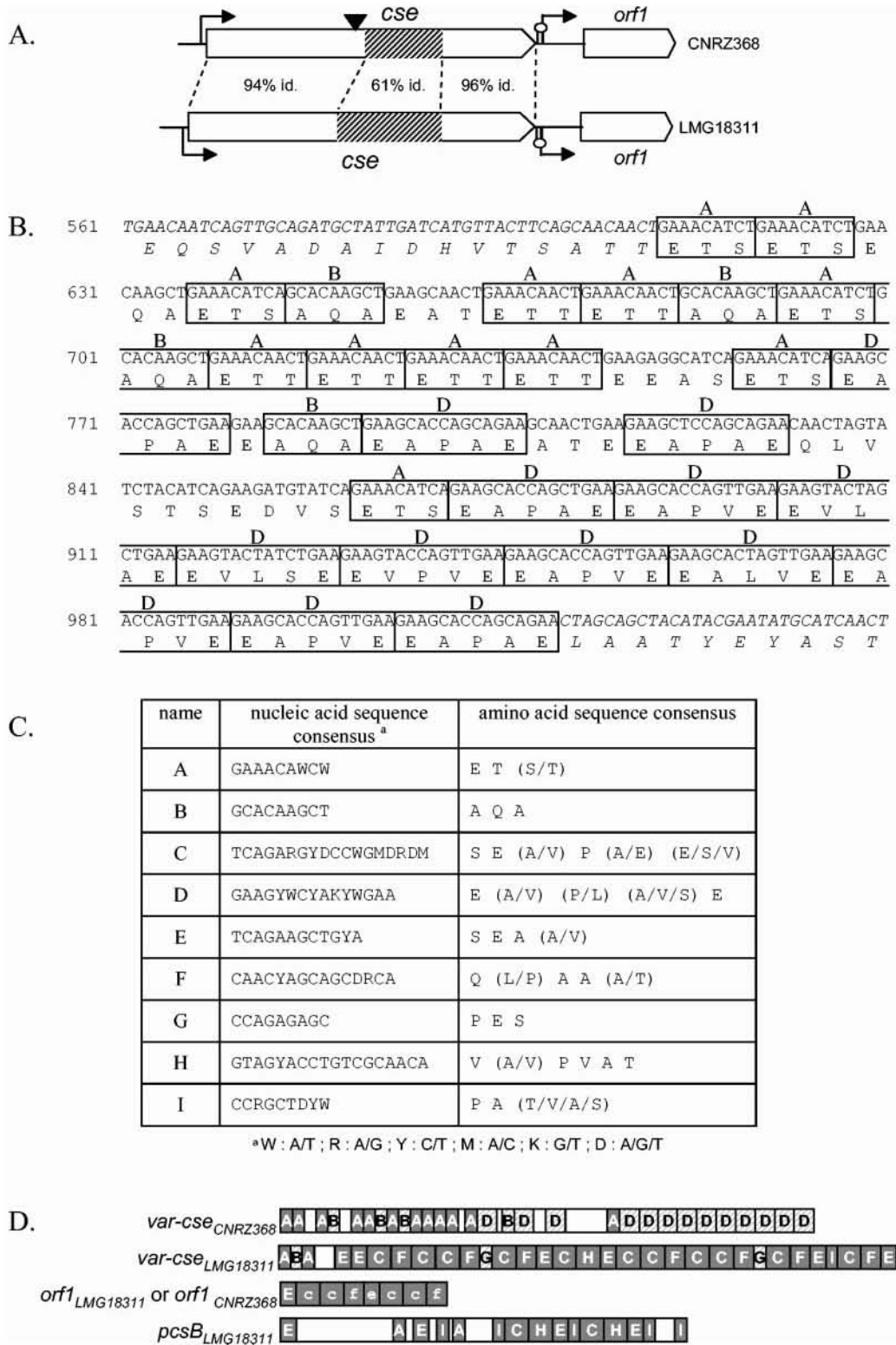


FIG. 2. *var-cse* variability and repeat content of *cse*, *pcsB*, and *orf1*. (A) Schematic representation of the *cse* locus from *S. thermophilus* strains CNRZ368 and LMG18311. Open arrows represent ORFs and indicate their reading direction, broken arrows indicate putative promoters, and hairpin loops 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*<sub>CNRZ368</sub> 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*.

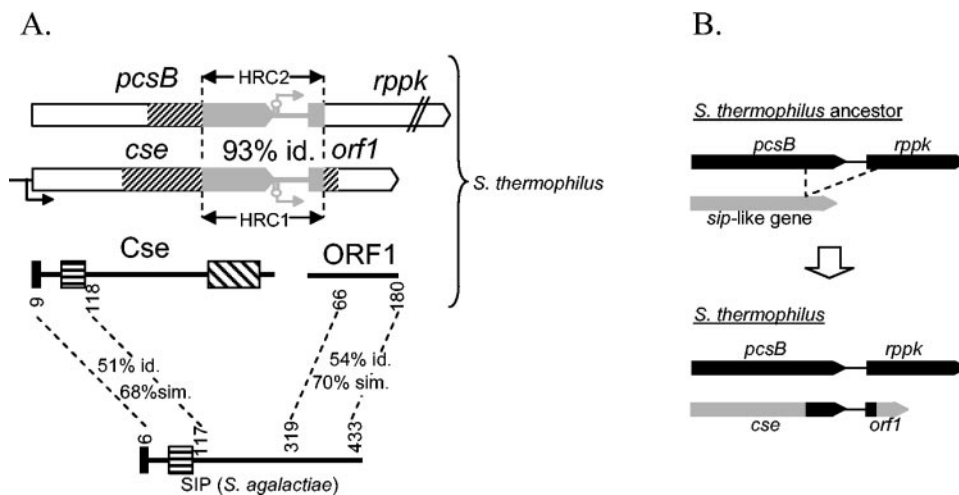


FIG. 3. *cse*, *orf1*, and their putative products are chimeric and may result from a duplication event. (A) Results from identity searches. Open arrows represent ORFs and indicate their reading directions, broken arrows indicate putative promoters, and hairpin loops symbolize putative rho-independent 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*<sub>LMG18311</sub>, *orf1*<sub>LMG18311</sub>, and *rppk*<sub>LMG18311</sub> were scanned for repeats. The analysis revealed that the *pcsB*<sub>LMG18311</sub> central region and *orf1*<sub>LMG18311</sub> 5' end also contain repeats (Fig. 3A) that are common to both *cse* alleles (Fig. 2D). More precisely, all repeats found in *pcsB*<sub>LMG18311</sub> (A, C, E, H, and I) are also represented in *var-cse*<sub>LMG18311</sub>, and one repeat (A) found in *pcsB*<sub>LMG18311</sub> is represented in *var-cse*<sub>CNRZ368</sub>. 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*<sub>LMG18311</sub> (Fig. 2D).

Interestingly, *var-cse*<sub>LMG18311</sub> and the *orf1*<sub>LMG18311</sub> 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*<sub>LMG18311</sub> 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 $\alpha$  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

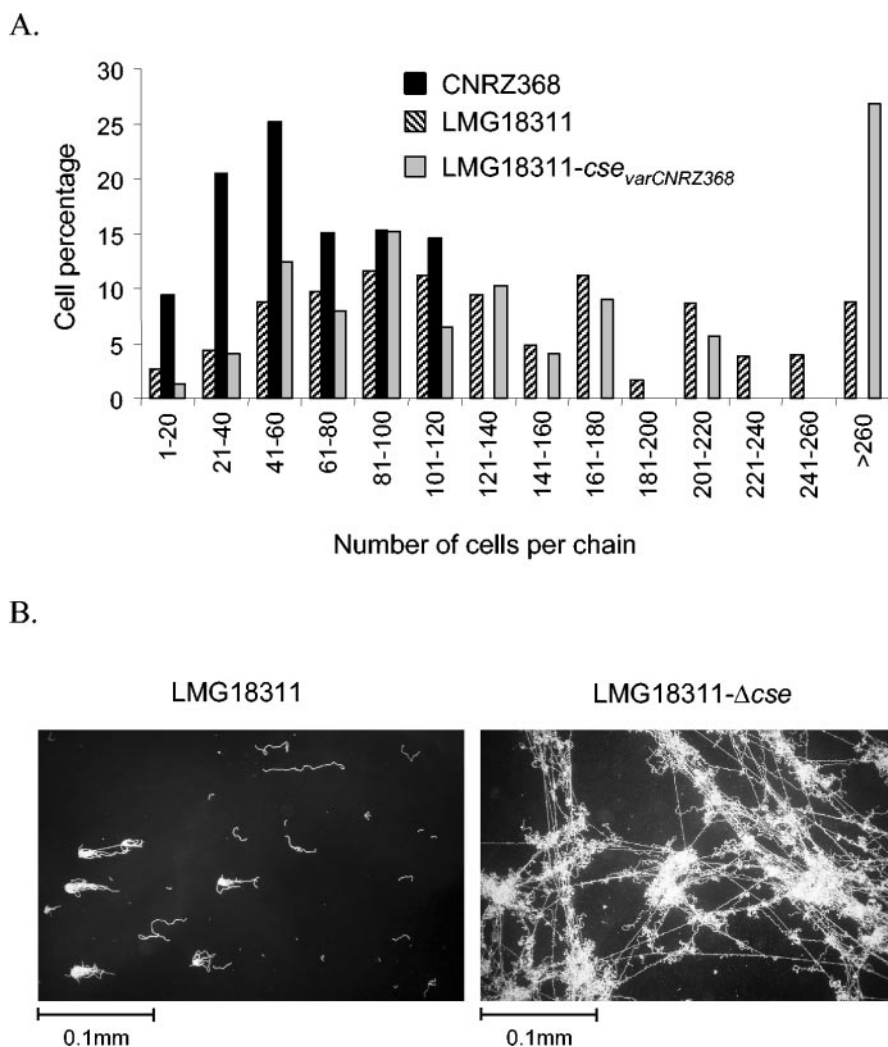


FIG. 4. Influence of *var-cse* variability on Cse cell segregation activity. (A) Results of counting numbers of cells per chain. A total of 1,000 cells were counted from three independent cultures. (B) Cell morphology of *S. thermophilus* strain LMG18311. Cells were photographed or counted in stationary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

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  $\Delta cse$  mutant of LMG18311 was found to exhibit a long-chain phenotype, as the CNRZ368- $\Delta cse$  strain does (Fig. 4B), confirming that the *cse*<sub>LMG18311</sub> 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 strain. No significant difference was seen between LMG18311 and LMG18311-*cse*<sub>varCNRZ368</sub> 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- $\Delta cse$  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 AcMB protein from *L. lactis* previously reported (16). The authors were able to circumvent this issue by overproducing AcMB 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 non-denaturing 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 µg of crude protein extract were added to a TPPY suspension of autoclaved *S. thermophilus* or *M. lysodeikticus* cells and incubated at 42°C. OD<sub>600</sub> 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*  $\Delta$ *acmA* was transformed with pNST260+::*cse*, and the number of cells per chain was counted. No difference in chain length was noticed between  $\Delta$ *acmA* and  $\Delta$ *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*  $\Delta$ *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*  $\Delta$ *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*<sub>LMG18311</sub> by *var-cse*<sub>CNRZ368</sub> 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*<sub>LMG18311</sub> by *var-cse*<sub>CNRZ368</sub> ending in a 60-bp shortening of *var-cse*, since *var-cse*<sub>CNRZ368</sub> is 60 bp shorter than *var-cse*<sub>LMG18311</sub>, 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*acm*Δ*I* 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

1. 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.
2. 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.
3. 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.
4. 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.
5. Bracqart, P. 1981. An agar medium for the differential enumeration of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* in yogurt. *J. Appl. Bacteriol.* **51**:303–305.
6. 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.
7. Burrus, V., Y. Roussel, B. Decaris, and G. Guédon. 2000. Characterization of a novel integrative element, ICESrI, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **66**:1749–1753.
8. 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.
9. 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.
10. 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.
11. Doolittle, R. F. 1995. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**:287–314.
12. 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. Piekarowicz, 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, 22<sup>nd</sup>, 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. Murl, 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. Hoerprich.** 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 PSPRED 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. **Mesnager, 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, I., 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. **Rigden, D. J., M. J. Jedrzejas, 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. **Stingele, 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* Sfi6. *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 inter-domain 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. **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**:3491–3501.
  59. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.



## II. Ce nouveau gène confère-t-il une fonction nouvelle à *Streptococcus thermophilus* ?

Article en préparation :

---

***cse*, a new gene recently created by domain shuffling has led to the emergence of a new function in *Streptococcus thermophilus*.**

Séverine Layec, Christine Delorme, Joëlle Gérard, Bernard Decaris et Nathalie Leblond-Bourget.

---

L'analyse des données de séquences du gène *cse* et de ses régions flanquantes a permis de proposer que ce gène résulte d'événements de réassortiment de modules dans le génome de *S. thermophilus*.

La recherche de *cse* au sein des génomes entièrement séquencés indique qu'il n'est présent qu'au sein de l'espèce *S. thermophilus*. Son absence dans le génome de tous les streptocoques pathogènes disponibles dans les banques de données EMBL/GenBank suggère que le réarrangement ayant conduit à la création de *cse* soit postérieur à la séparation de l'espèce *S. thermophilus* des streptocoques pathogènes. Afin de « dater » plus précisément la création du gène *cse*, sa présence a été recherchée au sein du groupe *salivarius* qui inclut les espèces *S. thermophilus*, *S. salivarius* et *S. vestibularis* (Kawamura et al., 1995 ; Poyart et al., 1998).

Par ailleurs, le réassortiment de modules à l'origine de la création de *cse* a conduit à l'association du module LysM de *sip* avec le module CHAP de *pcsB*. Ainsi, les protéines PcsB et Cse présentent le même domaine catalytique qui pourrait leur conférer une fonction cellulaire analogue. Toutefois, les extrémités N-terminales de ces deux protéines sont différentes, suggérant que Cse et PcsB puissent présenter des fonctions distinctes. Ceci nous a conduit à comparer le phénotype des mutants délétés des gènes *cse* et *pcsB*, afin de savoir si Cse et PcsB sont des analogues fonctionnels ou non, et par conséquent à déterminer si l'évènement à l'origine de la création de *cse* a conduit à l'émergence d'une protéine de fonction nouvelle chez *S. thermophilus*.

### 1. *cse*, est-il un gène de création récente ?

L'interrogation des banques de données EMBL/GenBank révèle l'absence du gène *cse* au sein de tous les génomes bactériens du genre *Streptococcus* entièrement séquencés, à l'exception des trois génomes séquencés de *S. thermophilus* LMG18311, CNZ1066 et LMD-9. Par ailleurs, la présence du gène *cse* a été retrouvée au sein de 19 souches de *S. thermophilus* testées par PCR (Borges et al., 2006). Ainsi, cette spécificité nous amène à privilégier l'hypothèse d'une acquisition récente de ce gène au sein de l'espèce *S. thermophilus*.

Toutefois, l'espèce *S. thermophilus* est phylogénétiquement apparentée aux espèces appartenant au groupe *salivarius* de la classe des « streptocoques viridans » (Sherman, 1937 ; Facklam, 2002). Ce groupe comprend deux autres espèces : *S. salivarius* et *S. vestibularis*. L'analyse comparative des ARN ribosomiques 16S et l'étude du polymorphisme intergénique du gène *sodA* de

*S. thermophilus*, *S. salivarius* et *S. vestibularis* indiquent que ces trois espèces sont étroitement apparentées génétiquement (Kawamura et al., 1995 ; Poyart et al., 1998).

Ainsi, contenu de la proximité phylogénétique de *S. thermophilus* avec ces deux espèces, on ne peut pas exclure que la formation du gène *cse* ait pu se produire antérieurement à l'émergence de l'espèce *S. thermophilus*. Afin de « dater » la création de ce gène au sein du génome de *S. thermophilus*, une recherche du gène *cse* dans les génomes de *S. salivarius* et *S. vestibularis* a été réalisée, en collaboration avec C. Delorme, du laboratoire de Génétique microbienne de l'INRA à Jouy-en-Josas.

### 1.1 Recherche *in silico* du gène *cse* dans les génomes de *S. salivarius*.

Les génomes de deux souches de *S. salivarius* JIM8777 et JIM8780 ont été séquencés et sont actuellement en cours d'assemblage et d'annotation (C. Delorme, communication personnelle). Une collaboration a été établie avec C. Delorme et P. Renault permettant d'avoir accès à ces séquences afin de rechercher *in silico* la présence du gène *cse* au sein de ces deux génomes.

Un blastX (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) de la séquence nucléotidique du gène *cse* de *S. thermophilus* LMG18311 sur les séquences génomiques des souches JIM8777 et JIM8780 de *S. salivarius* a révélé deux zones d'homologies partielles sur deux contigs différents (C. Delorme, communication personnelle). L'extrémité 3' du gène *cse* de la position 1101 à 1446 bp a été révélée sur l'un des deux contigs et correspond au domaine CHAP de la protéine PcsB de *S. salivarius* qui présente 93% d'identité avec celui de la protéine Cse de LMG18311. Sur l'autre contig, l'extrémité 5' du gène *cse* de la position 1 à 613 pb présente 87% d'identité avec celle d'une ORF<sub>x</sub> inconnue de 1260 pb codant une protéine putative de 420 aa. Le résultat de l'analyse comparative de cette ORF et de ses régions flanquantes avec l'environnement génique du gène *cse* de *S. thermophilus* LMG18311 est illustré dans la figure 21.

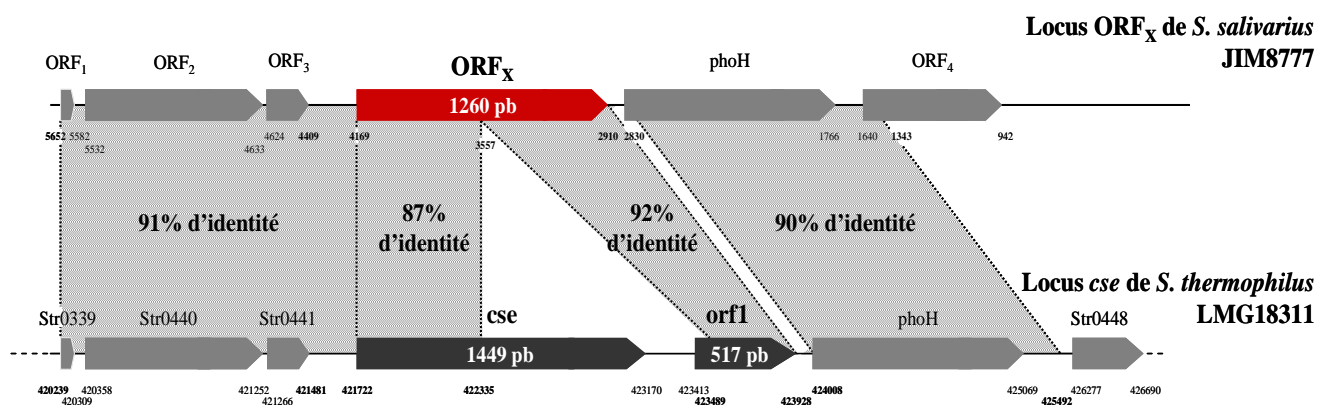


Figure 21 : Analyse de la syntonie de l'ORF<sub>x</sub> de *S. salivarius* JIM8777.

Alignement de l'ORF<sub>x</sub> et de ses régions flanquantes avec l'environnement génique du gène *cse* de *S. thermophilus* LMG18311. Les pourcentages d'identité en acides nucléotidiques de ces ORFs avec ceux du locus *cse* sont indiqués. Les chiffres indiquent les positions nucléotidiques des ORFs et ceux en gras indiquent le début et la fin des homologies observées.

L'ORF<sub>x</sub> de *S. salivarius* JIM8777 présente un environnement génique similaire à celui du gène *cse*. De plus, cette ORF présente une seconde homologie partielle (de la position 613 à 1260 pb) avec l'extrémité 3' de l'*orf1* de *S. thermophilus* LMG18311 (Figure 21). Ainsi, l'ensemble de ces données suggère que ORF<sub>x</sub> est probablement un gène *sip*. Un argument en faveur de cette hypothèse est que le BlastP (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) de la séquence protéique de cette ORF<sub>x</sub> a révélé une homologie sur toute la longueur avec la protéine Sip des streptocoques du groupe B avec 35% d'identité et un indice de confiance E-value = 3<sup>e-48</sup>. Par ailleurs, l'interrogation de la banque de données CDD par le logiciel Blast a indiqué la présence d'un domaine LysM du résidu aminé 53 à 96 à son extrémité N-terminale.

Ces résultats ont été confirmés avec ceux obtenus pour la souche *S. salivarius* JIM8780.

**Ainsi, cette recherche *in silico* montre que le gène *cse* est absent des génomes de *S. salivarius* JIM8777 et *S. salivarius* JIM8780. Par ailleurs, l'absence du gène *cse* est corrélée à la présence du gène *sip* renforçant l'hypothèse selon laquelle la création du gène *cse* résulte de l'insertion d'un fragment d'ADN au sein du gène *sip*.**

## **1.2 Recherche par PCR du gène *cse* dans le groupe *salivarius*.**

Le fait que le gène *cse* soit absent des génomes de deux souches de *S. salivarius* appartenant à la collection de souches du laboratoire de Génétique microbienne de l'INRA à Jouy-en-Josas, nous amène à la question de savoir si cet échantillon de population est représentatif de l'espèce *S. salivarius*. Par conséquent, une recherche du gène *cse*, par PCR sur colonies, a été réalisée pour 26 souches de *S. salivarius* ainsi qu'à 4 souches disponibles de l'espèce *S. vestibularis*. La présence du gène *cse* a également été recherchée au sein des génomes de 6 souches de *S. thermophilus* appartenant à la collection de souches du laboratoire de Génétique microbienne de l'INRA et considérées comme atypique de part leur origine géographique et leur niche écologique très diverses.

Ainsi, la présence du gène *cse* a été recherchée sur un total de 36 souches appartenant au groupe *salivarius*. Les souches étudiées sont référencées dans le tableau 1.

Tableau 1 : Liste des 36 souches testées par PCR

| Nom des souches               | Source  | Origine   |
|-------------------------------|---|---|
| <b><i>S. thermophilus</i></b> |   |   |
| CNR703                        | Souche sauvage isolée du lait fermenté            | Mongolie, Collection INRA <sup>a</sup> -CNRZ <sup>b</sup> |
| CNRZ759                       | Souche sauvage isolée du yaourt                   | Bulgarie, Collection INRA-CNRZ                            |
| CNRZ1202                      | Souche sauvage isolée du yaourt grec              | Grèce, Collection INRA-CNRZ                               |
| CNRZ385                       | Souche sauvage isolée du yaourt                   | Japon, Collection INRA-CNRZ                               |
| CNRZ1592                      | Souche sauvage isolée du yaourt                   | Allemagne, Collection INRA-CNRZ                           |
| JIM8232                       | Souche sauvage isolée du lait non pasteurisé      | France, Collection INRA-CNRZ                              |
| <b><i>S. salivarius</i></b>   |   |   |
| CIP53.158                     | Source non déterminée                             | Etats-Unis, Collection CIP <sup>c</sup>                   |
| CIP55.126                     | Source non déterminée                             | France, Collection CIP                                    |
| CIP102505                     | Source non déterminée                             | Etats-Unis, Collection CIP                                |
| CIP102503 <sup>T</sup>        | Souche sauvage, isolée du sang humain*            | Etats-Unis, Collection CIP                                |
| CIP1044994                    | Souche sauvage, isolée du sang humain*            | France, Collection INRA                                   |
| LMG14652                      | Souche sauvage, isolée du sang humain*            | Suède, Collection LMG <sup>d</sup>                        |
| LMG13104                      | Source non déterminée                             | Allemagne, Collection INRA                                |
| LMG13106                      | Source non déterminée                             | Royaume-Uni, Collection s LMG                             |
| LMG13108                      | Source non déterminée                             | Royaume-Uni, Collection LMG                               |
| LMG13109                      | Source non déterminée                             | Royaume-Uni, Collection LMG                               |
| JIM8421                       | Souche sauvage isolée du lait maternel            | Finlande, Collection INRA                                 |
| JIM8775                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8777                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8776                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8773                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8771                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8772                       | Souche sauvage isolée de la cavité buccale        | France, Collection INRA                                   |
| JIM8222                       | Souche sauvage isolée de la cavité buccale        | Allemagne, Collection INRA                                |
| JIM8223                       | Souche sauvage isolée de la cavité buccale        | Allemagne, Collection INRA                                |
| JIM8224                       | Souche sauvage isolée de la cavité buccale        | Allemagne, Collection INRA                                |
| JIM8221                       | Souche sauvage isolée de la cavité buccale        | Allemagne, Collection INRA                                |
| CCHSS1                        | Souche sauvage, isolée du sang humain*            | France, Hôpital Cochin                                    |
| CCHSS2                        | Souche sauvage, isolée du sang humain*            | France, Hôpital Cochin                                    |
| CCHSS3                        | Souche sauvage, isolée du sang humain*            | France, Hôpital Cochin                                    |
| CCHSS4                        | Souche sauvage, isolée du sang humain*            | France, Hôpital Cochin                                    |
| CCHSS7                        | Souche sauvage, isolée du sang humain*            | France, Hôpital Cochin                                    |
| <b><i>S. vestibularis</i></b> |   |   |
| CIP103363 <sup>T</sup>        | Souche sauvage isolée de la cavité buccale        | Royaume-Uni, Collection CIP                               |
| CIP105035                     | Souche sauvage isolée des muqueuses oropharyngées | Suède, Collection LMG                                     |
| CIP105040                     | Souche sauvage isolée des muqueuses oropharyngées | Suède, Collection LMG                                     |
| LMG14647                      | Souche sauvage isolée des muqueuses oropharyngées | Suède, Collection LMG                                     |

<sup>a</sup> Institut National de Recherche Agronomique ; <sup>b</sup> Centre National de la Recherche Zoologique ; <sup>c</sup> Collection de l'Institut Pasteur ; <sup>d</sup> Laboratoire de Génétique microbienne de l'INRA de Jouy-en-Josas ; <sup>T</sup> Souches de référence, L'astérisque indique les souches pathogènes.

La recherche du gène *cse* a été réalisée par PCR en utilisant des amorces spécifiques de *cse*, ainsi que le gène *sip* dont l'amplification sert de témoin négatif.

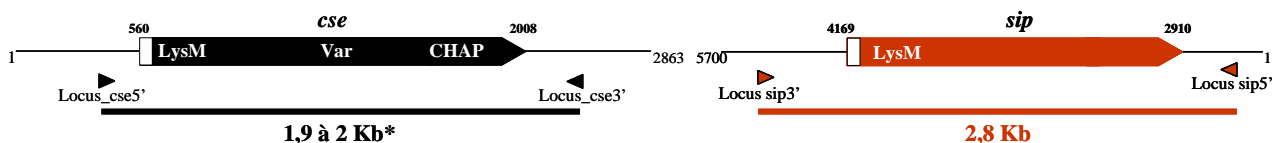
Pour ce faire, plusieurs couples d'amorces ont été testés en utilisant comme matrice, les ADN génomiques de *S. thermophilus* et *S. salivarius* (résultats non montrés) afin d'exclure toutes amorces permettant l'amplification à la fois du gène *cse* et du gène *sip* et de conserver uniquement celles permettant d'amplifier soit le gène *cse* ou le gène *sip*.

Les couples d'amorces donnant les meilleurs résultats sont le couple Locus\_cse5'/Locus\_cse3' permettant l'amplification exclusive du gène *cse* et le couple Locus\_sip5'/Locus\_Sip3' amplifiant uniquement le gène *sip* (Figure 22).

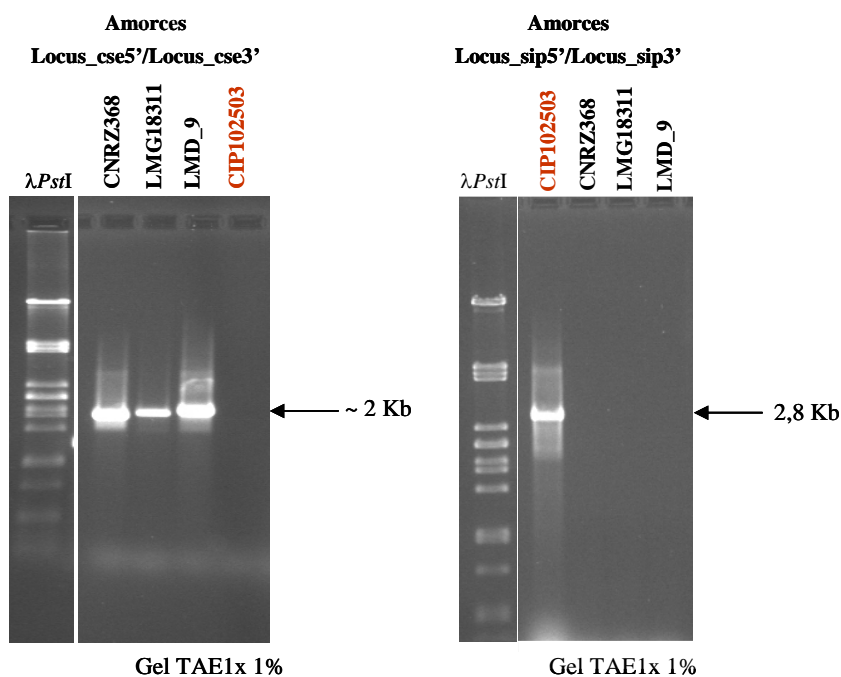
A.

**Locus *cse* de *S. thermophilus* LMG18311**

**Locus *sip* de *S. salivarius* JIM8777**



B.



**Figure 22 : Amplification spécifique du gène *cse* de *S. thermophilus* et du gène *sip* de *S. salivarius*.**

A. Position des amorces Locus\_cse5', Locus\_cse3', Locus\_sip5' et Locus\_sip3' utilisées lors de la réaction de PCR. Les modules de chacun des deux gènes sont indiqués. B. Electrophorèse des produits de PCR obtenus avec pour matrice l'ADN génomique de *S. thermophilus* CNRZ368, LMG18311, LMD-9 et *S. salivarius* CIP102503.

Les amorces Locus\_cse5', Locus\_cse3', Locus\_sip5' et Locus\_sip3' utilisées lors de la réaction de PCR ont respectivement pour séquence : 5' TTTAGCCGATTTGGCTTTTGAAG 3', 5' CCCATTATTTTCTCAGGATGAAT 3', 5' TTATTTTTTAAGATCCTCAAC 3' et 5' ATGAGAAAATCCTTTTATACTT 3'.

\* La différence de taille du produit de PCR de l'allèle *cse* entre les trois souches s'explique par l'important polymorphisme de la région centrale du gène *cse* entre ces souches.

Ces deux couples d'amorces ont été utilisés dans les réactions de PCR en utilisant comme matrices une colonie de chacune des 36 souches indiquées dans le tableau 1. Comme témoin positif de PCR, l'amplification du gène *ddlA* (459 pb) a été effectuée. Par ailleurs, le séquençage de l'extrémité 3' des produits d'amplification a été réalisé pour confirmer la nature des produits de PCR obtenus.

Un produit de PCR correspondant au locus *cse* a été détecté pour les 6 souches de *S. thermophilus* (CNRZ703, CNRZ759, CNRZ1202, CNRZ385, CNRZ1595 et JIM8232) (résultats non montrés) et confirme les résultats obtenus lors d'une étude précédente réalisée sur 19 autres souches de *S. thermophilus* (Borges et al., 2006). Tous les allèles *cse* de toutes les souches de *S. thermophilus* ont été partiellement (14/25) ou complètement séquencés (11/25). La comparaison des séquences nucléotidiques complètement séquencées a indiqué que les allèles *cse* diffèrent par la taille et exhibent une divergence en séquence allant de 3,3% à 24% (Tableau 2). Les différences de séquences sont principalement concentrées dans la région centrale nommée « var-*cse* » et peuvent atteindre un taux de divergence maximum de 50%. Au contraire, les régions flanquant la région « var-*cse* » sont bien conservées avec un maximum d'identité de 93% et 91% respectivement pour les extrémités 5' et 3' (Résultats non montrés).

Tableau 2 : Pourcentage de divergence des allèles *cse* totalement séquencés de 11 souches de *S. thermophilus*.

|                    | 368 <sup>a</sup> | 18311 <sup>b</sup> | 445 <sup>a</sup> | 19258 <sup>c</sup> | 702 <sup>a</sup> | 1066 <sup>a</sup> | 308 <sup>a</sup> | 1402 <sup>a</sup> | LMD_9         | 302 <sup>a</sup> | 8232 <sup>d</sup> |
|--------------------|------------------|--------------------|------------------|--------------------|------------------|-------------------|------------------|-------------------|---------------|------------------|-------------------|
| 368 <sup>a</sup>   |                  | 21,7%              | 9,22%            | 22,51%             | 19,65%           | 18,62%            | 20,23%           | 21,8%             | 21,8%         | 20,36%           | 21,29%            |
| 18311 <sup>b</sup> |                  |                    | 23,65%           | 15,52%             | 13,31%           | 10,64%            | 12,73%           | 3,4%              | 3,4%          | 6,67%            | 15,18%            |
| 445 <sup>a</sup>   |                  |                    |                  | 23,4%              | 21,52%           | 18,89%            | 22,61%           | <b>23,93%</b>     | <b>23,93%</b> | 22,42%           | 23,05%            |
| 19258 <sup>c</sup> |                  |                    |                  |                    | 10,38%           | 11,43%            | 12,32%           | 15,44%            | 15,44%        | 18,23%           | 14,95%            |
| 702 <sup>a</sup>   |                  |                    |                  |                    |                  | 9,74%             | 8,03%            | 13,64%            | 13,64%        | 14,05%           | 11,75%            |
| 1066 <sup>a</sup>  |                  |                    |                  |                    |                  |                   | 6,63%            | 10,62%            | 10,62%        | 10,84%           | 10,34%            |
| 308 <sup>a</sup>   |                  |                    |                  |                    |                  |                   |                  | 11,62%            | 11,62%        | 13,48%           | 14,23%            |
| 1402 <sup>a</sup>  |                  |                    |                  |                    |                  |                   |                  |                   | 0%            | <b>3,27%</b>     | 14,23%            |
| LMD_9              |                  |                    |                  |                    |                  |                   |                  |                   |               | <b>3,27%</b>     | 14,23%            |
| 302 <sup>a</sup>   |                  |                    |                  |                    |                  |                   |                  |                   |               |                  | 16,04%            |
| 8232 <sup>d</sup>  |                  |                    |                  |                    |                  |                   |                  |                   |               |                  |                   |

Les numéros en noir indiquent les noms des souches de *S. thermophilus*. <sup>a</sup>Les souches CNRZ, <sup>b</sup>La souche LMG18311, <sup>c</sup>La souche de référence ATCC19258 et <sup>d</sup>La souche JIM8777. Le pourcentage de divergence a été obtenu par alignement 2 à 2 des séquences nucléotidiques des gènes *cse* de *S. thermophilus* par le programme ClustalW du logiciel BioEdit (Thompson et al., 1994).

En revanche, la recherche des gènes *cse* et *sip* au sein des génomes de 17 souches de *S. salivarius* et de 4 souches de *S. vestibularis* a conduit exclusivement à l'amplification du locus *sip* et démontre l'absence de *cse* au sein de ces souches. Toutefois, pour 9 autres souches de *S. salivarius*, aucun produit de PCR n'a été obtenu. L'absence de produit de PCR pour ces 9 souches peut s'expliquer soit par l'absence du gène *sip* au sein de leur génome ou encore par la non hybridation des amorces Locus\_sip5' et Locus\_sip3' résultant d'un polymorphisme de séquences.

Les allèles *sip* de 9 souches de *S. salivarius* ainsi que ceux de 3 souches de *S. vestibularis* pris au hasard ont été entièrement séquencés et ont révélé moins de 13,4% de divergence (Tableau 3). Le taux

de polymorphisme du gène *sip* est compatible avec la moyenne des divergences en séquence qui avait été obtenu par la technique MLST, méthode de typage basée sur l'identification des variations nucléotidiques au sein des gènes de ménages de *S. salivarius* et *S. vestibularis* (Delorme et al., 2007).

**Tableau 3 : Résultat du pourcentage de divergence pour les allèles *sip* séquencés des 12 souches.**

|      | 8777 | 8780 | 8114  | 8223  | 8399  | 8403  | 8421  | 8771  | 8772  | 8229   | 8230   | 8405   |
|------|------|------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
| 8777 |      | 5,7% | 6,34% | 4,84% | 5,15% | 5,01% | 4,69% | 5,95% | 5,86% | 6,75%  | 7,94%  | 7,94%  |
| 8780 |      |      | 4,54% | 4,92% | 5,39% | 3,92% | 5,08% | 3,08% | 6,64% | 6,71%  | 7,74%  | 7,74%  |
| 8114 |      |      |       | 5%    | 5,47% | 2,19% | 5,08% | 3,08% | 6,64% | 9,7%   | 10,7%  | 7,74%  |
| 8223 |      |      |       |       | 0,79% | 4,06% | 0,32% | 5,62% | 2,13% | 9,29%  | 10,46% | 10,46% |
| 8399 |      |      |       |       |       | 4,37% | 0,47% | 5,62% | 2,28% | 9,21%  | 10,54% | 10,54% |
| 8403 |      |      |       |       |       |       | 4,22% | 5,4%  | 5,7%  | 8,69%  | 9,86%  | 9,86%  |
| 8421 |      |      |       |       |       |       |       | 5,39% | 1,81% | 8,90%  | 10,31% | 10,31% |
| 8771 |      |      |       |       |       |       |       |       | 6,79% | 5,93%  | 7,27%  | 7,27%  |
| 8772 |      |      |       |       |       |       |       |       |       | 13,38% | 11,71% | 11,71% |
| 8229 |      |      |       |       |       |       |       |       |       |        | 1,93%  | 2,16%  |
| 8230 |      |      |       |       |       |       |       |       |       |        |        | 0,24%  |
| 8405 |      |      |       |       |       |       |       |       |       |        |        |        |

Les numéros en vert et en rouge indiquent respectivement les noms des souches de *S. salivarius* et *S. vestibularis* de la collection JIM du Laboratoire de Génétique microbienne de l'INRA de Jouy-en-Josas. Le pourcentage de divergence a été obtenu par alignement 2 à 2 des séquences nucléotidiques des gènes *sip* de *S. salivarius* et *S. vestibularis* par le programme ClustalW du logiciel BioEdit (Thompson et al., 1994).

Par ailleurs, les résultats obtenus seraient en concordance avec les récentes analyses MLST par Delorme et al. (2007) sur la structure de la population et la diversité génétique au sein du groupe *salivarius*. Les auteurs ont révélé que les deux espèces *S. salivarius* et *S. vestibularis* présentent une divergence intraspécifique plus importante que l'espèce *S. thermophilus* suggérant que *S. thermophilus* représente une population clonale qui a récemment émergé du groupe *salivarius* (Delorme et al., 2007).

**Ainsi, l'ensemble de ces résultats montre que la présence du gène *cse* est restreinte à l'espèce *S. thermophilus*. Par ailleurs, les récentes données concernant l'émergence de *S. thermophilus* (Delorme, 2007), nous permet d'imaginer que la création du gène *cse* aurait été concomitante avec l'apparition de cette espèce dans le groupe *salivarius*. Ces données permettent ainsi de « dater » la création du gène *cse* chez *S. thermophilus* qui est estimé à près de 7000 ans quand les activités de fabrication de produits laitiers fermentés de l'homme ont commencé (Bolotin et al., 2004 ; Fox , 1993).**

**En conclusion, le gène *cse* est le premier exemple de création récente d'un gène chimérique fonctionnel apparu au cours de l'évolution du génome de *S. thermophilus*.**

## 2. Les protéines Cse et PcsB, ont-elles des fonctions cellulaires identiques ?

Le gène *cse* créé par réassortiment de modules est présent dans le génome de toutes les souches de *S. thermophilus*. Ainsi, une question d'importance se pose alors à savoir quelle pression de sélection a pu s'exercer ou s'exerce pour permettre le maintien de ce nouveau gène chez *S. thermophilus*. La réponse à cette question revient dans un premier temps à rechercher si ce gène nouveau confère ou non une nouvelle fonction à *S. thermophilus*.

L'extrémité 3' du gène *cse* est homologue à celle du gène *pcsB*, ayant en commun le module codant le domaine CHAP, porteur de l'activité catalytique de la protéine (Borges et al., 2005). Ainsi, nous pouvons émettre l'hypothèse d'une éventuelle redondance fonctionnelle des protéines Cse et PcsB, partageant le même domaine catalytique. Cependant, l'extrémité N-terminale de ces deux protéines ne présente pas d'identité de séquences significatives, suggérant que les protéines Cse et PcsB pourraient avoir une fonction différente chez *S. thermophilus*. Si tel est le cas, la création du gène *cse* aurait contribué à l'émergence d'une fonction nouvelle chez *S. thermophilus*.

Afin d'étayer cette dernière hypothèse, la construction d'un mutant délété du gène *pcsB* et d'un double mutant  $\Delta\Delta cse-pcsB$  de *S. thermophilus* a été entreprise pour analyser le phénotype associé et de le comparer avec celui du mutant  $\Delta cse$  de *S. thermophilus*.

### 2.1 Analyse du locus *pcsB* de *S. thermophilus* LMG18311.

Le locus *pcsB* de *S. thermophilus* LMG18311 dont la séquence génomique est disponible dans la banque de données EMBL/GenBank (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>) a été analysé. L'environnement génique de l'ORF *pcsB* est illustré dans la figure 23A.

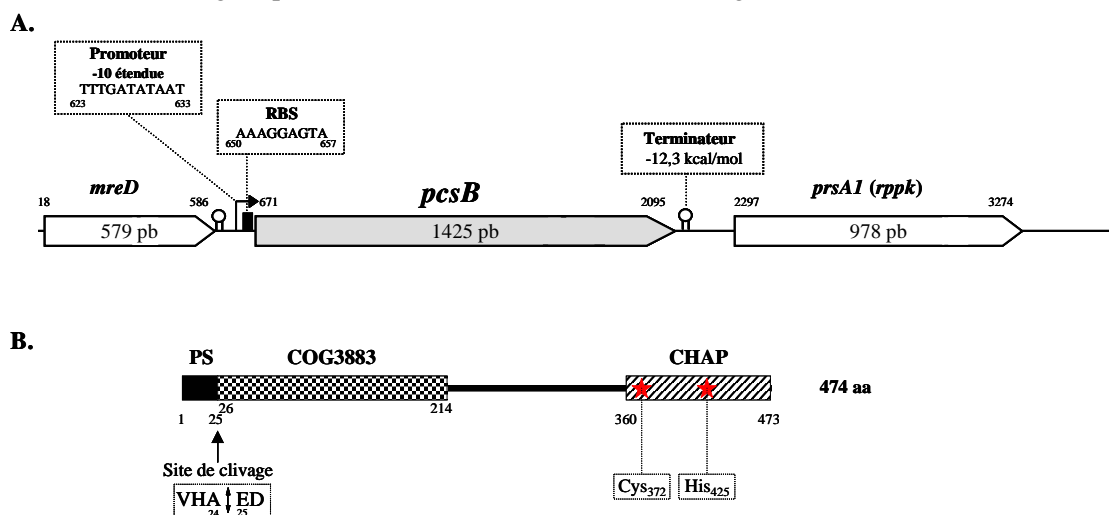


Figure 23 : Représentation schématique de l'environnement génique de l'ORF *pcsB* et de la structure protéique de PcsB de *S. thermophilus* LMG18311.

A. Environnement génique du gène *pcsB* de *S. thermophilus* LMG18311. Le gène *pcsB* est précédé par un RBS potentiel et par un promoteur constitué d'une boîte -10 étendue. Le codon d'initiation est un codon ATG. Un terminateur de transcription potentiel est situé en aval du gène *pcsB*. La position des gènes et de leurs signaux de transcription et de traduction est indiquée par les chiffres. B. Organisation modulaire de la protéine codée par le gène *pcsB*. La position de la séquence signal (PS), de la région conservée et du domaine CHAP est indiquée par les chiffres. De même que le site de clivage reconnu par la peptidase signal qui hydrolyse la liaison peptidique entre les résidus 24 et 25. Les résidus du site actif du domaine CHAP sont représentés par une étoile rouge.



L'ORF *pcsB*, de 1425 pb, est précédée d'un site putatif de fixation de ribosomes (RBS), de séquence AAAGGAGTA. Les signaux d'initiation et de terminaison de la transcription ont été identifiés. En amont de *pcsB*, un promoteur potentiel de transcription constitué d'une boîte -10 étendue a été retrouvé. De même que le terminateur de transcription potentiel constitué d'une séquence pouvant adopter une structure tige-boucle et d'une succession de 9 résidus thymine a été retrouvé 14 pb en aval du codon stop de l'ORF *pcsB*. Ces données de séquences suggèrent que l'ORF *pcsB* serait transcrite de façon monocistronique.

Cette ORF code une protéine putative de 474 aa (Figure 23B) dont la masse théorique est 48,1 kDa et le pI théorique de 4,55. L'analyse de la séquence protéique codée par *pcsB*, réalisée à l'aide du logiciel SignalP3.0 server disponible sur le site <http://www.cbs.dtu.dk/services/SignalP>, révèle la présence, à son extrémité N-terminale, d'une séquence signal potentielle de 25 acides aminés. Le logiciel SignalP3.0 prédit également l'existence d'un site de clivage entre les résidus 24 et 25. La peptidase signal reconnaît la séquence Val-His-Ala-Gln-Asp et hydrolyse la liaison peptidique entre Ala et Gln (Figure 23B) (Nielsen *et al.*, 1997).

L'interrogation de la banque de données CDD (pour Conserved Domain Database) par le logiciel Blast (<http://ncbi.nih.gov/structure/cdd>) révèle à son extrémité N-terminale, la présence d'une région conservée nommée COG3883 de fonction inconnue, en aval de la séquence signal de la position 26 à 214. Cette région s'aligne de façon fiable (Expect  $1^{e-27}$ ) avec 38% d'identité en acides aminés avec la séquence COG3883. Et à son extrémité C-terminale, le domaine CHAP du résidu 360 à 473 avec un indice de confiance de  $3^{e-14}$  et s'aligne avec 37,8% d'identité en acides aminés de la séquence pfam05227 (CHAP).

**Ainsi, l'analyse *in silico* de la séquence du gène *pcsB* de *S. thermophilus* suggère que celui-ci coderait une protéine extracellulaire pouvant également être impliquée dans l'hydrolyse du peptidoglycane via son domaine CHAP.**

## 2.2 *cse* et *pcsB*, deux gènes de fonction identique ou différente ?

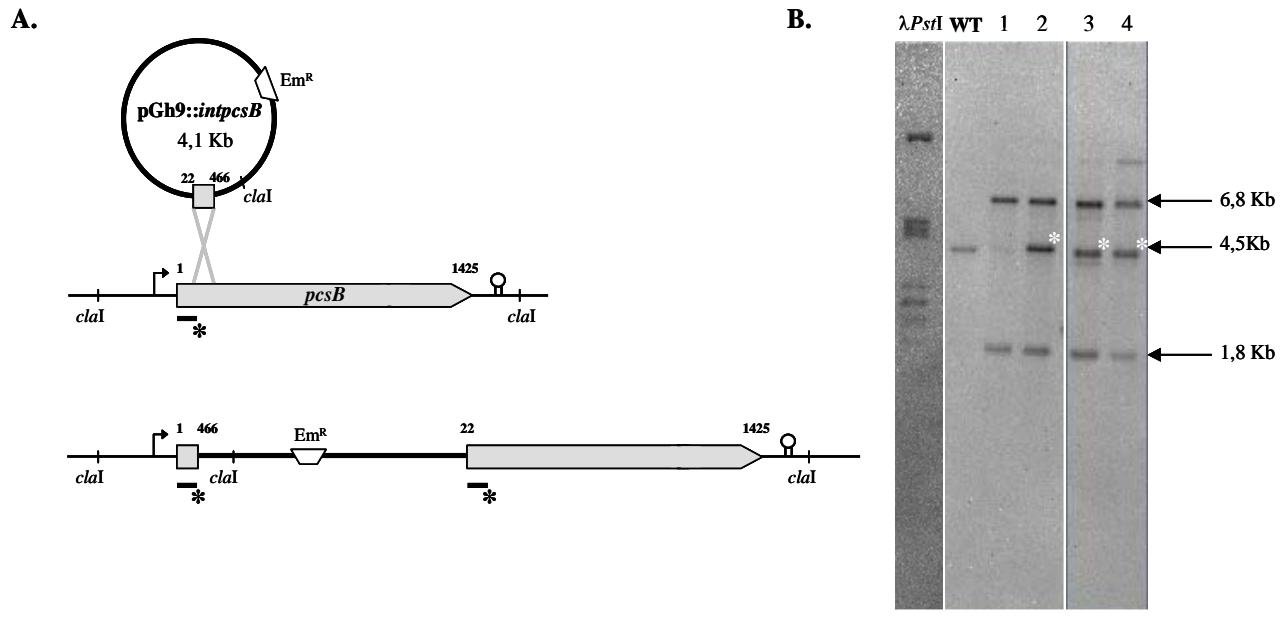
### a. Inactivation du gène *pcsB* chez *S. thermophilus*.

Afin de savoir si le gène *cse* confère ou non une fonction nouvelle à *S. thermophilus*, la construction d'un mutant délété de la quasi-totalité du gène *pcsB* en respectant la phase de lecture a été entreprise chez la souche *S. thermophilus* LMG18311 et le mutant LMG18311 $\Delta$ *cse*, et ceci par un remplacement allélique du gène *pcsB* avec la mutation  $\Delta$ *pcsB* portée par le plasmide pGh9 :: $\Delta$ *pcsB*.

Néanmoins, malgré de nombreuses tentatives dans différentes conditions testées, aucun mutant présentant l'allèle  $\Delta$ *pcsB* au locus chromosomique n'a été obtenu pour *S. thermophilus* LMG18311 et LMG18311 $\Delta$ *cse*, suggérant que le gène *pcsB* joue un rôle essentiel pour la croissance de *S. thermophilus* et que sa délétion serait létale.

Ainsi, la construction d'un mutant présentant une interruption du gène *pcsB* a été envisagée comme stratégie alternative. L'interruption de *pcsB* a été réalisée par insertion du plasmide pGh9 ::*intpcsB* (Figure 24A). Ce plasmide porte une région interne spécifique du gène *pcsB* de *S. thermophilus*, ceci afin d'éviter toute interruption non spécifique et notamment dans le gène *cse*.

Ainsi l'insertion de ce plasmide se fait par recombinaison homologue entre la séquence interne du plasmide et celle du gène *pcsB* sur le chromosome de *S. thermophilus*.

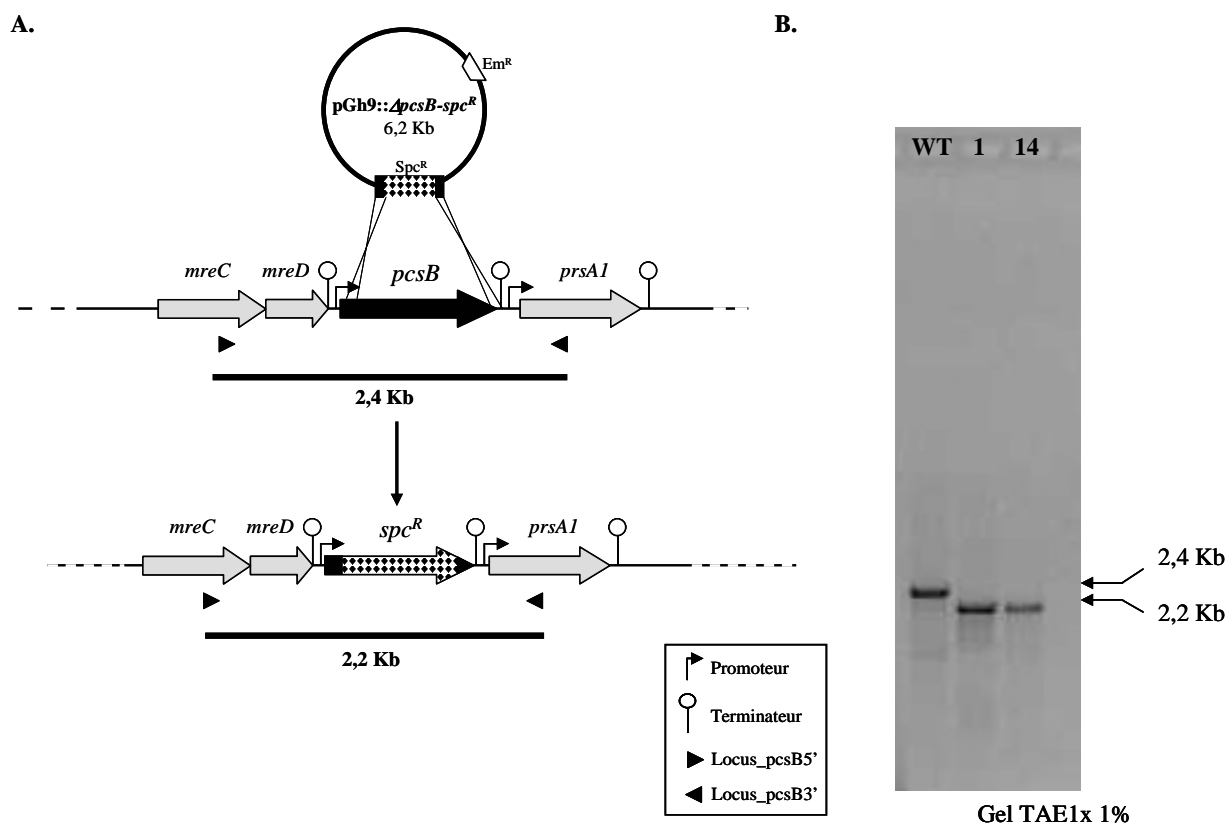


**Figure 24 : Interruption du gène *pcsB* chez *S. thermophilus* LM18311.**

A. Représentation schématique du plasmide pGh9::intpcsB, du gène *pcsB* et de l'interrupteur. B. Profil d'hybridation de l'ADN génomique de la souche LMG18311 (WT) et les interruptants de LMG18311pGh9::intpcsB (pistes n°1 et 2) et de LMG18311ΔcsepGh9::intpcsB (pistes n°3 et 4) digéré par l'enzyme de restriction *clalI* avec la sonde intpcsB. La sonde correspond à la séquence interne (de la position 22 à 466) spécifique du gène *pcsB*. Elle est indiquée par un trait noir-étoile sur la partie A. De même que les sites de restriction de l'enzyme *clalI*. L'insertion en tandem du plasmide a été constatée sur les profils d'hybridation. La bande correspondante est indiquée par une étoile blanche.

Parallèlement, une stratégie alternative pour l'obtention d'un mutant Δ*pcsB* a été entreprise et consiste au remplacement allélique de la quasi-totalité du gène *pcsB* en respectant la phase de lecture par une cassette de résistance à la spectinomycine. Pour ce faire, le plasmide pGh9::Δ*pcsB*-*spc<sup>R</sup>* a été construit. Ce plasmide porte un gène de résistance à la spectinomycine (*spc<sup>R</sup>*) incluant ses signaux de transcription, et de part et d'autre de ce gène, des séquences qui sont homologues aux régions flanquantes du gène *pcsB* à déléter (Figure 25A).

Par conséquent, la transformation de *S. thermophilus* LMG18311 et LMG18311Δ*cse* indépendamment avec les plasmides pGh9::intpcsB et pGh9::Δ*pcsB*-*spc<sup>R</sup>* a été réalisée, suivie de la sélection des clones résistants à l'érythromycine pour les interruptants et des clones résistants à la spectinomycine pour la délétion du gène *pcsB*.



**Figure 25 : Délétion du gène *pcsB* de *S. thermophilus* LMG18311.**

A. Représentation schématique de la stratégie de remplacement du gène *pcsB* par le gène de résistance à la spectinomycine (*spc<sup>R</sup>*). B. Electrophorèse des produits de PCR réalisés sur l'ADN génomique de la souche LMG18311 et des mutants LMG18311 $\Delta pcsB$  obtenus. Les produits ont été générés à l'aide des amorces Locus\_*pcsB5'* et Locus\_*pcsB3'*. La position des amorces ainsi que la taille des produits de PCR attendue, sont indiquées dans la partie A.

L'insertion du plasmide pGh9 ::*intpcsB* dans le gène *pcsB* a été vérifiée par la méthode du Southern-blotting sur un profil d'hybridation d'ADN génomique de *S. thermophilus* digéré par l'enzyme de restriction *ClaI* avec la sonde *intpcsB* (Figure 24A). Des mutants interrompus dans le gène *pcsB* présentant le profil d'hybridation attendu ont été obtenus pour *S. thermophilus* LMG18311 et LMG18311 $\Delta cse$  (Figure 24B).

Par ailleurs, la stratégie de remplacement allélique du gène *pcsB* par la cassette de résistance à la spectinomycine a permis l'obtention de deux clones à partir de la souche *S. thermophilus* LMG18311 (Figure 25B). Les mutants LMG18311 $\Delta pcsB$  présentent un allèle mutant constitué des 15 premiers et des 12 derniers codons du gène *pcsB*, eux-mêmes flanquant le gène de résistance à la spectinomycine. Afin de s'assurer de la délétion, l'allèle porteur du remplacement pour les deux mutants a été séquencé et n'a présenté aucune mutation ni de décalage du cadre de lecture (Résultats non montrés).

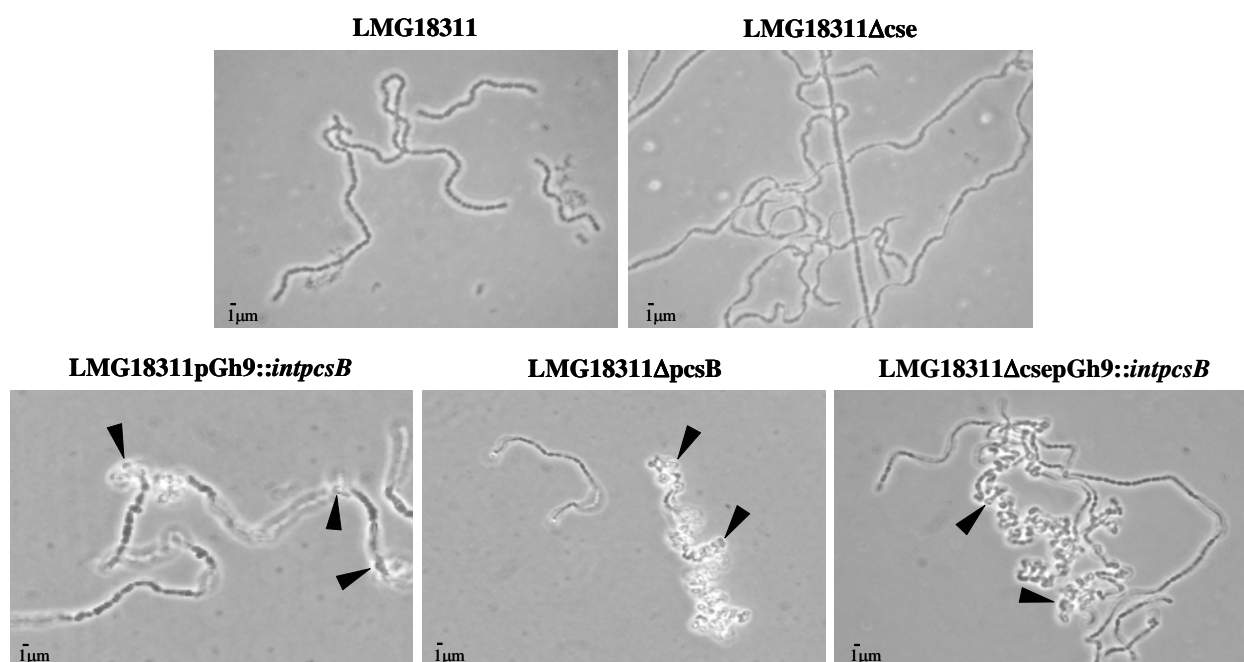
Les tentatives de construction d'un mutant  $\Delta pcsB$  à partir de la souche *S. thermophilus* LMG18311 $\Delta cse$ , n'ont donné aucun résultat malgré de nombreux essais et ceci dans différentes conditions. Inversement, le plasmide pGh9 ::*Δcse* qui avait servi à la construction du mutant LMG18311 $\Delta cse$  (Borges et al., 2005) a été utilisé pour sélectionner la délétion du gène *cse* chez le mutant LMG18311 $\Delta pcsB$ . De même que précédemment, aucun clone n'a été obtenu. Par conséquent,

nous avons entrepris la construction du plasmide pGh9::*Acse-Cm<sup>R</sup>*. Ce plasmide porte le gène de résistance au chloramphénicol et ses signaux de transcription, et de part et d'autre de ce gène, des séquences homologues aux régions flanquantes du gène *cse* à déléter. Le plasmide a été introduit chez le mutant LMG18311 $\Delta$ *pcsB* et en parallèle au sein de la souche sauvage LMG18311 utilisée comme témoin. De nouveau, aucun double mutant n'a été obtenu, tandis que des mutants délétés du gène *cse* et ayant acquis une résistance au chloramphénicol ont été sélectionnés à partir de la souche LMG18311 (Résultats non montrés).

Par conséquent, ces résultats laissent à penser que la délétion concomitante des gènes *cse* et *pcsB* serait contre sélectionnée chez *S. thermophilus*.

**b. Implication du gène *pcsB* dans l'organisation des septa chez *S. thermophilus*.**

Les mutants LMG18311 et LMG18311 $\Delta$ *cse* interrompus dans le gène *pcsB* ont été nommés respectivement LMG18311pGh9::*intpcsB* et LMG18311 $\Delta$ *csepGh9*::*intpcsB*. Le phénotype de ces deux mutants et du mutant LMG18311 $\Delta$ *pcsB* a été observé par microscopie photonique et comparé avec ceux de *S. thermophilus* LMG18311 et LMG18311 $\Delta$ *cse* (Figure 26).



**Figure 26 : Phénotype des interruptants LMG18311pGh9::*intpcsB*, LMG18311 $\Delta$ *csepGh9*::*intpcsB* et du mutant LMG18311 $\Delta$ *pcsB*.**

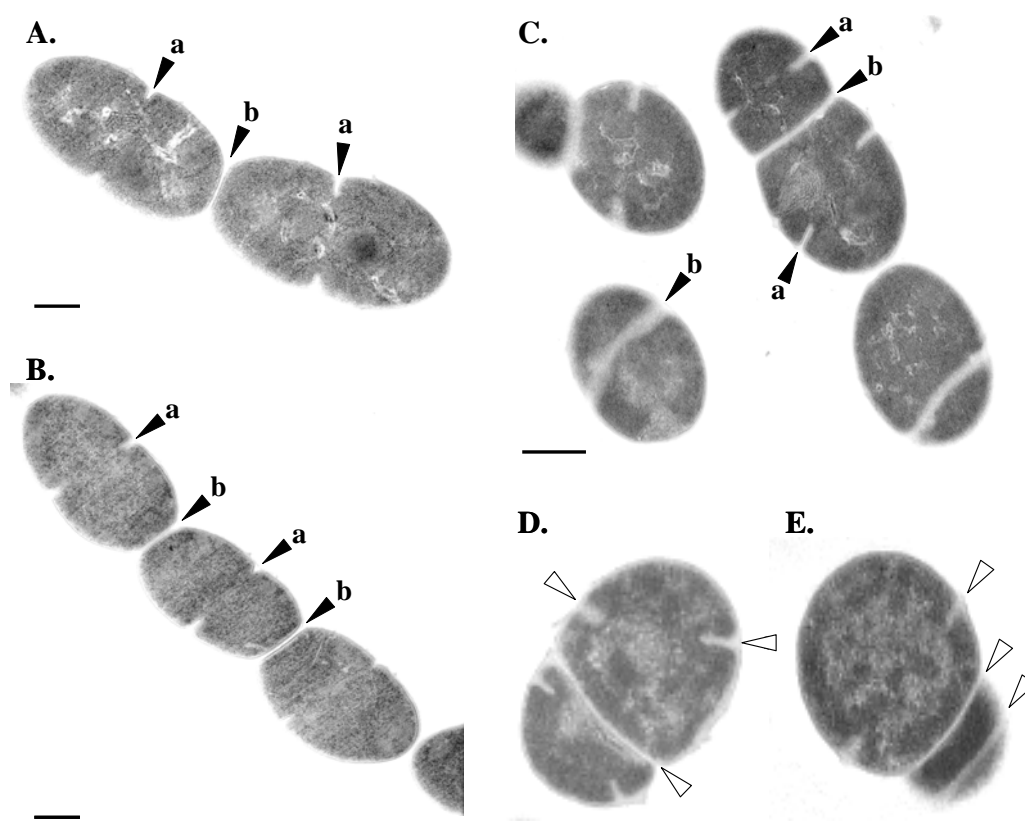
Les photographies ont été prises en phase stationnaire après 20h de cultures en HJL à 42°C. Les flèches pointent sur des cellules difformes et des entortillements.

Ces observations montrent que les chaînes de cellules des mutants LMG18311pGh9::*intpcsB* et LMG18311 $\Delta$ *pcsB* sont caractérisées par la présence de cellules anormales d'aspect globulaire et irrégulières (indiquées par les flèches sur la figure 26), alors que les chaînes de cellules de *S. thermophilus* LMG18311 sont le résultat d'enchaînements réguliers de diplocoques dont les cellules sont ovoïdes et uniformes. La longueur des chaînes de cellules des mutants LMG18311pGh9::*intpcsB* et LMG18311 $\Delta$ *pcsB* ne semble pas être significativement

différente de celle de la souche sauvage (Figure 26). Cependant, il est difficile d'estimer le nombre de cellules par chaînes de ces deux mutants dans la mesure où celles-ci sont entortillées sur elles-mêmes. Ainsi, l'inactivation du gène *pcsB* chez *S. thermophilus* LMG18311 n'induit pas un phénotype « longue chaîne de cellules » semblable au mutant  $\Delta$ *se* (Figure 26). Par conséquent, le gène *pcsB* n'est pas impliqué dans la séparation des cellules de *S. thermophilus*. Un argument en faveur de cette affirmation est que le mutant LMG18311 $\Delta$ *csepGh9* ::*intpcsB* présente à la fois des cellules difformes donnant un aspect « torturé » des chaînes de cellules et de longues chaînes de cellules (Figure 26). Par conséquent, le phénotype des trois mutants LMG18311pGh9 ::*intpcsB*, LMG18311 $\Delta$ *pcsB* et LMG18311 $\Delta$ *csepGh9* ::*intpcsB* est très différent de celui du mutant  $\Delta$ *se*.

Une complémentation avec une copie du gène *pcsB* porté par le plasmide pNST260<sup>+</sup> a été réalisée chez le mutant LMG18311 $\Delta$ *pcsB*. Elle conduit à une restauration du phénotype sauvage (Résultats non montrés).

Afin de préciser l'impact de la délétion du gène *pcsB* sur les chaînes de cellules de *S. thermophilus* LMG18311, une analyse sur des coupes des chaînes de cellules de la souche sauvage et des mutants (LMG18311 $\Delta$ *se* et LMG18311 $\Delta$ *pcsB*) par microscopie électronique à transmission (MET) a été réalisée (Figure 27).



**Figure 27 : Le gène *pcsB* est impliqué dans le positionnement et l'orientation des septa chez *S. thermophilus* LMG18311.**

Les photographies ont été prises sur des coupes de cellules au microscope électronique à transmission.

A= LMG18311 ; B= LMG18311 $\Delta$ *cse* ; C-E= LMG18311 $\Delta$ *pcsB*. Les lettres a et b correspondent respectivement au début de l'invagination de la paroi et au septum. Les septa anormaux, de même que les invaginations de la paroi positionnées au mauvais endroit sont indiqués par les flèches blanches.

La barre correspond à 0,5  $\mu$ m.

Ces observations montrent que les cellules du mutant LMG18311 $\Delta$ pcsB (Figure 27C-E) sont très différentes de celles du sauvage (Figure 27A) et du mutant LMG18311 $\Delta$ cse (Figure 27B). Les chaînes de cellules du sauvage ou du mutant  $\Delta$ cse résultent de l'enchaînement de deux cellules ovoïdes uniformes séparées à l'identique par un septum. Et dans une même cellule, la formation du futur site de septation marqué par le début de l'invagination de la paroi est régulière et partage la cellule en son centre en deux parties égales au cours de la division cellulaire (Figure 27A et B). Dans le cas du mutant LMG18311 $\Delta$ pcsB, les septa ne sont pas correctement positionnés conduisant à un aspect difforme des cellules et à des tailles de cellules variables. Par conséquent, l'orientation des plans de septation est altérée expliquant l'aspect « entortillé » des chaînes de cellules des mutants LMG18311pGh9::*intpcsB*, LMG18311 $\Delta$ pcsB et LMG18311 $\Delta$ csepGh9::*intpcsB* observé en microscopie photonique (Figure 26). De même que le mutant LMG18311 $\Delta$ pcsB, des observations similaires ont été obtenus pour le mutant LMG18311 $\Delta$ csepGh9::*intpcsB* en MET (Résultats non montrés). Par conséquent, ces résultats suggèrent que le gène *pcsB* soit impliqué principalement dans l'organisation des plans de septation lors de la division cellulaire de *S. thermophilus*.

Des résultats analogues chez trois streptocoques pathogènes : *S. agalactiae*, *S. mutans* et *S. pneumoniae* (Reinscheid et al., 2001 ; Chia et al., 2001 ; Ng et al., 2004) ont déjà été obtenus. Les données bibliographiques concernant les streptocoques pathogènes indiquent que l'inactivation du gène *pcsB* conduit également à la présence de cellules difformes et à une mauvaise organisation des plans de division. Par ailleurs, ce gène est essentiel à la croissance de ces streptocoques et plus particulièrement à *S. pneumoniae* dont l'inactivation est létale (Ng et al., 2004).

**Ainsi, l'ensemble de ces résultats montre que les mutants  $\Delta$ cse et  $\Delta$ pcsB ont des phénotypes différents. Et par conséquent, Cse et PcsB sont des protéines de fonctions cellulaires différentes.**

Ces résultats indiquent que *cse* est un gène de création récente dans le génome de *S. thermophilus*. Le fait que celui-ci soit absent des génomes de *S. salivarius* et *S. vestibularis*, espèces les plus proches phylogénétiquement de *S. thermophilus*, a permis de « dater » la création de ce gène chez *S. thermophilus*.

Par ailleurs, nous avons montré que les protéines Cse et PcsB ne sont pas des analogues fonctionnels. La protéine PcsB est une protéine de division cellulaire, nécessaire au positionnement et à l'orientation correcte des septa, tandis que Cse interviendrait dans le mécanisme de séparation des cellules de *S. thermophilus*. Néanmoins, cela n'exclut pas que ces deux protéines aient toutes deux une activité d'hydrolyse du peptidoglycane, conférée par leur domaine catalytique CHAP.

En conclusion, le fait que le gène *cse* ait une fonction différente du gène *pcsB* suggère que la création du gène *cse* par réassortiment de modules aurait conduit à l'émergence d'une protéine de fonction nouvelle chez *S. thermophilus*.

### III. Le gène *cse* code une enzyme de séparation cellulaire chez *Streptococcus thermophilus*.

Article en soumission

---

#### **The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote *S. thermophilus* cell separation.**

Séverine Layec, Joëlle Gérard, Valérie Legué, Marie-Pierre Chapot-Chartier, Pascal Courtin, Frédéric Borges, Bernard Decaris et Nathalie Leblond-Bourget.

**Molecular Microbiology**

---

La division cellulaire dont la plupart des différentes étapes est clairement établie, est un processus complexe dans lequel un septum de division est formé au futur site de séparation (pour revue Rothfield et *al.*, 1999 ; Errington et *al.*, 2003 ; Buddelmeijer et Beckwith, 2002). Après que la scission de la cellule mère en deux cellules filles ait abouti, la disjonction des cellules filles va nécessiter un clivage rigoureux et précis du peptidoglycane afin de préserver son intégrité et éviter la lyse cellulaire. Dans la littérature, ce processus implique l'action d'hydrolases du peptidoglycane. Toutefois, bien qu'un grand nombre d'enzymes capables de cliver les liaisons du peptidoglycane ait été identifié jusqu'à ce jour, peu d'entre elles semblent être impliquées dans la séparation des cellules. Les gènes codant ces enzymes de séparation cellulaire qui ont été identifiés, sont tous absents du génome de *S. thermophilus*.

Ainsi, le phénotype « longue chaîne de cellules » du mutant  $\Delta cse$  et le peu de connaissance concernant le mécanisme de séparation cellulaire chez les streptocoques, nous a conduit à étudier la fonctionnalité de la protéine Cse dans la séparation des cellules de *S. thermophilus*.

Une hypothèse simple concernant le rôle de la protéine Cse serait son implication dans le métabolisme du peptidoglycane. Les arguments en faveur de cette hypothèse proviennent de l'analyse de la séquence protéique de Cse. En effet, son extrémité C-terminale est caractérisée par la présence d'un domaine catalytique, le domaine CHAP (pour Cysteine Histidine dependent Aminohydrolase/Peptidase) (Figure 28). Ce domaine est exclusivement retrouvé au sein des protéines ayant une activité d'hydrolyse du peptidoglycane (Bateman et Rawlings, 2003). Par ailleurs, son extrémité N-terminale est caractérisée par un domaine de liaison à la paroi, le domaine LysM (pour Lysine motif) (Figure 28) connu comme étant une caractéristique partagée par de nombreuses hydrolases du peptidoglycane (Buist et *al.*, 2008).

Ainsi, l'ensemble de ces données permet de formuler l'hypothèse selon laquelle la protéine Cse serait une protéine de surface, probablement liée de façon non covalente à la paroi de *S. thermophilus* et qu'elle hydrolyserait une ou plusieurs liaisons du peptidoglycane qui restent à déterminer.

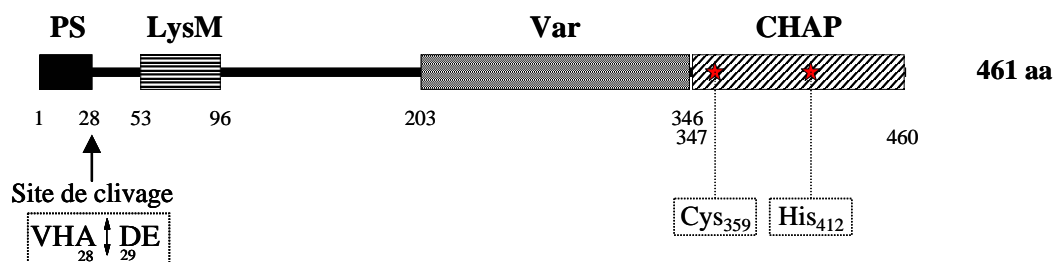


Figure 28 : Représentation schématique de l'architecture modulaire de la protéine Cse de *S. thermophilus* CNRZ368.

Le gène *cse* de *S. thermophilus* CNRZ368 (1383 pb) code une protéine de 461 acides aminés dont la masse moléculaire théorique est de 48,5 kDa et le pI (point isoélectrique) théorique est de 4,04. La région centrale riche en motifs répétés est symbolisée par un rectangle nommé « Var ». Les chiffres indiquent la position du peptide signal (PS), des domaines et de la région centrale. De même que le site de clivage (Val-His-Ala-Asp-Glu) reconnu par la peptidase signal qui hydrolyse la liaison peptidique entre les résidus 28 et 29. Les résidus du site actif du domaine CHAP sont représentés par une étoile en rouge.

### 1. Cse est une protéine extracellulaire.

Afin de tester l'hypothèse selon laquelle la protéine Cse serait extracellulaire, le gène *cse* a été fusionné au gène rapporteur *nuc*, délété de son peptide signal ( $\Delta_{Spnuc}$ ). Ce gène code une nucléase de *Staphylococcus aureus* (Le Loir et al., 1994) dont l'activité enzymatique peut être détectée sur milieu TBD-agar contenant de l'ADN de sperme de saumon couplé à du bleu de toluidine. Un halo rose est visualisé si la protéine de fusion est excrétée et dégrade l'ADN contenu dans le milieu (Lachica et al., 1971).

Le gène  $\Delta_{Spnuc}$  qui n'est sous la dépendance d'aucun promoteur, est porté par le vecteur pFUN (Poquet et al., 1998). Ainsi, la région contenant le promoteur et le gène *cse* en entier à l'exception du codon stop a été clonée dans le vecteur pFUN pour générer la fusion traductionnelle *cse*- $\Delta_{Spnuc}$ . Le plasmide pFUN::*cse* qui en résulte a ensuite été introduit dans *S. thermophilus* CNRZ368.

Les cellules transformées sont alors étalées sur boîte afin de générer des colonies après incubation à 30°C qui sont alors recouvertes d'une couche de milieu TBD-agar pour tester leur activité nucléasique potentielle.

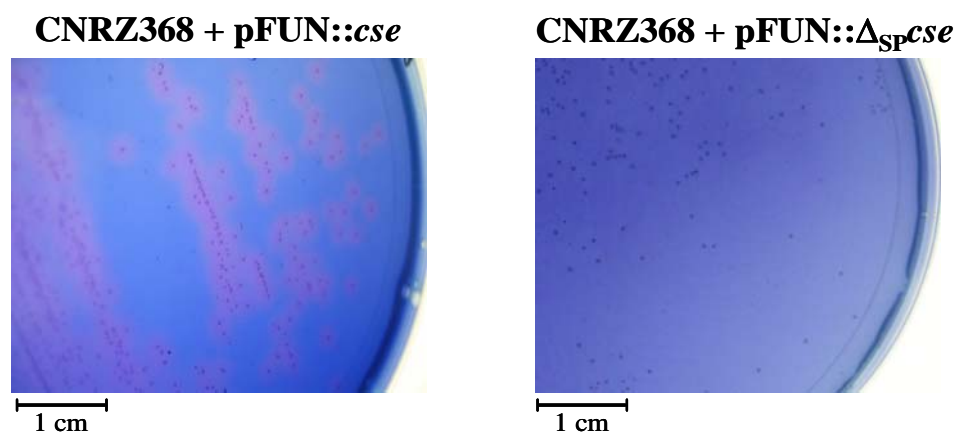


Figure 29 : Mise en évidence de l'exportation de la protéine de fusion Cse- $\Delta_{SpNuc}$  à l'extérieur des cellules de *S. thermophilus* CNRZ368.

L'activité nucléasique est visualisée après 2h d'incubation à 37°C grâce au milieu TBD-agar sur des colonies cultivées sur milieu TPPY solide supplémenté en érythromycine à 2µg/ml pendant 16h à 37°C.

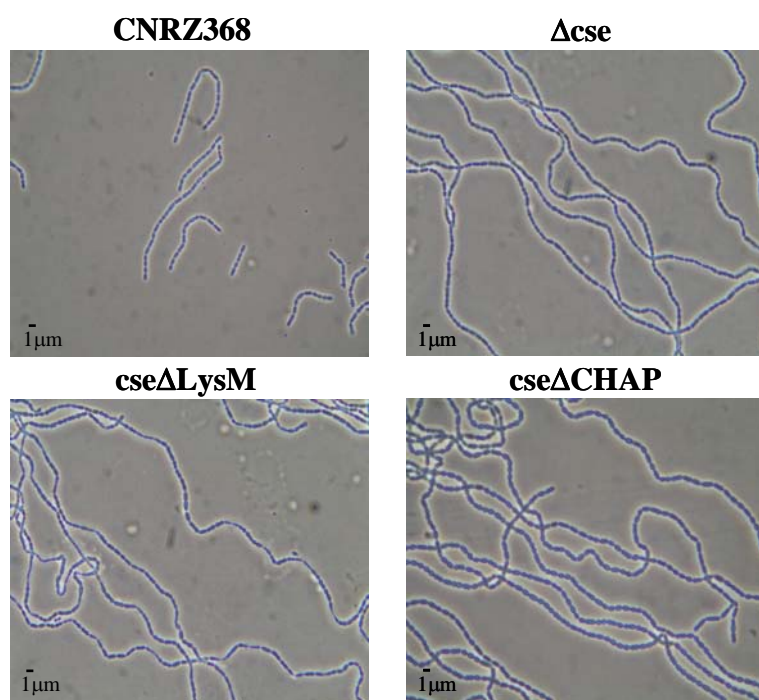


Un halo rose est visualisé autour des colonies du transformant CNRZ368+pFUN::cse (Figure 29), indiquant la présence de la protéine de fusion Cse- $\Delta_{SP}Nuc$  dans le milieu. Un témoin négatif a également été effectué par transformation de *S. thermophilus* CNRZ368 avec le plasmide pFUN:: $\Delta_{SP}cse$ , contenant la fusion traductionnelle  $\Delta_{SP}cse$ - $\Delta_{SP}nuc$  dans lequel le gène *cse* présente une délétion en phase de lecture de son peptide signal. Dans ce cas, aucun halo autour des colonies n'est observé (Figure 29), montrant que l'exportation de la protéine Cse est conférée par son peptide signal.

**Ainsi, ces résultats indiquent que le gène *cse* code une protéine extracellulaire : le peptide signal de la protéine Cse étant nécessaire à son exportation à l'extérieur de la cellule de *S. thermophilus*.**

## 2. Les domaines LysM et CHAP sont essentiels à l'activité de la protéine Cse.

Afin de connaître l'impact des domaines LysM et CHAP dans la fonction de séparation des cellules de la protéine Cse, des mutants délétés indépendamment de chacun des deux domaines et respectant la phase de lecture du gène *cse* de *S. thermophilus* CNRZ368 ont été construits. Les mutants *cse* $\Delta$ LysM et *cse* $\Delta$ CHAP présentent un phénotype « longue chaîne de cellules » semblable au mutant délété du gène *cse*,  $\Delta cse$  (Figure 30).



**Figure 30 : Phénotype des mutants *cse* $\Delta$ LysM et *cse* $\Delta$ CHAP.**

Les photographies ont été prises en phase stationnaire après 20h de cultures en HJL à 42°C

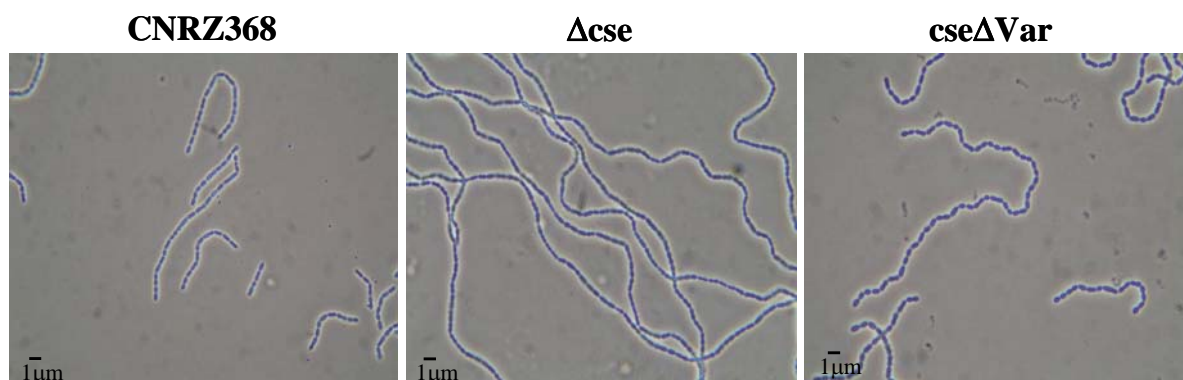
Par ailleurs, les chaînes de cellules des mutants *cse* $\Delta$ LysM et *cse* $\Delta$ CHAP sont constituées en moyenne de 400 cellules, contre en moyenne 40 cellules pour les chaînes de cellules de la souche sauvage.

La complémentation de ces mutants à l'aide d'un plasmide porteur du gène *cse* permet de restaurer le phénotype « courte chaîne de cellules » semblable au sauvage (Résultats non montrés).

Ainsi, la délétion des domaines LysM et CHAP a contribué à une perte de l'activité de séparation des cellules chez *S. thermophilus* suggérant que ces deux domaines sont essentiels à l'activité de la protéine Cse.

### 3. La région centrale n'est pas essentielle à l'activité de Cse.

Une autre caractéristique intéressante de la protéine Cse, concerne sa région centrale nommée « Var » riche en motifs répétés. L'interrogation des banques de données, en prenant pour cible cette région centrale n'a pas permis de révéler de domaine connu ni d'homologie de séquence. Afin de savoir si la région centrale est importante pour la fonction de Cse, la construction d'un mutant délété à 80% de cette région, en respectant la phase de lecture du gène *cse* de *S. thermophilus* CNRZ368 a été réalisée. Le mutant *cse* $\Delta$ Var qui en résulte, présente un phénotype dont les chaînes de cellules sont plus courtes que celles du mutant  $\Delta$ *cse* (Figure 31).



**Figure 31 : Influence de la délétion de la région centrale sur l'activité de séparation des cellules.**  
Les photographies ont été prises en phase stationnaire après 20h de cultures en HJL à 42°C.

La distribution du nombre de cellules par chaînes indique que 20% des chaînes de cellules du mutant *cse* $\Delta$ Var sont constituées de plus de 200 cellules, contre 78% des chaînes de cellules du mutant  $\Delta$ *cse* (Figure 32). Par conséquent, contrairement aux mutants *cse* $\Delta$ LysM et *cse* $\Delta$ CHAP, la délétion partielle de la région centrale n'a pas conduit à la perte de l'activité de Cse dans la séparation des cellules de *S. thermophilus*.

Par ailleurs, seulement 24% des chaînes de cellules du mutant *cse* $\Delta$ Var sont constituées de moins de 40 cellules contre 70% pour les chaînes de cellules de la souche sauvage (Figure 32). Ainsi, cette délétion ne s'est pas non plus traduite par un phénotype « courte chaîne de cellules » semblable au phénotype sauvage.

En revanche, 56% des chaînes de cellules du mutant *cse* $\Delta$ Var comptent 41 à 200 cellules, se traduisant par un phénotype dit « intermédiaire » entre le phénotype « courtes chaînes de cellules » de la souche sauvage et le phénotype « longues chaînes de cellules » que l'on observe pour les mutants  $\Delta$ *cse*, *cse* $\Delta$ LysM et *cse* $\Delta$ CHAP.

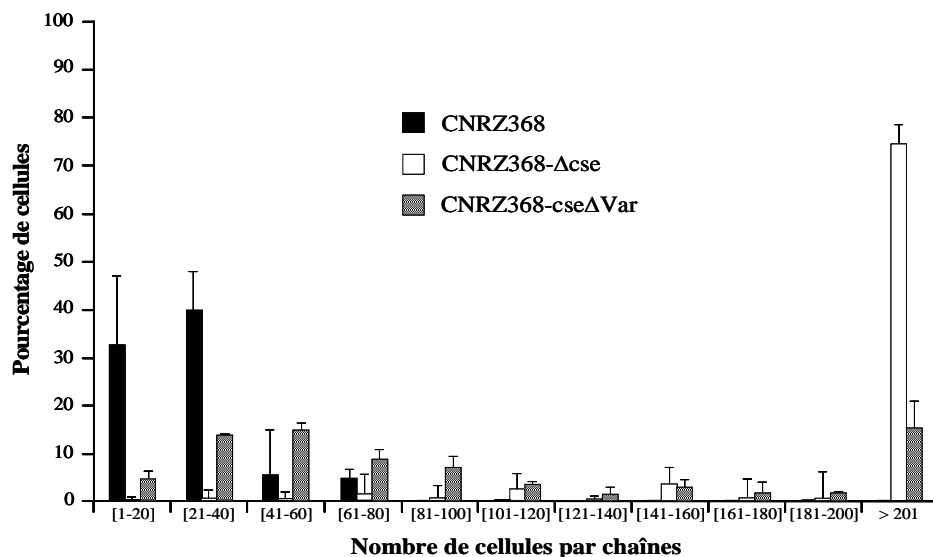


Figure 32 : Distribution du nombre de cellules par chaînes du mutant *cseΔvar*. Le comptage a été réalisé sur trois expériences indépendantes d'une culture en phase stationnaire après 20h d'incubation en HJL à 42°C.

Ainsi, le fait que la délétion de la région centrale n'ait pas conduit à perte de la fonction de séparation des cellules suggère que cette région ne soit pas indispensable à l'activité de la protéine Cse.

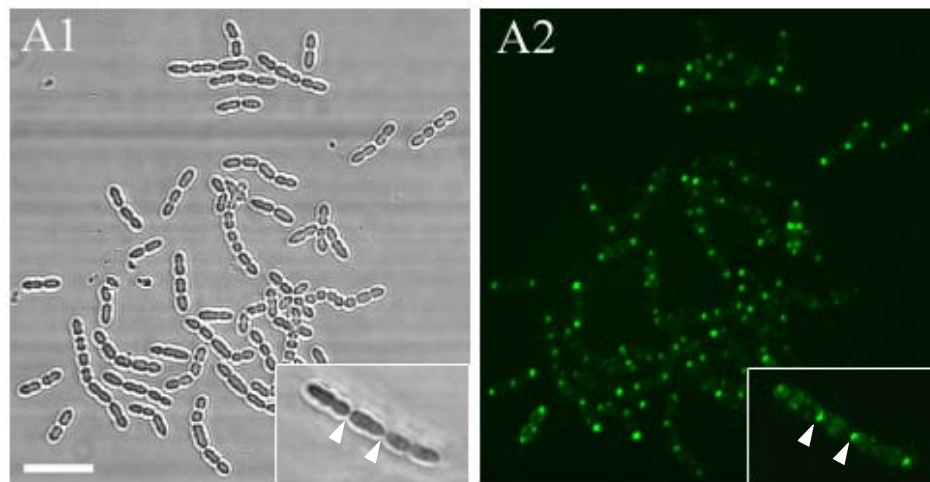
#### 4. Cse est une protéine de surface.

Les données bibliographiques concernant le domaine LysM indiquent qu'il présente une structure secondaire de type  $\beta\alpha\beta$  d'environ 40 acides aminés (Bateman et Bycroft, 2000). Ce domaine constitue un domaine de liaison non covalent, des protéines à la surface des cellules bactériennes (Joris et al. 1992 ; Steen et al., 2003).

Le fait que la protéine Cse soit extracellulaire et qu'elle possède le domaine LysM à son extrémité N-terminale, suggère que cette dernière serait liée à la surface des cellules de *S. thermophilus*. Afin de tester cette hypothèse, la localisation de la protéine Cse a été étudiée à la surface des cellules de *S. thermophilus* CNRZ368 par immunofluorescence et immunogold.

Pour cela, dans un premier temps, des anticorps dirigés contre Cse, et plus précisément contre l'extrémité N-terminale de la protéine Cse (du résidu aminé 1 à 183) ont été produits. C'est ainsi que le plasmide pET15b ::lysM<sub>Cse</sub> a été construit pour exprimer la protéine recombinante LysM<sub>Cse</sub>-His<sub>6</sub>. Cette protéine couplée à une queue de 6 résidus histidines à l'extrémité N-terminale a été purifiée par chromatographie d'affinité sur résine Ni<sup>2+</sup> en conditions natives. Puis celle-ci est injectée à un lapin afin d'obtenir des anticorps anti-LysM<sub>Cse</sub> dirigés contre la protéine Cse.

Les signaux de fluorescence (Figure 33) montrent que la protéine Cse est une protéine de surface liée à la paroi de *S. thermophilus* uniquement au niveau des septa matures (indiqués par les flèches sur la chaîne de cellules de *S. thermophilus* CNRZ368).



**Figure 33 : Immunolocalisation de la protéine Cse à la surface des cellules de *S. thermophilus* CNRZ368.**

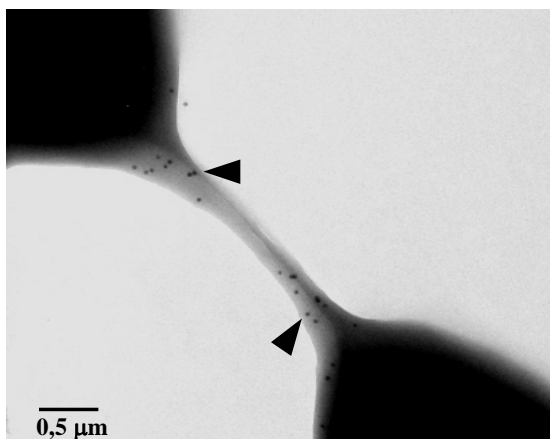
A1. La photographie des chaînes de cellules de *S. thermophilus* a été prise en phase de contraste au microscope confocal à balayage laser. A2. La protéine Cse est révélée par marquage en immunofluorescence avec les anticorps Alexa-fluor488 qui s'hybrident sur les anticorps primaire, anti-LysM<sub>Cse</sub>.

La barre correspond à 25 µm.

Ainsi, Cse présente une localisation spécifique. Contrairement aux protéines de surface à domaine LysM telles que AcmA de *L. lactis* et Sip de *S. agalatae* qui ont une localisation aux pôles cellulaires (Steen et al., 2005 ; Rioux et al., 2001), Cse est exclusivement localisée au niveau des septa matures des chaînes de cellules de *S. thermophilus* (Figure 33).

Les observations réalisées par microscopie électronique révèlent que les cellules filles sont connectées entre-elles par un pont constitué de peptidoglycane (Tomasz et Fischer, 2000) : leur séparation nécessitant vraisemblablement la rupture du lien reliant les deux cellules.

Le fait que l'on retrouve la protéine Cse au niveau des septa matures, c'est-à-dire au niveau du lien de peptidoglycane extrêmement fin entre deux cellules (Figure 34), suggère que Cse agisse probablement par clivage du peptidoglycane de ce lien fin.



**Figure 34 : Localisation de Cse sur le lien fin de peptidoglycane reliant deux cellules.**

La localisation de la protéine Cse est indiquée par des flèches pointées sur les grains d'or couplés aux anticorps secondaires qui se sont hybridés aux anticorps anti-LysM<sub>Cse</sub>. Cette photographie a été prise en microscopie électronique à transmission.

**Ainsi, ces résultats montrent que Cse est une protéine liée à la surface des cellules de *S. thermophilus* via son domaine LysM. Par ailleurs, l'étude de la localisation de la protéine Cse a révélé que celle-ci est localisée spécifiquement au niveau des septa matures de *S. thermophilus*,**

**suggérant que Cse serait vraisemblablement nécessaire aux étapes tardives de la séparation des cellules.**

## **5. Cse a une activité d'hydrolyse du peptidoglycane.**

La présence du domaine CHAP est une caractéristique partagée par des protéines ayant une activité d'hydrolyse du peptidoglycane (Bateman et Rawlings, 2003). Les données bibliographiques concernant le domaine CHAP indiquent que celui-ci présente deux résidus conservés du site actif : une cystéine et une histidine qui sont également retrouvés au sein de la structure primaire du domaine CHAP de la protéine Cse (Cys359 et His412).

Par ailleurs, la délétion du domaine CHAP ou la mutagenèse ciblée des résidus cystéine et histidine conservés a révélé une perte d'activité des hydrolases du peptidoglycane des phages  $\phi 11$  des staphylocoques (Navarre et al., 1999), B30 de *S. agalactiae* (Pritchard et al., 2004 ; Donovan et al., 2006) et PlyC des streptocoques (Nelson et al., 2006). De plus, il a été démontré que le domaine CHAP est impliqué dans le clivage de liens du peptidoglycane. Cependant, les liaisons peptidiques clivées par le domaine CHAP des protéines  $\phi 11$ , Sle1 de *Staphylococcus aureus* et B30 de *S. agalactiae* ne sont pas les mêmes (Navarre et al., 1999 ; Kajimura et al., 2005 ; Baker et al., 2006).

Ainsi, le fait que ce domaine soit retrouvé dans la protéine Cse, suggère que Cse serait probablement une hydrolase du peptidoglycane. Afin de confirmer cette hypothèse, l'étude de l'activité catalytique de la protéine Cse a été entreprise et la spécificité de clivage de Cse au sein de la paroi de *S. thermophilus* a été recherchée.

### **5.1 Mise en évidence de l'activité d'hydrolyse du peptidoglycane de Cse**

La technique la plus couramment employée pour démontrer l'activité d'hydrolyse du peptidoglycane est la technique SDS-PAGE (pour Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis) renaturante. Il s'agit d'un gel de polyacrylamide-SDS incluant un substrat rendant le gel opaque. Ce substrat peut être soit des cellules bactériennes lyophilisées et autoclavées ou une préparation de paroi purifiée. Après migration des extraits protéiques, les gels sont traités par un tampon de renaturation. Si une hydrolase du peptidoglycane se renature de façon native, elle pourra potentiellement entraîner la lyse des cellules ou paroi incluses dans le gel et un éclaircissement localisé à l'endroit de migration de la protéine hydrolytique sera observé.

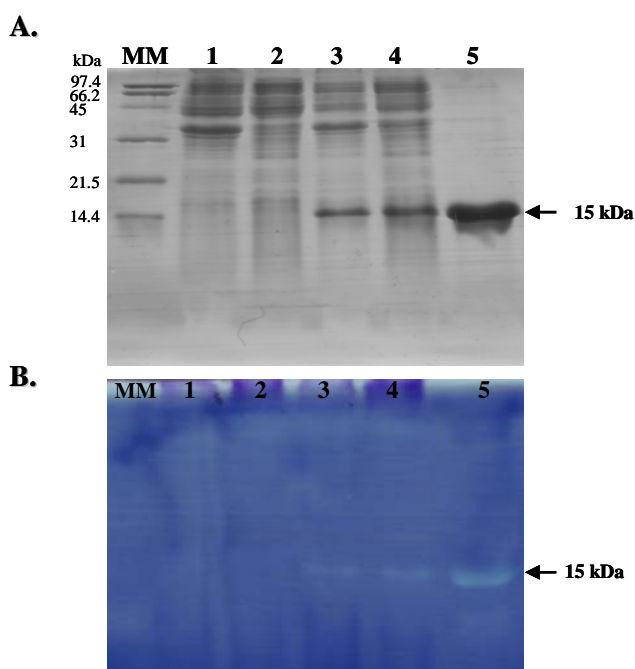
De nombreux essais de zymogramme ont été réalisés avec la protéine Cse en entier dans différentes conditions expérimentales. Cependant, toutes les tentatives ont échouées. N'arrivant pas à exprimer la protéine Cse entière à un niveau suffisant pour détecter son activité lytique, nous avons alors cherché à exprimer l'extrémité C-terminale (du résidu 341 à 361) contenant uniquement le domaine catalytique, CHAP de la protéine Cse de *S. thermophilus* CNRZ368.

Afin d'étudier l'activité d'hydrolyse du peptidoglycane de Cse par la technique SDS-PAGE renaturante, la construction du plasmide pET15b ::*chap* a été réalisée pour exprimer la protéine recombinante CHAP-His<sub>6</sub> possédant une queue histidine en N-terminale facilitant sa purification. Le taux de pureté de la protéine recombinante a été estimé à plus de 96% sur gel SDS-PAGE après coloration au bleu de Coomassie (Figure 35A, piste 5). Par ailleurs, la séquence protéique de la protéine recombinante de 15,1 kDa a été confirmée par analyse des spectres obtenus en HPLC-UV

après digestion de la protéine CHAP-His<sub>6</sub> avec la trypsine selon le protocole décrit dans la littérature (Miclo *et al.*, 1995 ; Perrin *et al.*, 1995) et par spectrométrie de masse MALDI-TOF (Ultraflex de Bruker Daltonics) (résultats non montrés).

Ainsi, des tests d'activité d'hydrolyse de la protéine recombinante CHAP-His<sub>6</sub> sur des gels SDS-15% PAGE contenant comme substrat des cellules lyophilisées et autoclavées de *Micrococcus lysodeikticus* ATCC 4698 (Sigma) et du mutant  $\Delta$ cse de *S. thermophilus* CNRZ368 ont été réalisés. Cette activité s'est traduite par l'apparition d'une zone d'éclaircissement correspondant à la lyse des cellules de *M. lysodeikticus* (Figure 35B, Pistes 3 à 5) ou de *S. thermophilus* au niveau de la co-migration de la protéine recombinante (~ 15 kDa) sur le gel SDS-15% PAGE.

L'activité de la protéine recombinante CHAP-His<sub>6</sub> n'a été détectée qu'à pH 4 et pH 4,5 avec 25mM de sodium citrate/50mM de sodium phosphate ou 50mM de KH<sub>2</sub>PO<sub>4</sub> de tampon additionné de 0,2% Triton-X100 après 16h de renaturation à 30°C.



**Figure 35 : Mise en évidence de l'activité lytique de la protéine recombinante CHAP-His<sub>6</sub>.**

A. Electrophorèse SDS-15% PAGE et coloration de bleu de Coomassie. B. Electrophorèse SDS-15% PAGE renaturante. Le gel contient 0,2% [wt/vol] des cellules autoclavées de *M. lysodeikticus* ATCC 4698 comme substrat. La renaturation a été réalisée en tampon 25mM de sodium citrate/50 mM de sodium phosphate pH 4 additionné de 0,2% de triton\_X100. La zone d'éclaircissement a été visualisée après coloration avec une solution de bleu de méthylène+ KOH.

Les pistes 1 et 2 correspondent aux extraits protéiques de la fraction insoluble et soluble respectivement du témoin négatif (*E. coli* BL21(DE3) + pET15b) ; les pistes 3 et 4 à des extraits protéiques de la fraction insoluble et soluble de *E. coli* exprimant la protéine recombinante et la piste 5 à la protéine recombinante purifiée.

Cependant, le fait que la protéine recombinante CHAP-His<sub>6</sub> présente une activité lytique avec des cellules lyophilisées et autoclavées comme substrat, ne prouve pas qu'elle agisse spécifiquement sur la paroi bactérienne. Par conséquent, une expérience de SDS-PAGE renaturante avec cette fois-ci 0,8 mg de paroi purifiée du mutant *S. thermophilus*  $\Delta$ cse comme substrat, a été réalisée dans les mêmes conditions que précédemment.

L'activité d'hydrolyse de la protéine recombinante CHAP-His<sub>6</sub> sur la paroi du mutant  $\Delta$ cse s'est traduite par la détection d'une zone d'éclaircissement (Figure 36) confirmant ainsi que cette protéine recombinante présente bien une activité d'hydrolase du peptidoglycane.

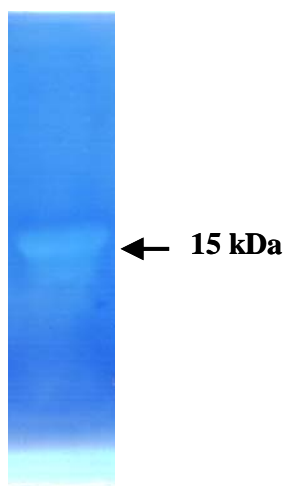


Figure 36 : Détection de l'activité d'hydrolyse du peptidoglycane de la protéine recombinante CHAP-His<sub>6</sub> sur la paroi de *S. thermophilus* Δcse. L'expérience a été réalisée par électrophorèse SDS-15% PAGE renaturante contenant de la paroi purifiée. La renaturation a été effectuée en tampon 25 mM de sodium citrate/50 mM sodium phosphate pH 4,0 additionné de 0,2% de Triton\_X100 pendant 16h à 30°C.

Ainsi, ces résultats confirment que la protéine Cse est bien une hydrolase du peptidoglycane dont l'activité est conférée par son domaine catalytique, CHAP.

## 5.2 Spécificité de clivage du peptidoglycane du domaine CHAP

Les hydrolases du peptidoglycane peuvent être classées en cinq classes selon la spécificité de la liaison qu'elles hydrolysent au sein du peptidoglycane (Figure 37) (Ghuysen 1968 ; Shockman et Barrett, 1983 ; Shockman et Høltje 1994).

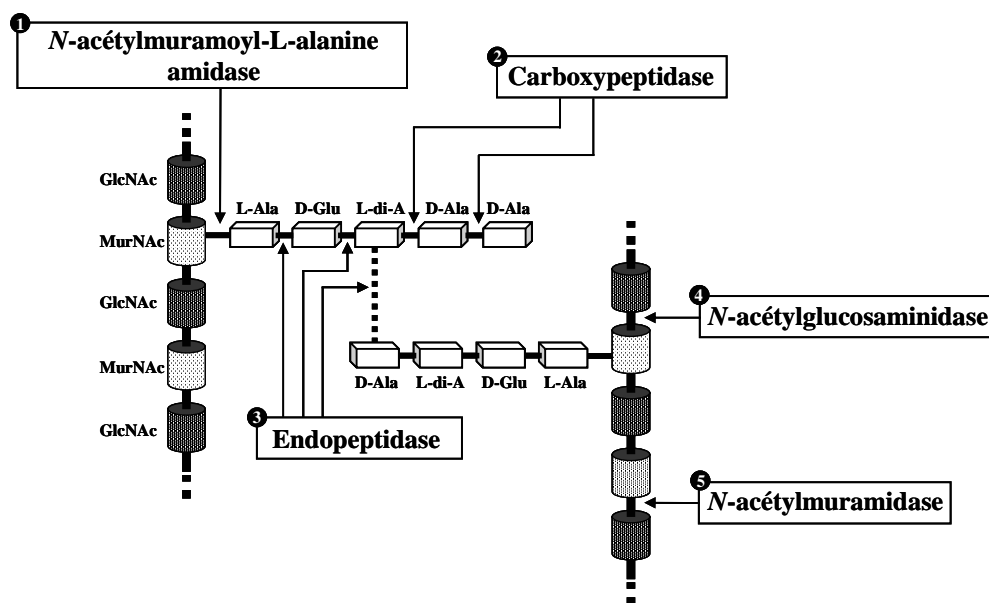


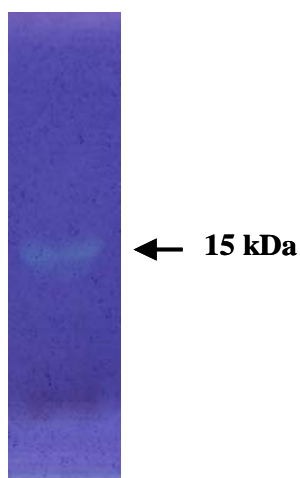
Figure 37 : Spécificité de clivage des hydrolases du peptidoglycane au sein de la paroi bactérienne.

Le peptidoglycane est un polymère constitué d'un motif de base composé de *N*-acetylglucosamine (GlcNAc) et de l'acide *N*-acétylmuramique (MurNAc) reliés ensemble par une liaison glycosidique de type  $\beta$  (1→4). A ces unités sont attachées des chaînes peptidiques latérales dont la séquence est L-ala-D-Glu-L-Lys-D-Ala-(D-Ala) pour la plupart des bactéries à coloration Gram positive, à l'exception de *Bacillus* et *Listeria* où le résidu Lys est remplacé par le résidu acide *méso*-diaminopimélique (mDAP). Ces chaînes latérales sont reliées par un pont interpeptidique qui est représenté par un trait en pointillé puisque sa structure est très variable (Scheilfer et Kandler, 1972).

Parmi ces différentes classes, on distingue les *N*-acétylmuramoyl-L-alanine amidases (ou amidases) qui clivent la liaison amide entre le groupe lactyl de l'acide *N*-acétylmuramique et le premier résidu de la chaîne peptidique latérale, le résidu L-alanine. Les *N*-acétylglucosaminidases et les *N*-acétylmuramidases hydrolysent quant à elles, la liaison glycosidique  $\beta$  (1 $\rightarrow$ 4) entre les unités disaccharidiques du peptidoglycane (Figure 37). On trouve également les carboxypeptidases qui clivent les liaisons peptidiques entre les résidus L-diA (L-Lys ou mDAP) et D-Ala qui le suit ou entre les deux résidus D-Ala des chaînes latérales. Et enfin, les endopeptidases qui sont capables d'hydrolyser les liaisons peptidiques de la chaîne latérale et du pont interpeptidique (Figure 37).

Afin de connaître à quelle classe d'hydrolases du peptidoglycane, la protéine Cse appartient, nous avons cherché à déterminer la spécificité de clivage de la protéine recombinante CHAP-His<sub>6</sub>, en collaboration avec Marie-Pierre Chapot-Chartier et Pascal Courtin du laboratoire de Biochimie bactérienne de l'INRA à Jouy-en-Josas.

En absence de données de cartographie de la paroi de *S. thermophilus*, une alternative était de tester l'activité de Cse sur le peptidoglycane de *Bacillus subtilis* dont la structure est connue (Atrih *et al.*, 1999). Cela nécessitait donc dans un premier temps de déterminer si le domaine CHAP de la protéine Cse était capable d'hydrolyser la paroi de *B. subtilis*. Pour ce faire, l'activité d'hydrolyse du peptidoglycane de la protéine recombinante a été testée sur la paroi purifiée de *B. subtilis* en SDS-15% PAGE renaturante dans les mêmes conditions que celles décrites dans le paragraphe 5.1. Ceci a permis de détecter une bande d'activité sur gel (Figure 38) indiquant que le domaine CHAP de Cse est capable d'hydrolyser le peptidoglycane de *B. subtilis*.



**Figure 38 : Détection de l'activité d'hydrolyse du peptidoglycane de la protéine recombinante CHAP-His<sub>6</sub> sur la paroi de *B. subtilis*.** L'expérience a été réalisée par électrophorèse SDS-15% PAGE renaturante contenant de la paroi purifiée. La renaturation a été effectuée en tampon 25 mM de sodium citrate/50 mM sodium phosphate pH 4,0 additionné de 0,2% de Triton\_X100 pendant 16h à 30°C.

Afin de déterminer le site de clivage de la protéine Cse, les muropeptides obtenus par digestion de la paroi de *B. subtilis* avec la protéine CHAP-His<sub>6</sub> ont été séparés par RP-HPLC et analysés par spectrométrie de masse MALDI-TOF. Les résultats obtenus ont révélé que la protéine recombinante coupe la liaison peptidique entre le résidu acide *méso*-diaminopimelique et le résidu alanine (Figure 39). Ainsi, ces résultats indiquent que la protéine recombinante a une activité endopeptidase. Après analyse des muropeptides, la liaison peptidique clivée par le domaine CHAP de la protéine Cse se trouve soit au niveau du pont interpeptidique ou au niveau de la chaîne peptidique latérale (Figure 39).



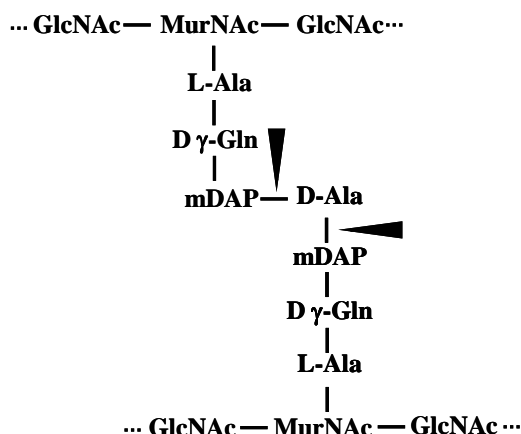


Figure 39 : Site de clivage de la protéine recombinante CHAP-His6 sur la paroi de *B. subtilis*. Les deux possibilités de liaison clivée par la protéine recombinante sont indiquées par les flèches.

Cependant, d'après Scheilfer et Kandler (1972), le peptidoglycane des streptocoques est dépourvu du résidu acide *méso*-diaminopimélique. Les chaînes peptidiques latérales sont constituées par l'enchaînement L-Ala-D-Gln-L-Lys-D-Ala-(D-Ala). Par analogie avec le peptidoglycane de *S. agalactiae* (Scheilfer et Kandler, 1972), il est probable que les chaînes peptidiques latérales de *S. thermophilus* soient reliées par un pont interpeptidique constitué de 2 ou 3 résidus L-Ala reliant le résidu D-Ala en position 4 d'une chaîne au résidu L-Lys en position 3 d'une autre chaîne.

Afin de déterminer le site de clivage du peptidoglycane de *S. thermophilus*, la paroi purifiée du mutant *S. thermophilus* Δcse a été digérée par la mutanolysine de *Streptomyces globispirus*. Les produits de digestion qui en résultent ont été analysés par RP-HPLC, et trois muuropeptides identifiés par spectrométrie de masse MALDI-TOF ont été retenus (Figure 40) et digérés par la protéine recombinante CHAP-His6.

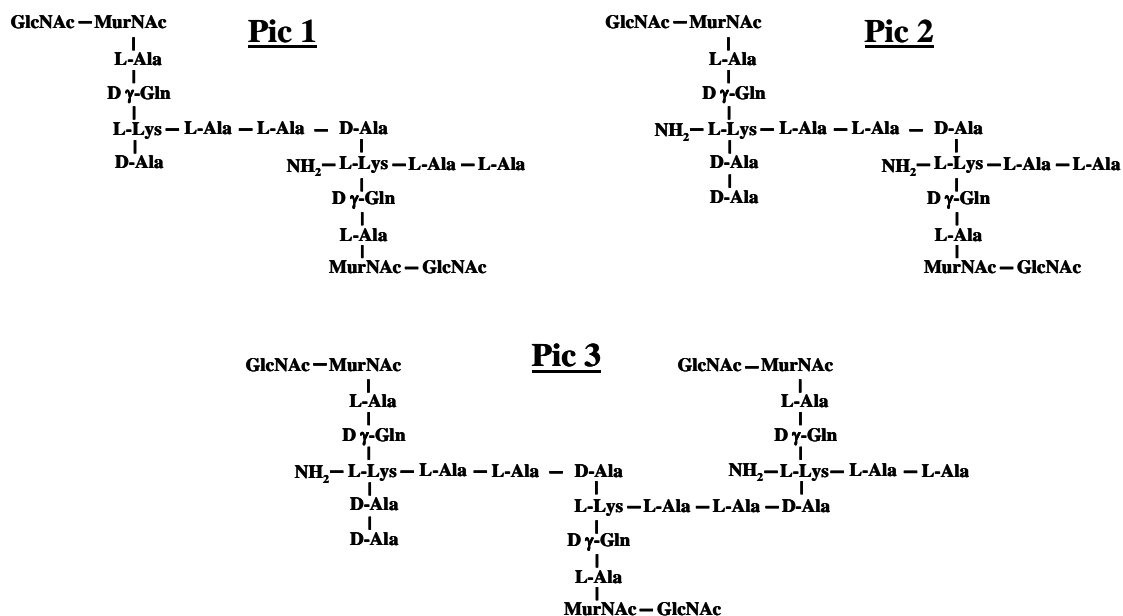
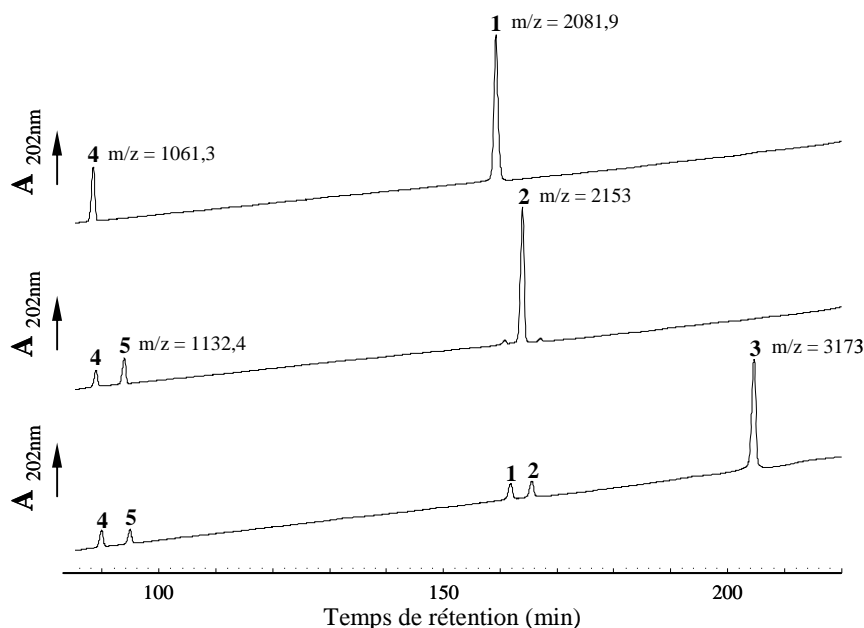


Figure 40 : Structure des trois muuropeptides de *S. thermophilus*.

Le muuropeptide 1 correspond au dimère DS-Tetra AA-DS-Tetra-AAa ; le muuropeptide 2 correspond au dimère DS-Penta-AAa-DS-TetraAAa et le muuropeptide 3 correspond au trimère DS-Penta-AAa-DS-Tetra-AAa-DS-Tetra-AAa. DS= disaccharides GlcNAc-MurNAc ; Tetra = L-Ala-Dγ-Gln-L-Lys-D-Ala ; Penta = L-Ala-Dγ-Gln-L-Lys-D-Ala-D-Ala ; AA = L-Ala-L-Ala et a = amidation.

Les muuropeptides correspondent aux pics 1, 2 et 3 de la figure 41.

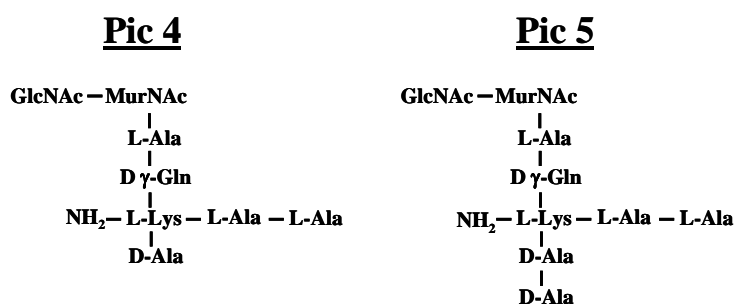
Les produits de digestion de ces trois muuropeptides par la protéine recombinante CHAP-His<sub>6</sub> ont été analysés par RP-HPLC et identifiés par spectrométrie de masse MALDI-TOF (Figure 41).



**Figure 41 : Profil d'élu­tion chromatographique des trois muuropeptides issus de la digestion par la protéine recombinante CHAP-His<sub>6</sub>.**

Les trois muuropeptides testés sont le dimère DS-Tetra-AA-DS-Tetra-AAa (pic 1), le dimère DS-Penta-AAa-DS-Tetra-AAa (pic 2) et le trimère DS-Penta-AAa-DS-Tetra-AAa-DS-Tetra-AA (pic 3). Les produits de digestion sont les monomères DS-Tetra-AAa (pic 4) et DS-Penta-AAa (Pic 5) et des dimères lors de digestions partielles. DS = disaccharide GlcNAc-MurNAc ; Tetra = L-Ala-Dγ-Gln-L-Lys-D-Ala ; Penta = L-Ala-Dγ-Gln-L-Lys-D-Ala-D-Ala ; AA = L-Ala-L-Ala et a signifie amidation.

Ces résultats montrent que la protéine recombinante CHAP-His<sub>6</sub> est capable d'hydrolyser ces trois muuropeptides issus de la paroi de *S. thermophilus*. La structure des produits de digestion qui en résulte est schématisée dans la figure 42.



**Figure 42 : Structure des produits de digestion correspondant aux pics 4 et 5 après hydrolyse des trois muuropeptides par le domaine CHAP de la protéine Cse.**

Le muuropeptide correspondant au pic 4 est un monomère DS-Tetra-AAa, tandis que le muuropeptide du pic 5 correspond au monomère DS-Penta-AAa.

L'analyse de ces résultats révèle que le domaine CHAP de Cse a une activité endopeptidase. Il clive la liaison peptidique entre le résidu D-alanine de la chaîne peptidique latérale et le résidu L-alanine du pont interpeptidique du peptidoglycane de *S. thermophilus* (Figure 43).

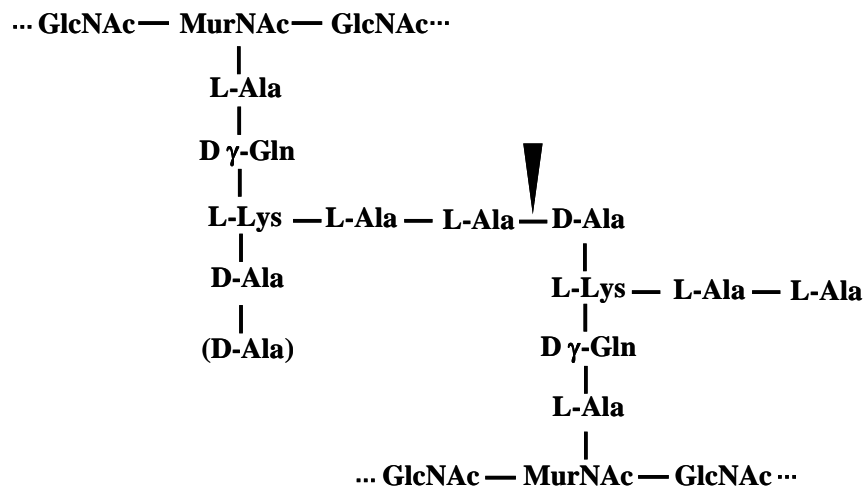


Figure 43 : Site de clivage de la protéine recombinante CHAP-His<sub>6</sub> sur la paroi de *S. thermophilus*. La liaison peptidique hydrolysée par la protéine recombinante est indiquée par la flèche.

Ainsi, ces résultats indiquent que la protéine recombinante CHAP-His<sub>6</sub> hydrolyse la liaison peptidique du pont interpeptidique. Par conséquent, la protéine Cse est une D-Alanyl-L-Alanine endopeptidase dont l'activité est conférée par le domaine CHAP.

Le gène *cse* code l'enzyme majeure qui assure la séparation des cellules chez *S. thermophilus*. Il s'agit d'une endopeptidase qui catalyserait vraisemblablement l'ultime étape de la séparation des cellules, au niveau des septa matures. Par conséquent, la séparation des cellules est un processus actif qui nécessite l'action d'hydrolases du peptidoglycane agissant non pas sur tout le pourtour des cellules mais à un endroit précis ciblant le site de séparation des cellules filles.



## **The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote *Streptococcus thermophilus* cell separation.**

Séverine Layec <sup>1</sup>, Joëlle Gérard <sup>2</sup>, Valérie Legué <sup>2</sup>, Marie-Pierre Chapot-Chartier<sup>3</sup>, Pascal Courtin<sup>3</sup>, Frédéric Borges <sup>1f</sup>, Bernard Decaris <sup>1</sup> and Nathalie Leblond-Bourget <sup>1\*</sup>

<sup>1</sup>Laboratoire de Génétique et Microbiologie, UMR INRA/UHP 1128, IFR 110, Nancy-Université, BP 239, 54506 Vandoeuvre-lès-Nancy, France.

<sup>f</sup> present address: LSGA- Laboratoire de Science et Génie Alimentaire, ENSAIA-INPL, BP172, 54506 Vandoeuvre-lès-Nancy, France.

<sup>2</sup>Génomique, Ecophysiologie et Ecologie Fonctionnelles, « Interaction Arbres/Micro-organismes », UMR INRA/UHP 1136, IFR 110, Nancy-Université, BP239, 54506 Vandoeuvre-lès-Nancy, France.

<sup>3</sup>INRA, UR477 Biochimie Bactérienne, F-78350 Jouy-en-Josas, France

\*Corresponding author : Phone: (33) 3 83 68 42 10  
Fax : (33) 3 83 68 44 99  
E-mail : [bourget@nancy.inra.fr](mailto:bourget@nancy.inra.fr)

Keywords: CHAP domain; cell wall hydrolase; mature septa; daughter cell separation; *S. thermophilus*

### **SUMMARY**

Cell separation is dependent on cell wall hydrolases that cleave the peptidoglycan shared between daughter cells. In *Streptococcus thermophilus*, this step is performed by the Cse protein whose depletion resulted in the formation of extremely long chains of cells. Cse, a natural chimeric enzyme created by domain shuffling, carries at least two important domains for its activity: the LysM expected to be responsible for the cell wall-binding and the CHAP domain predicted to contain the active center. Accordingly, the localization of Cse on *S. thermophilus* cell surface has been undertaken by immunogold electron and immunofluorescence microscopies using of antibodies raised against the N-terminal end of this protein. Immunolocalization shows the presence of the Cse protein at mature septa. Moreover, the CHAP domain of Cse overexpressed in *Escherichia coli* exhibits a lytic activity on the cell wall of several Gram-positive bacteria, including *S. thermophilus*. Additionally, RP-HPLC analysis of mucopeptides released from *Bacillus subtilis* and *S. thermophilus* cell wall after digestion with the CHAP domain shows that Cse is a peptidoglycan hydrolase that can function as an endopeptidase. Altogether, these results suggest that Cse is a major cell wall hydrolase involved in daughter cell separation of *S. thermophilus*.

## INTRODUCTION

Cell separation can be defined as the mechanism needed to disconnect the two new daughter sacculi after the cell division is completed (or at the very late step of cell division). This is still a poorly understood process that requires the involvement of peptidoglycan hydrolases (PGHs). Cell separating PGHs were mostly identified by phenotypic analysis, their absence resulting in the formation of chains of unseparated cells (Borges *et al.*, 2005; Buist *et al.*, 1995; Fukushima *et al.*, 2006; Garcia *et al.*, 1999; Heidrich *et al.*, 2001; Margot *et al.*, 1998; Margot *et al.*, 1999; Yoshimura *et al.*, 2006) or in a highly aggregated phenotype (Kajimura *et al.*, 2005). In *Streptococcus thermophilus*, the deletion of *cse* led to a dramatic increase of the chain length indicating that Cse is the key *S. thermophilus* cell separating enzyme (Borges *et al.*, 2005).

The Cse protein consists of four regions: a signal peptide and a LysM domain at its N-terminus, a catalytic domain at its C-terminus, both domains being separated by a central variable region (Borges *et al.*, 2005; Borges *et al.*, 2006). The LysM motif functions as a general peptidoglycan binding module (Eckert *et al.*, 2006; Steen *et al.*, 2005a) that may play important roles in appropriate localization and/or recognition of substrates (Fukushima *et al.*, 2006; Kajimura *et al.*, 2005; Steen *et al.*, 2005a; Yamamoto *et al.*, 2003).

The C-terminal catalytic domain of Cse carries a putative Cysteine, Histidine-dependent Amidohydrolases/Peptidases (CHAP) domain conferring to the protein a potential catalytic activity (Borges *et al.*, 2005; Layec *et al.*, 2008a). Point mutation of the conserved Cys and His residues or total deletion of this domain reveals that it retains peptidoglycan hydrolase activity (Donovan *et al.*, 2006; Navarre *et al.*, 1999; Nelson *et al.*, 2006; Pritchard *et al.*, 2004; Yokoi *et al.*, 2005).

Among the Firmicutes, few bacterial and phage-encoded CHAP proteins have been characterized and participate either in cell division (Chia *et al.*, 2001; Mattos-Graner *et al.*, 2006; Ng *et al.*, 2004; Reinscheid *et al.*, 2001) daughter cell separation (Borges *et al.*, 2005; Kajimura *et al.*, 2005) or in cell lysis (Donovan *et al.*, 2006; Huard *et al.*, 2003; Navarre *et al.*, 1999; Pritchard *et al.*, 2004; Yokoi *et al.*, 2005).

By analogy with other cell separating enzymes and taking into account the involvement of Cse in cell disconnection, we postulate that Cse is a peptidoglycan hydrolase probably acting at cell separation sites. However, there has been no direct evidence concerning either the Cse catalytic activity or its localization at the *Streptococcus thermophilus* cell surface. In the present report, we provide genetic and biological evidence that Cse separates daughter cells by cleaving the bond between D-Ala of the stem peptide and L-Ala of the cross-bridge in *S. thermophilus* peptidoglycane at mature septa.

## RESULTS

### The LysM and CHAP domains are essential for Cse cell separating activity.

To get more insight into the biological role of the LysM and CHAP domains of the Cse protein, in frame deletion mutants were constructed resulting in the *cse*ΔLysM and *cse*ΔCHAP mutant strains (Fig. 1A). Visual inspection of overnight standing HJL cultures revealed that turbidity in the cultured medium was only observed for the CNRZ368 control strain whereas the cells of the Δ*cse*, *cse*ΔLysM and *cse*ΔCHAP mutants all flocculate at the bottom of the tube (Fig. 1B). To determine if the sedimentation properties of these mutants correlate with an increase of their chain length, aliquots of each culture were observed under light microscopy followed by the counting of the number of cells per chain. As observed in figure 1C, chains of the wild-type strain were rather small since 70% of them were constituted by less than 40 cells. By contrast, chains of the 3 *cse* deletion mutants were mostly composed of more than 201 cells. Complementation experiments were performed by transformation of the *cse*ΔLysM and *cse*ΔCHAP mutants with pNST260<sup>+</sup>::*cse*, a vector containing an intact copy of *cse* that we developed previously (Borges *et al.*, 2005). Complemented strains exhibited a wild-type phenotype with respect to the number of cells per chain and to the sedimentation property while the negative control maintained a mutant phenotype (data not shown). Thus, deletion of the

LysM or the CHAP domain of the *cse* gene had a clear effect on the chain length indicating that both domains were required for Cse efficient cell separation activity.

### **Cse is not involved in cell wall synthesis at septa.**

In a previous work, we demonstrated that the catalytic CHAP module of *cse* is homologous to that of the *pcsB* gene of group B streptococci (GBS) (Borges *et al.*, 2005). Downexpression of *pcsB* from *Streptococcus pneumoniae* and *Streptococcus agalactiae* resulted in the formation of cell with abnormal morphology and led to uncontrolled cell wall synthesis at misplaced septa (Ng *et al.*, 2004; Reinscheid *et al.*, 2001). In *Streptococcus mutans*, the homolog of PcsB is GbpB an essential protein involved in peptidoglycan synthesis and cell division (Chia *et al.*, 2001; Mattos-Graner *et al.*, 2006).

Transmission electron microscopy of both the wild-type and the  $\Delta cse$  mutant revealed that they formed linear chains that appeared like linear necklaces of diplococcal cells (Fig. 2). As observed for *S. pneumoniae*, progression of the *S. thermophilus* septum synthesis is materialised by three kinds of septa (Fig. 2C and F): (i) nascent septum (a) located equidistant from the cell poles and initiated by the constriction ring, (ii) intermediate septum (b) generated by further constriction of the cell walls and (iii) mature septum (c) observed at the end of the division process and corresponding to an extension of the septum holding daughter cells together. No impairment of the septum formation was observed in the  $\Delta cse$  mutant suggesting that Cse had no role in cell wall synthesis at septa. The only difference between the  $\Delta cse$  mutant (Fig. 2A to C) and the wild type strains (Fig. 2D to F) is the regularity of the cell morphology. The mutant displayed a regular succession of diplococcal cells with mature septa (c) found every four cells. This suggests a better synchronization of the mutant cell division compared to wild-type.

Since deletion of *cse* led to a dramatic reduction of daughter cell separation and because the material linking daughter cells was expected to be peptidoglycan, we postulate that Cse acts as a late peptidoglycan hydrolase that is able to bind *S. thermophilus* cell surface and separate divided cells.

### **Localization of Cse and its derivatives on the cell surface.**

In a previous study, we demonstrate that Cse is an extracellular protein whose export is signal peptide dependent (Borges *et al.*, 2005). Moreover, a search of the Conserved Domain Database shows that the N-terminal parts of Cse contain a LysM domain potentially involved in Cse attachment to the cell wall. Thus, to visualize the specific localization of Cse on the *S. thermophilus* cell surface, we utilized immunogold electron microscopy. Antibodies were raised only against the N-terminal end of Cse (from 1 to 183 aa) to avoid the simultaneous detection of Cse and PcsB sharing the same C-terminal end (Borges *et al.*, 2005). The amino-acid sequence of Cse from position 1 to 183 was used in a BLASTP search against the genome of *S. thermophilus* LMG18311 and allowed the retrieval of Cse exclusively. As shown in figure 3, the labelled gold particles were localized on the fine link connecting daughter cells and within adjacent cell poles (Fig. 3A and B). However, because of the darkness of the cells, we could not determine if labelled particles were present elsewhere. We used next immunofluorescence microscopy on the wild-type strain, and confirmed that Cse binding was highly localized at mature septa with no or little Cse protein at nascent septa (Fig. 4A2 and E2). No fluorescence was detected on the *cse* $\Delta$ LysM mutant (Fig. 4C) or on the wild-type strain stained with preimmune serum (Fig. 4B) used as negative controls. Fluorescence analysis of the *cse* $\Delta$ CHAP mutant (Fig. 4D) showed that the Cse derivative was capable of recognizing the same targets but surprisingly the patterns of fluorescence was more diffuse than that of the wild type strain. Staining of the wild type *S. thermophilus* cells with Calcofluor fluorescent dye (Fig. 4, E4) which stains cell walls blue confirmed the specific binding of Cse to the peptidoglycan of the mature septa.

### **Cse is a D,L-endopeptidase.**

#### **(i) The CHAP domain of Cse confers to the protein its cell-wall peptidoglycan hydrolyzing activity.**

In a previous work, we tried to overexpress the entire Cse protein to characterize its enzymatic activity but our attempts failed. This time, only the CHAP domain of Cse (from position in aa 341 to 461) was expressed in *E. coli*. To facilitate the purification of the CHAP protein, a hexahistidine tag was introduced at its N-terminus. On SDS-PAGE, the purified CHAP-His<sub>6</sub>-tagged protein gave a band of about 15 kDa estimated to be more than 96% pure (Fig. 5, lane 5). This band

corresponded to the cell-wall hydrolyzing band observed on zymogram (Fig. 5, lane 6) and was also revealed by Western immunoblot analysis using an anti-His tag monoclonal antibody (supplementary data Fig. 8A). Zymograms were done using lyophilised and autoclaved whole cells to *Micrococcus lysodeikticus* ATCC 4698 (Supplementary data Fig. 8B) and of the *S. thermophilus*  $\Delta$ cse mutant (Fig. 5) as substrates and revealed that the CHAP domain of Cse was able to cleave cells of both species. That Cse acts as a peptidoglycan hydrolase was further confirmed using purified insoluble cell-wall peptidoglycan from *S. thermophilus*  $\Delta$ cse mutant (Fig. 5, lane 6) and from *B. subtilis* 168 HR (data not shown). The CHAP-His<sub>6</sub>-tagged protein activity was efficient with incubation buffer containing 0.2% Triton\_X100 at pH 4.0 or pH 4.5. No activity was detected at other pH (data not shown).

#### (ii) Determination of the peptidoglycan hydrolytic bond specificity of the CHAP domain.

The hydrolytic bond specificity of the CHAP-His<sub>6</sub>-tagged protein was investigated on *B. subtilis* cell-wall peptidoglycan whose structure was established previously (Atrih *et al.*, 1999). Therefore, purified CHAP-His<sub>6</sub>-tagged protein was used to digest purified cell-wall peptidoglycan of *B. subtilis*. After incubation, the digestion mixture was centrifuged and both supernatant and insoluble fractions were analyzed. No muropeptide or peptide could be detected in the supernatant when analyzed by RP-HPLC as described in Material and Methods (data not shown). The pellet obtained after CHAP-His<sub>6</sub>-tagged treatment of peptidoglycan was further digested with mutanolysin and the soluble muropeptides were analyzed by RP-HPLC. Comparison of the muropeptide profile of the peptidoglycan double digest with the muropeptide profile obtained by mutanolysin digestion of *B. subtilis* peptidoglycan (Fig. 6A) revealed a large increase of monomers in parallel with a clear decrease of dimers and trimers in the CHAP-His<sub>6</sub>-tagged treated sample (Table 3A). Especially, the amount of the major dimers, Peak 2 and 3 (Fig. 6A) corresponding to disaccharide-tripeptide-disaccharide-tetrapeptide with one or two amidations respectively, decreased strikingly in *B. subtilis* peptidoglycan treated with the CHAP-His<sub>6</sub>-tagged protein, whereas the amount of monomers in peaks 1, 4, 5 that are disaccharide-tripeptide, disaccharide-tetrapeptide without an amidation, and disaccharide-tetrapeptide with an amidation, respectively, increased clearly (Fig. 6A and Table 3B). From these results, we can conclude that the CHAP-His<sub>6</sub>-tagged protein cleaves the linkage between *meso*-diaminopimelic acid and D-alanyl residues either in the cross-bridge or in the stem peptide of *B. subtilis* cell-wall peptidoglycan (Fig. 6C1). Thus, the CHAP-His<sub>6</sub>-tagged protein has endopeptidase activity.

Whereas *B. subtilis* peptidoglycan contains *meso*-diaminopimelic acid at the third position of the stem peptide and direct cross-link of peptide chains, *S. thermophilus* peptidoglycan contains a Lys residue in position three and two (or three) L-Ala in the interpeptide bridge (Schleifer and Kandler, 1972). To identify the CHAP-His<sub>6</sub>-tagged protein cleavage site on *S. thermophilus* peptidoglycan, three muropeptides purified from a mutanolysin digest of *S. thermophilus* (Table 4) were used as substrates. After incubation with CHAP-His<sub>6</sub>-tagged, analysis of the reaction mixtures by RP-HPLC revealed several peaks. In Figure 6B, the peaks with the highest retention times (peak 8 in a) ; peak 9 in b) and peak 10 in c) correspond to the intact muropeptides whereas the others are consistent with being the CHAP-His<sub>6</sub>-tagged protein digestion products. Thus, although being partial, this digestion further confirmed that the CHAP-His<sub>6</sub>-tagged protein is able to cleave *S. thermophilus* peptidoglycan. According to the retention times of each peak and the masses determined by mass spectrometry, the structures in Table 4 can be proposed for the different muropeptides. From these results, it appears that the CHAP recombinant protein cleaves the bond between D-Ala of the stem peptide and L-Ala of the cross-bridge in *S. thermophilus* peptidoglycan (Fig. 6C2) and that Cse is a D,L-endopeptidase.



## DISCUSSION

Within the *Firmicutes*, only a few CHAP proteins have been characterized that function as peptidoglycan hydrolases. Among them are PcsB/GbpB from *Streptococcus* involved in cell division (Chia *et al.*, 2001; Mattos-Graner *et al.*, 2006; Ng *et al.*, 2004; Reinscheid *et al.*, 2001), Sle1 from *Staphylococcus aureus* (Kajimura *et al.*, 2005) and Cse from *S. thermophilus* that are necessary for daughter cell separation (Borges *et al.*, 2005). Acmb from *Lactococcus lactis* contributes to cellular autolysis (Huard *et al.*, 2003). Six other phage-encoded CHAP proteins are involved in bacterial cell lysis (Donovan *et al.*, 2006; Llull *et al.*, 2006; Navarre *et al.*, 1999; Nelson *et al.*, 2006; Pritchard *et al.*, 2004; Rashel *et al.*, 2008; Yokoi *et al.*, 2005). This indicates that CHAP proteins present diverse roles within the cell.

In a previous work, the C-terminal end of Cse, including the CHAP domain, was demonstrated to be homologous with that of PcsB/GbpB (Borges *et al.*, 2005). Many attempts were done to demonstrate the potential peptidoglycan hydrolytic activity of these enzymes, but all failed (Borges *et al.*, 2005; Ng *et al.*, 2004; Reinscheid *et al.*, 2001). This time, by over expressing the CHAP domain of Cse, we succeeded to demonstrate its endopeptidase activity confirming that Cse is indeed a peptidoglycan hydrolase. Accordingly, the CHAP domain of several phage-encoded lytic proteins also displays an endopeptidase activity (Baker *et al.*, 2006; Donovan *et al.*, 2006; Navarre *et al.*, 1999; Nelson *et al.*, 2006; Pritchard *et al.*, 2004; Yokoi *et al.*, 2005). An alternative catalytic activity was reported for the CHAP proteins Sk1 from the *Streptococcus mitis* phage SK137 and Sle1 from *S. aureus* that both function as *N*-acetylmuramoyl-L-alanine amidase (Kajimura *et al.*, 2005; Llull *et al.*, 2006).

The process of daughter cell separation can be viewed as an active process that takes place at the end or after the completion of the cell division process. Indeed, cells lacking cell-separating enzyme are inhibited only in cell separation and are not inhibited in cell division, with normal septa still being formed (Ohnishi *et al.*, 1999). The cell separation process requires dedicated PGHs that cleave the sacculi connecting daughter cells. In their vast majority, cell separating enzymes are composed of a unique CHAP, NLPC/P60 or Glucosaminidase (more rarely Glyco\_hydro\_25) catalytic domain (Layec *et al.*, 2008b). The catalytic domain diversity is correlated with the diversity of the peptidoglycan linkage that they cleave. For instance, AcmA from *L. lactis* (Buist *et al.*, 1995; Steen *et al.*, 2005a) and LytB from *S. pneumoniae* (De Las Rivas *et al.*, 2002; Garcia *et al.*, 1999) carry a glucosaminidase domain conferring the *N*-acetylglucosaminidase activity. Ami A, B and C from *E. coli* and Sle1 from *S. aureus* are *N*-acetylmuramoyl-L-alanine amidases. Aml from *Streptococcus mutans*, carrying a “Glyco\_hydro\_25” domain, presents a *N*-acetylmuramidase activity (Yoshimura *et al.*, 2006). At last, several cell separating enzymes display an endopeptidase activity. That is the case of Cse from *S. thermophilus* (this work) and of LytE, LytF and CwlS (YolJ) from *B. subtilis* (Fukushima *et al.*, 2006; Margot *et al.*, 1998; Margot *et al.*, 1999). The hydrolytic specificity of the CHAP domain of Cse is distinct from the NLPC/P60 domain of LytF and CwlS since hydrolyzing the linkage between *meso*-diaminopimelic acid and D-alanyl residues (this work) rather than the linkage of D- $\gamma$ -glutamyl-*meso*-diaminopimelic acid of the *B. subtilis* cell wall (Fukushima *et al.*, 2006; Margot *et al.*, 1999). Collectively, all these data clearly show that evolution has selected several alternative enzymes exhibiting distinct catalytic activity to perform the daughter cell separation process.

All the cell separating enzymes described so far are cell surface proteins and Cse does not escape this rule. Cse is thought to bind peptidoglycan via its LysM domain. The LysM domain is one of the most common attachment modules in bacterial cell separating enzymes. Thus, AcmA from *L. lactis*, Sle1 from *S. aureus*, LytE, LytF and CwlS from *Bacillus* contain 2 to 5 tandem repeats of the LysM motif (Fukushima *et al.*, 2006; Ishikawa *et al.*, 1998; Kajimura *et al.*, 2005; Margot *et al.*, 1998; Margot *et al.*, 1999; Ohnishi *et al.*, 1999; Steen *et al.*, 2003). In addition to its role in peptidoglycan binding, LysM is important for biological functioning (Buist *et al.*, 2008). The chaining phenotype of the *cse* $\Delta$ LysM mutant corroborated these results since LysM was found to be necessary for efficient Cse cell separating activity.

The location of only a few cell separating enzymes has been studied so far but reveals that all of them are positioned at cell separation sites and cell poles (Fukushima *et al.*, 2006; Kajimura *et al.*, 2005; Steen *et al.*, 2003). The peculiarity of Cse is that this protein settled mainly (if not only) at mature septa and questions about the reasons leading to such exclusive location.

One explanation is to consider that Cse cleavage specificity would be ensured by the restricted binding of the enzyme to its cleavage site thanks to its LysM domain. Another possibility to explain the exclusive binding of Cse to mature septa is to consider that a specific mechanism targets Cse to its site of action. Recently, Carlsson et al (2006) (Carlsson *et al.*, 2006) demonstrated that the signal sequence of M protein promotes secretion at the division septum, whereas that of PrtF preferentially promotes secretion at the old pole. Thus, the possibility that the Cse signal peptide governs its secretion to mature septum is not excluded.

Within Cse protein, other regions were found to be important for Cse positioning. Indeed, depletion of the CHAP domain (Fig. 4D) or of the Var region (supplementary data Fig. 7C) led to a modification of Cse binding on *S. thermophilus* cell walls. Although we cannot exclude a direct role for these regions on cell wall attachment and/or substrate recognition, these data suggest that the tridimensional conformation of the protein could be involved in protein positioning on the cell surface. At last, not only the structure protein but also the cell wall constituents and probably its architecture possibly interfere with protein attachment. Indeed, studies of AcmA led to the conclusion that the binding of this autolysin over the entire surface is hindered by lipoteichoic acid a cell wall compound associated with glycan molecules (Steen *et al.*, 2005b).

Data from this work indicate that Cse is not involved in cell wall synthesis but only takes part in the very late stage of (or after) the cellular division to catalyse the separation of daughter cells. It demonstrates that cell separation is an active process that requires the action of dedicated PGHs.

## EXPERIMENTAL PROCEDURES

### Bacterial strains, plasmids, media and growth conditions.

Bacterial strains and plasmids used in this study are listed in Table 1 (Supplementary data). *S. thermophilus* CNRZ368 and its derivatives were cultivated in milk medium, M17 (Terzaghi and Sandine, 1975) or HJL (Stingele and Mollet, 1996). Milk medium was used for strain storage, M17 for mutant generation, and Hogg Jago Lactose (HJL) for phenotypic analysis. When required, erythromycin (final concentration 2  $\mu\text{g}\cdot\text{ml}^{-1}$ ) was added. *S. thermophilus* strains containing pG<sup>+</sup>host 9 (Maguin *et al.*, 1992) or pNST260<sup>+</sup> 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. Phenotypic analyses were all performed at 42°C. *S. thermophilus* CNRZ368 $\Delta$ cse mutant was cultured in HJL at 42°C and was used (i) as a substrate in renaturing SDS-PAGE and (ii) to prepare cell-wall extracts for the zymography assays. *E. coli* strains were grown in Luria-bertani (LB) medium (Sambrook *et al.*, 1989) at 37°C with shaking. When required erythromycin or ampicillin was added to a final concentration of 150  $\mu\text{g}/\text{ml}$ . *E. coli* EC101 (Leenhouts, 1995) was used as a host of recombinant plasmids derived from pG<sup>+</sup>host 9. *E. coli* strain DH5 $\alpha$  was used as a host of pET15b derivatives, and *E. coli* strain BL21(DE3) (Stratagene) was used as a host to overexpress the LysM<sub>Cse</sub> and CHAP fusion (His<sub>6</sub>-tagged) proteins. *B. subtilis* 168 HR was cultured in LB medium at 37°C and was used to prepare cell-wall extracts for the study of (i) lytic activity by zymogram and (ii) hydrolytic bond specificity.

### DNA manipulations

Plasmids and primers used in this study are listed in Table 1 and Table 2 (Supplementary data), respectively.

#### (i) Construction of *S. thermophilus* CNRZ368 cse $\Delta$ LysM and cse $\Delta$ CHAP mutant strains.

The *S. thermophilus* CNRZ368 strain was used in this study as the wild-type strain. *S. thermophilus* mutant strains used in this study are listed in Table 1. The LysM and the CHAP module of the *cse* gene were deleted by allelic replacement as previously described (Thibessard *et al.*, 2004). To make deletions, the two regions flanking the LysM or CHAP domains were independently amplified by PCR using the following primers pairs: dellysME5'/dellysMI5' and dellysMI3'/dellysME3' or delchapE5'/delchapI5' and delchapI3'/delchapE3', respectively (Table 2). The PCR products were digested by appropriate restriction enzymes, joined together and ligated with the plasmid pG<sup>+</sup>host9 (Maguin *et al.*, 1992). The ligation products were used to transform *E. coli* EC101. After transformation of *S. thermophilus* with the recombinant plasmid, deleted mutants displaying an *EcoRI* restriction site replacing 459 pb of the LysM module or 318 bp of the CHAP module were selected. Deletion mutants were checked by PCR, Southern hybridization and sequencing (data not shown).

#### (ii) cse $\Delta$ LysM and cse $\Delta$ CHAP complementation.

The complementation of cse $\Delta$ lysM or cse $\Delta$ CHAP mutants was carried out by inserting the pNST260+::cse plasmid in their chromosome as previously described (Borges *et al.*, 2005).

#### (iii) Construction of pET15b::lysM<sub>cse</sub> and pET15b::chap for LysM<sub>Cse</sub>- and CHAP-His<sub>6</sub>-tagged proteins overexpression.

*S. thermophilus* CNRZ368 DNA fragments encoding (i) the 5' end of *cse* including the cell wall-binding LysM module (from 1 to 549 bp) and (ii) the catalytic CHAP module (from 1021 to 1383 bp) of *cse* were amplified by PCR. The amplified DNA fragments were purified, double digested with *NdeI* and *BamHI*, repurified, and fused in-frame to the 3'end His<sub>6</sub>-tagged of pET15b (Novagen), generating the pET15b::lysM<sub>cse</sub> and pET15b::chap plasmids, respectively (Table 1). The in-frame fusion was confirmed by DNA sequencing.

#### Overexpression, purification of His<sub>6</sub>-tagged proteins and western blotting.

Expression of the His<sub>6</sub>-tagged proteins was induced in *E. coli* BL21(DE3) containing pET15::lysM<sub>Cse</sub> or pET15b::chap by adding IPTG (to 1 mM final concentration) at OD<sub>600</sub> = 0.5 for 4h at 28°C. For

extraction, cells were harvested by centrifugation and disrupted on ice with a microtip of Sonifier 250 (Branson Ultrasonics). The soluble fraction containing the recombinant protein was collected by centrifugation at 14,000 rpm for 30 min at 4°C. The His<sub>6</sub>-tagged proteins were purified by a HiTrap chelating HP column according to the manufacturer's instructions (5ml, Amersham Biosciences). The LysM<sub>Cse</sub>- and CHAP-His<sub>6</sub>-tagged proteins have a predicted molecular mass of 21.5 kDa and 15.1 kDa, respectively. The purity of the His<sub>6</sub>-tagged proteins was confirmed by SDS-15% [wt/vol] polyacrylamide gel electrophoresis (PAGE) and was estimated >96% by quantity one software (BioRad). The LysM<sub>Cse</sub>-His<sub>6</sub>-tagged protein was used for antibodies generation; the CHAP-His<sub>6</sub>-tagged protein was used to demonstrate Cse enzymatic properties.

For Western blotting, proteins were transferred onto PVDF membrane (Roche) using a semi-dry blotting system (Bio-Rad) following standard protocols (Towbin *et al.*, 1992). To control the presence of the His-tag on the fusion proteins, immunodetection assays were performed using rabbit IgG against RGS-His tag (Qiagen, at dilution 1:2000). The ECL Western blotting detection was done according to the manufacturer's instructions (Roche) and visualized by FluorS' max (Bio-Rad).

### Cse immunolocalization.

#### (i) Production of anti-LysM<sub>Cse</sub> antibodies.

Polyclonal anti-LysM<sub>Cse</sub> were obtained from New Zealand White Rabbit inoculated with a mixture of 500 µg of purified LysM<sub>Cse</sub>-His<sub>6</sub>-tagged protein and Freund's complete adjuvant (Sigma-Aldrich) followed by booster two times with LysM<sub>Cse</sub>-His<sub>6</sub>-tagged protein and incomplete adjuvant every three weeks. The purified LysM<sub>Cse</sub>-His<sub>6</sub>-tagged protein was separated by SDS-15% [wt/vol] polyacrylamide gel, cut from the gel, fragmented and used for immunization. The specificity of the LysM<sub>Cse</sub> antibodies was verified on the purified LysM<sub>Cse</sub>-His<sub>6</sub>-tagged protein (at dilution 1:10,000) and *S. thermophilus* total proteins extract (at dilution 1:5000) by Western blotting analysis. In the soluble fraction of *S. thermophilus* CNRZ368 cells extract, one approximately 45.5 kDa band compatible with Cse molecular mass without its signal peptide (theoretical molecular mass with signal peptide, 48.5 kDa) was detected with polyclonal anti-LysM<sub>Cse</sub> antibodies (data not shown). No band was discovered in *S. thermophilus* CNRZ368Δcse mutant cells extract, used as a negative control (data not shown).

#### (ii) Preparation of *S. thermophilus* samples for Cse immunolocalization.

After overnight growth at 42°C in HJL medium, samples (2 ml) of *S. thermophilus* were pelleted by centrifugation (3000 rpm for 20 min at 4°C) and washed two times with 2 ml phosphate buffered saline (81 mM NaCl, 1.62 mM KCl, 0.9 mM KH<sub>2</sub>PO<sub>4</sub>, 4.8 mM K<sub>2</sub>HPO<sub>4</sub> [pH 7.2]; PBS buffer). The cell pellets were resuspended in 2 ml PBS buffer and fixed by addition of 4% [vol/vol] final concentration paraformaldehyde and 0.2% [vol/vol] final concentration glutaraldehyde for 3 h of incubation on ice. Cells were harvested by centrifugation and washed four times in 2 ml PBS buffer containing 0.05 M glycine. Cells were resuspended in 1 ml PBS buffer. A sample of 100 µl was centrifugated and resuspended in 50 µl PBS buffer containing polyclonal anti-LysM<sub>Cse</sub> primary antibodies and incubated for 12 h at 4°C with gentle rolling. After washing, cells were resuspended in 50 µl of PBS buffer containing Alexa fluor 488-conjugated anti-rabbit IgG antibodies (1:400) (Molecular-Probes) and observed by Fluorescence Microscopy. For Transmission Electron Microscopy (TEM), cells were resuspended in 50 µl of PBS buffer containing anti-rabbit IgG conjugated with 10 nm colloidal gold secondary antibodies (1:50) (Sigma-Aldrich). The fluorochrome calcofluor (Sigma-Aldrich) was used on *S. thermophilus* cells with a 0.01% [wt/vol] solution to reveal peptidoglycan.

#### (iii) Microscopy.

Immunofluorescence microscopy was performed with a Bio-Rad LRC Radiance 2100 (Elexience, Paris, France) confocal laser scanning microscope (CLSM), attached to Nikon TE2000-U inverted microscope, using an excitation at 488 nm from 100 mW argon ion laser and appropriate barrier filters. Cells were observed through x60 Nikon immersion oil plan apo (NA 1.40). Images were contrast enhanced in Adobe Photoshop (Adobe Systems Inc., Mountain View, USA). A sample of the bacterial cell suspension was deposited onto formvar carbon-coated copper nickel grids. Cells were examined with a Zeiss 902A transmission electron microscope at 80kV (magnification x 20,000 or

x 30,000). *S. thermophilus* chaining phenotypes were observed with a Nikon OPTIPHOT microscope mounted with phase-contrast equipment “Ph”. Cells were photographed with a Nikon coolpix 5400 digital camera.

### **Peptidoglycan hydrolase activity.**

#### **(i) SDS-PAGE and zymography.**

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were carried out as previously described (Laemmli, 1970) with a protean II minigel system (Bio-Rad). Zymograms were performed with 15% [wt/vol] polyacrylamide separating gels containing either lyophilized of autoclaved cells (from *M. lysodeikticus* ATCC4698 (Sigma, 0.2% [wt/vol]) or from *S. thermophilus* CNRZ368 $\Delta$ cse mutant (0.4% [wt/vol])) or purified cell-wall peptidoglycan 0.08% [wt/vol] (from *S. thermophilus* CNRZ368 $\Delta$ cse mutant or *B. subtilis*). Peptidoglycan from *S. thermophilus* and *B. subtilis* 168 HR vegetative cells were prepared as described previously (Courtin *et al.*, 2006; Meyrand *et al.*, 2007).

After electrophoresis, gels were washed two times for 30 min. in 3% [vol/vol] Triton X-100 and then incubated in renaturation buffer overnight at 30°C. Proteins renaturation was performed with: 25 mM sodium citrate/ 50 mM sodium phosphate (Huard *et al.*, 2003) or 50 mM KH<sub>2</sub>PO<sub>4</sub> buffer containing 0.2% [vol/vol] Triton X-100 at pH 3 to 7. The gels were stained with 0.1% [wt/vol] methylene blue in 0.01% [wt/vol] KOH for 4 h at room temperature and destained in deionized H<sub>2</sub>O. Gel images were generated with GS-800 Calibrated Densitometer Scanner (Bio-Rad).

#### **(ii) Determination of CHAP hydrolytic specificity.**

Peptidoglycan from *B. subtilis* 168 HR and *S. thermophilus* was extracted as described previously (Meyrand *et al.*, 2007). *B. subtilis* peptidoglycan (4 mg) was incubated overnight at 37°C with purified CHAP-His<sub>6</sub>-tagged protein (500  $\mu$ g) in a final volume of 500  $\mu$ l buffer (25 mM sodium citrate/ 50 mM sodium phosphate containing 0.2% Triton X-100 at pH 4.0). The sample was boiled for 3 min to stop the reaction and then centrifuged at 14 000 g for 15 min. The insoluble material as well as half of the soluble fraction, both recovered after incubation with CHAP-His<sub>6</sub>-tagged protein, were further treated with mutanolysin (2 500 U.ml<sup>-1</sup>) overnight at 37°C. A control digest of *B. subtilis* peptidoglycan was obtained by mutanolysin digestion as described (Courtin *et al.*, 2006). The soluble muropeptides obtained after the digestions were reduced with sodium borohydride and separated by reverse phase HPLC using a Hypersil ODS column (C18, 250 x 4.6 mm; 5  $\mu$ m; ThermoHypersil-Keystone) at 50°C using ammonium phosphate buffer and a methanol linear gradient as described previously (Atrih *et al.*, 1999; Courtin *et al.*, 2006). Peaks were analysed without desalting by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry with a Voyager DE STR mass spectrometer (Applied Biosystems) as described previously (Courtin *et al.*, 2006).

*S. thermophilus* peptidoglycan was digested with mutanolysin overnight at 37°C and the resulting soluble muropeptides were reduced and separated by RP-HPLC as described previously (Hebert *et al.*, 2007). Three muropeptides were collected, dried in a Speed-vac concentrator under vacuum. The purified muropeptides were incubated with 50  $\mu$ g of purified CHAP-His<sub>6</sub>-tagged protein in a total volume of 200  $\mu$ l in the conditions described above for *B. subtilis* peptidoglycan. The reaction products were then analyzed by RP-HPLC and MALDI-TOF mass spectrometry as described above.

## **ACKNOWLEDGMENTS**

This study is supported by grants from the Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche.

We thank Amand Chesnel for polyclonal antibodies production and Gérard Guédon for allowing us to use pNST260<sup>+</sup> prior to publication.

## REFERENCES

- Atrih, A., Bacher, G., Allmaier, G., Williamson, M.P., and Foster, S.J. (1999) Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *J Bacteriol* **181**: 3956-3966.
- Baker, J.R., Liu, C., Dong, S., and Pritchard, D.G. (2006) Endopeptidase and glycosidase activities of the bacteriophage B30 lysin. *Appl Environ Microbiol* **72**: 6825-6828.
- Biswas, I., Gruss, A., Ehrlich, S.D., and Maguin, E. (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol* **175**: 3628-3635.
- Borges, F., Layec, S., Thibessard, A., Fernandez, A., Gintz, B., Hols, P., Decaris, B., and Leblond-Bourget, N. (2005) *cse*, a Chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J Bacteriol* **187**: 2737-2746.
- Borges, F., Layec, S., Fernandez, A., Decaris, B., and Leblond-Bourget, N. (2006) High genetic variability of the *Streptococcus thermophilus cse* central part, a repeat rich region required for full cell segregation activity. *Antonie Van Leeuwenhoek* **90**: 245-255.
- Buist, G., Kok, J., Leenhouts, K.J., Dabrowska, M., Venema, G., and Haandrikman, A.J. (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.
- Buist, G., Steen, A., Kok, J., and Kuipers, O.P. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* **68**: 838-847.
- Carlsson, F., Stalhammar-Carlemalm, M., Flardh, K., Sandin, C., Carlemalm, E., and Lindahl, G. (2006) Signal sequence directs localized secretion of bacterial surface proteins. *Nature* **442**: 943-946.
- Chia, J.S., Chang, L.Y., Shun, C.T., Chang, Y.Y., Tsay, Y.G., and Chen, J.Y. (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.
- Courtin, P., Miranda, G., Guillot, A., Wessner, F., Mezange, C., Domakova, E., Kulakauskas, S., and Chapot-Chartier, M.P. (2006) Peptidoglycan structure analysis of *Lactococcus lactis* reveals the presence of an L,D-carboxypeptidase involved in peptidoglycan maturation. *J Bacteriol* **188**: 5293-5298.
- De Las Rivas, B., Garcia, J.L., Lopez, R., and Garcia, P. (2002) Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* **184**: 4988-5000.
- Donovan, D.M., Foster-Frey, J., Dong, S., Rousseau, G.M., Moineau, S., and Pritchard, D.G. (2006) The cell lysis activity of the *Streptococcus agalactiae* bacteriophage B30 endolysin relies on the cysteine, histidine-dependent amidohydrolase/peptidase domain. *Appl Environ Microbiol* **72**: 5108-5112.
- Eckert, C., Lecerf, M., Dubost, L., Arthur, M., and Mesnage, S. (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. *J Bacteriol* **188**: 8513-8519.
- Fukushima, T., Afkham, A., Kurosawa, S., Tanabe, T., Yamamoto, H., and Sekiguchi, J. (2006) A new D,L-endopeptidase gene product, YojL (renamed CwlS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis*. *J Bacteriol* **188**: 5541-5550.
- Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R., and Garcia, J.L. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol* **31**: 1275-1281.
- Hebert, L., Courtin, P., Torelli, R., Sanguinetti, M., Chapot-Chartier, M.P., Auffray, Y., and Benachour, A. (2007) *Enterococcus faecalis* constitutes an unusual bacterial model in lysozyme resistance. *Infect Immun* **75**: 5390-5398.
- Heidrich, C., Templin, M.F., Ursinus, A., Merdanovic, M., Berger, J., Schwarz, H., de Pedro, M.A., and Holtje, J.V. (2001) Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**: 167-178.
- Huard, C., Miranda, G., Wessner, F., Bolotin, A., Hansen, J., Foster, S.J., and Chapot-Chartier, M.P. (2003) Characterization of AcnB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*. *Microbiology* **149**: 695-705.

- Ishikawa, S., Hara, Y., Ohnishi, R., and Sekiguchi, J. (1998) Regulation of a new cell wall hydrolase gene, *cwlF*, which affects cell separation in *Bacillus subtilis*. *J. Bacteriol.* **180**: 2549-2555.
- Kajimura, J., Fujiwara, T., Yamada, S., Suzawa, Y., Nishida, T., Oyamada, Y., Hayashi, I., Yamagishi, J., Komatsuzawa, H., and Sugai, M. (2005) Identification and molecular characterization of an *N*-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of *Staphylococcus aureus*. *Mol Microbiol* **58**: 1087-1101.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Danchin, A., and et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Layec, S., Decaris, B., and Leblond-Bourget, N. (2008a) Characterization of proteins belonging to the CHAP-related superfamily within the *Firmicutes*. *J Mol Microbiol Biotechnol* **14**.
- Layec, S., Decaris, B., and Leblond-Bourget, N. (2008b) Diversity of *Firmicutes* peptidoglycan hydrolases and specificities of those involved in daughter cell separation. *Res Microbiol.* (in press)
- Leenhouts, K. (1995) Integration strategies and vectors. *Dev Biol Stand* **85**: 523-530.
- Llull, D., Lopez, R., and Garcia, E. (2006) Skl, a novel choline-binding *N*-acetylmuramoyl-L-alanine amidase of *Streptococcus mitis* SK137 containing a CHAP domain. *FEBS Lett* **580**: 1959-1964.
- Maguin, E., Duwat, P., Hege, T., Ehrlich, D., and Gruss, A. (1992) New thermosensitive plasmid for gram-positive bacteria. *J Bacteriol* **174**: 5633-5638.
- Margot, P., Wahlen, M., Gholamhoseinian, A., Piggot, P., and Karamata, D. (1998) The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. *J. Bacteriol.* **180**: 749-752.
- Margot, P., Pagni, M., and Karamata, D. (1999) *Bacillus subtilis* 168 gene *lytF* encodes a gamma-D-glutamate-meso-diaminopimelate muropeptidase expressed by the alternative vegetative sigma factor, sigmaD. *Microbiology* **145**: 57-65.
- Mattos-Graner, R.O., Porter, K.A., Smith, D.J., Hosogi, Y., and Duncan, M.J. (2006) Functional analysis of glucan binding protein B from *Streptococcus mutans*. *J Bacteriol* **188**: 3813-3825.
- Meyrand, M., Boughammoura, A., Courtin, P., Mezange, C., Guillot, A., and Chapot-Chartier, M.P. (2007) Peptidoglycan *N*-acetylglucosamine deacetylation decreases autolysis in *Lactococcus lactis*. *Microbiology* **153**: 3275-3285.
- Navarre, W.W., Ton-That, H., Faull, K.F., and Schneewind, O. (1999) Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity. *J Biol Chem* **274**: 15847-15856.
- Nelson, D., Schuch, R., Chahales, P., Zhu, S., and Fischetti, V.A. (2006) PlyC: a multimeric bacteriophage lysin. *Proc Natl Acad Sci U S A* **103**: 10765-10770.
- Ng, W.L., Kazmierczak, K.M., and Winkler, M.E. (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.
- Ohnishi, R., Ishikawa, S., and Sekiguchi, J. (1999) Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J Bacteriol* **181**: 3178-3184.
- Pritchard, D.G., Dong, S., Baker, J.R., and Engler, J.A. (2004) The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology* **150**: 2079-2087.
- Rashel, M., Uchiyama, J., Takemura, I., Hoshihara, H., Ujihara, T., Takatsuji, H., Honke, K., and Matsuzaki, S. (2008) Tail-associated structural protein gp61 of *Staphylococcus aureus* phage phi MR11 has bifunctional lytic activity. *FEMS Microbiol Lett* **284**: 9-16.
- Reinscheid, D.J., Gottschalk, B., Schubert, A., Eikmanns, B.J., and Chhatwal, G.S. (2001) Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J Bacteriol* **183**: 1175-1183.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*.

- Schleifer, K.H., and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36**: 407-477.
- Steen, A., Buist, G., Leenhouts, K.J., El Khattabi, M., Grijpstra, F., Zomer, A.L., Venema, G., Kuipers, O.P., and Kok, J. (2003) Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J Biol Chem* **278**: 23874-23881.
- Steen, A., Buist, G., Horsburgh, G.J., Venema, G., Kuipers, O.P., Foster, S.J., and Kok, J. (2005a) AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. *Febs J* **272**: 2854-2868.
- Steen, A., Palumbo, E., Deghorain, M., Cocconcelli, P.S., Delcour, J., Kuipers, O.P., Kok, J., Buist, G., and Hols, P. (2005b) Autolysis of *Lactococcus lactis* is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids. *J Bacteriol* **187**: 114-124.
- Stingele, F., and Mollet, B. (1996) Disruption of the gene encoding penicillin-binding protein 2b (*pbp2b*) causes altered cell morphology and cease in exopolysaccharide production in *Streptococcus thermophilus* Sfi6. *Mol Microbiol* **22**: 357-366.
- Terzaghi, B.E., and Sandine, W.E. (1975) Improved Medium for Lactic Streptococci and Their Bacteriophages. *Appl Microbiol* **29**: 807-813.
- Thibessard, A., Borges, F., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2004) Identification of *Streptococcus thermophilus* CNRZ368 Genes Involved in Defense against Superoxide Stress. *Appl Environ Microbiol* **70**: 2220-2229.
- Towbin, H., Staehelin, T., and Gordon, J. (1992) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24**: 145-149.
- Yamamoto, H., Kurosawa, S., and Sekiguchi, J. (2003) Localization of the vegetative cell wall hydrolases LytC, LytE, and LytF on the *Bacillus subtilis* cell surface and stability of these enzymes to cell wall-bound or extracellular proteases. *J Bacteriol* **185**: 6666-6677.
- Yokoi, K.J., Kawahigashi, N., Uchida, M., Sugahara, K., Shinohara, M., Kawasaki, K., Nakamura, S., Taketo, A., and Kodaira, K. (2005) The two-component cell lysis genes holWMY and lysWMY of the *Staphylococcus warneri* M phage varphiWMY: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. *Gene* **351**: 97-108.
- Yoshimura, G., Komatsuzawa, H., Hayashi, I., Fujiwara, T., Yamada, S., Nakano, Y., Tomita, Y., Kozai, K., and Sugai, M. (2006) Identification and molecular characterization of an N-Acetylmuraminidase, Aml, involved in *Streptococcus mutans* cell separation. *Microbiol Immunol* **50**: 729-742.



## LEGEND FIGURES

### FIG. 1. Mutation of *cse* alters *S. thermophilus* cell separation.

(A.) Detail of the *cse* structure in the wild type and mutant strains. The positions in nucleotides delimit the signal peptide (SP), the LysM cell wall-binding module, the central variable region (Var) and the catalytic CHAP module of *S. thermophilus* CNRZ368 *cse* gene. Arrows and hairpin loops symbolize putative promoters and putative rho-independent transcriptional terminators, respectively. Positions of the initial and of the final deleted nucleotides are indicated. (B.) Picture of standing HJL cultures and phase contrast microscopy of the strains. Photographs were taken after 20 h of growth at 42°C in HJL medium. Magnification, x1,000. (C.) Chain length analysis. Cells were counted in stationary phase after 20 h of growth at 42°C in HJL medium. The error bars represent standard deviations of three independent experiments.

**FIG. 2. *S. thermophilus* chaining phenotype observed by transmission electron microscopy (TEM).** The  $\Delta cse$  mutant chaining phenotype is shown in panels A to C, that of the wild type in D to F. Cells were visualized after 20 h of growth at 42°C in HJL medium. Arrowheads in panels C and F pointed distinct septa. Diplococcal dividing cells are characterized by the in-growth of the septum (nascent septum, a) initiated by the constriction ring located equidistant from the cell poles. The progression of the septal peptidoglycan synthesis is materialized by the further constriction of the cell walls (intermediate septum, b). At the end of the division process, diplococcal cells are held together by a thin peptidoglycan filamentous corresponding to an extension of the septum (mature septum, c). Bars = 0.5  $\mu$ m.

### FIG. 3. Immunolocalization of Cse on *S. thermophilus* cell surface.

The TEM pictures of the wild type strain are shown in A to D and that of the  $\Delta cse$  mutant in E and F. Cells were successively probed with anti-LysM<sub>Cse</sub> rabbit polyclonal antibodies (A,B,C,E and F) or with preimmun serum (D) and with anti-rabbit IgG conjugated with 10 nm colloidal gold secondary antibody. The localization of Cse protein is indicated by the arrowheads pointed on gold-conjugates secondary antibodies (A to C). No labelling was observed on the negative controls (D to F). Bars = 0.1  $\mu$ m.

**FIG. 4. Exclusive Cse localization on *S. thermophilus* CNRZ368 wild-type cell-wall peptidoglycan mature septa.** CNRZ368 wild-type (A, B and E), *cse* $\Delta$ LysM (C) and *cse* $\Delta$ CHAP (D) cells are shown. Surface bound Cse and derivatives proteins were detected by immunofluorescence microscopy using anti-LysM<sub>Cse</sub> antibodies Phase-contrast images (A1 to E1) and fluorescence images (A2 to B2 and E2 to E4) of the bacterial cells in the same field are shown. The arrowheads pointed mature septa. No fluorescence staining on the wild-type cell surface was observed using the preimmun serum (B2), as well as on the *cse* $\Delta$ LysM mutant cells with the anti-LyM<sub>Cse</sub> antibodies (C2), as negative controls. Bars = 25  $\mu$ m.

E2 shows a single-labeled with anti-LysM<sub>Cse</sub> antibodies that indicates the localization of Cse exclusively on mature septa (Bars = 5  $\mu$ m). E3 shows the fluorochrome calcofluor having an affinity for peptidoglycan. E4 is the result double-labelled.

**FIG. 5. Zymographic analysis of CHAP domain activity.** (a) SDS-15% PAGE with Coomassie brilliant blue staining and (b) Zymogram assay was performed with 15  $\mu$ g of purified CHAP-His<sub>6</sub>-tagged protein and 0.08% *S. thermophilus* CNRZ368 $\Delta cse$  cell-wall peptidoglycan as substrate. Renaturing buffer contained 0.2% TritonX-100. CHAP domain activity is visible as clearing zones in the gel (lane 6). Total proteins of *E. coli* BL21(DE3) + pET15b (lane 1: cell extract and lane 2: supernatant) and *E. coli* BL21(DE3) + pET15b::chap (lane 3: cell extract and lane 4: supernatant) and purified CHAP-His<sub>6</sub>-tagged protein (lane 5) are shown. Molecular masses of standard proteins (MM) are shown on the left.

**FIG. 6. Determination of the cleavage site of the CHAP catalytic domain of Cse by RP-HPLC.** (A) The cell wall of *B. subtilis* was digested with mutanolysin (a) or the CHAP-His<sub>6</sub>-tagged protein and mutanolysin (b). The numbers indicate the peaks analyzed by MALDI-TOF MS referred in Table 3B. (B) *S. thermophilus* mucopeptides (Table 4) digested by the CHAP-His<sub>6</sub>-tagged protein. (C) Schematic structure of *B. subtilis* (C1) and *S. thermophilus* (C2) cell-wall peptidoglycan and sites of cleavage of the CHAP domain.

GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid, mDAP, *meso*-diaminopimelic acid.

**FIG. 7. Immunolocalization of Cse protein derivatives on the cell surface of *S. thermophilus* (Supplementary data).**

*S. thermophilus* CNRZ368 cells of the  $\Delta cse$  (A), the  $cse\Delta CHAP$  (B) and the  $cse\Delta Var$  (C and D) mutants are shown. Cells were visualized with a laser scanning confocal microscope. Phase-contrast images (A1 to D1) and fluorescence images (A2 to D2) of the bacterial cells are shown in the same field. No fluorescence was detected on negative controls:  $\Delta cse$  cells with anti-LysM<sub>Cse</sub> antibodies (A2); from  $cse\Delta CHAP$  to  $cse\Delta var$  cells with preimmun serum (B2 and D2). C2 indicates the localization of Cse derivative protein using the LysM<sub>Cse</sub> antibodies. (Bars = 25  $\mu$ m)

**FIG. 8. Overexpression, purification, western-blotting and lytic activity of the CHAP-His<sub>6</sub>-tagged protein. (Supplementary data).**

(A) Western blot obtained using RGS-His tag antibodies raised against CHAP-His<sub>6</sub>-tagged protein. (B) Cell wall hydrolytic activity of the CHAP-His<sub>6</sub>-tagged protein in renaturing SDS-15% polyacrylamide gel containing 0.2% [wt/vol] *M. lysodeikticus* ATCC4698 autoclaved cells. The CHAP-His<sub>6</sub>-tagged protein activity is visible as clearing zones. Lanes 1 and 2, cell extract and supernatant of *E. coli* BL21(DE3) containing pET15b. Lanes 3 and 4, cell extract and supernatant of *E. coli* BL21(DE3) containing pET15b::chap. Lane 5, purified CHAP-His<sub>6</sub>-tagged protein. Lane MM, the molecular masses of standard proteins. The arrow indicates the position of purified CHAP-His<sub>6</sub>-tagged protein.

**Table 1.** Bacterial strains and plasmids used in this study (Supplementary data).

| Strain or plasmid                     | Genotype/phenotype/ source   | Origin/reference   |
|---------------------------------------|--|--|
| <b><i>S. thermophilus</i> strains</b> |  |  |
| CNRZ368                               | Wild-type; isolated from yogurt.   | INRA-CNRZ, strain collection, France                       |
| CNRZ368 $\Delta$ cse                  | Derivative of CNRZ368 carrying a 1359 bp deletion in the cse gene  | (Borges <i>et al.</i> , 2005)                              |
| CNRZ368cse $\Delta$ LysM              | Derivative of CNRZ368 carrying a 459 bp deletion in the cse LysM module  | This study   |
| CNRZ368cse $\Delta$ Var               | Derivative of CNRZ368 carrying a 345 bp deletion in the var- cse region  | (Borges <i>et al.</i> , 2005)                              |
| CNRZ368cse $\Delta$ CHAP              | Derivative of CNRZ368 carrying a 318 bp deletion in the cse CHAP module  | This study   |
| <b><i>E. coli</i> strains</b>         |  |  |
| EC101                                 | <i>supE hsd-5 thi <math>\Delta</math>(lac-proAB) F'(traD6 proAB<sup>+</sup> lacI<sup>q</sup> lacZ<math>\Delta</math>M15) repA<sup>+</sup></i> , derivative of TG1 strain | (Leenhouts, 1995; Sambrook <i>et al.</i> , 1989)           |
| DH5 $\alpha$                          | F <sup>-</sup> , <i>endA1, gyrA96, thi-1, hsdR17(r<sub>K</sub><sup>-</sup>, m<sub>K</sub><sup>+</sup>), supE44, rel A1, M15, <math>\Delta</math>(lacZYA-argF), U169</i>  | (Sambrook <i>et al.</i> , 1989)                            |
| BL21(DE3)                             | <i>E. coli</i> B, F <sup>-</sup> , <i>dcm, ompT, hsdS(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>), gal<math>\lambda</math></i> (DE3)                            | Stratagene   |
| <b><i>B. subtilis</i></b>             |  |  |
| 168 HR                                | <i>trpC2</i>   | (Kunst <i>et al.</i> , 1997)                               |
| <b>Plasmids</b>                       |  |  |
| pG <sup>+</sup> host9                 | Thermosensitive plasmid, used for gene replacement; Ery <sup>R</sup>   | (Biswas <i>et al.</i> , 1993; Maguin <i>et al.</i> , 1992) |
| pGh9 ::cse $\Delta$ lysM              | Derivative of pG <sup>+</sup> host9 carrying a 1098 bp for LysM deletion   | This study   |
| pGh9 ::cse $\Delta$ CHAP              | Derivative of pG <sup>+</sup> host9 carrying a 924 bp for CHAP deletion  | This study   |
| pNST260+                              | Derivative of pG <sup>+</sup> host9 carrying a <i>int</i> gene coding a integrase and an <i>attI</i> from ICES $\Delta$ 1, specific-site of recombinaison                | G. Guédon, personal communication                          |
| pNST260+ ::cse                        | pNST260+ with CNRZ368 <i>cse</i> gene, its putative promoter and terminator  | (Borges <i>et al.</i> , 2005)                              |
| pET15b                                | Expression vector for N-terminal His <sub>6</sub> -tagged fusion; Amp <sup>R</sup>   | Novagen  |
| pET15b :: lysM <sub>cse</sub>         | Derivative of pET15b carrying a 549 bp insert coding a LysM <sub>cse</sub> His <sub>6</sub> -tagged protein  | This study   |
| pET15b :: chap                        | Derivative of pET15b carrying a 369 bp insert coding a CHAP His <sub>6</sub> -tagged protein   | This study   |

**Table 2.** Primers used in this study (Supplementary data).

| Name  | Sequence (5' to 3') <sup>a</sup>                                  | Restriction sites |
|---|---|-------------------|
| <b>Construction of <i>S. thermophilus</i> cseΔlysM mutant</b> |   |                   |
| Del_lysME5'   | CCC CCC CCA <b><u>AGC TT</u></b> A CAA GAT CAG TTT GGA AC         | <i>Hind</i> III   |
| Del_lysMI5'   | CCC CCC <b><u>GAA TTC</u></b> ACG TGC TGT CCA GTG A               | <i>Eco</i> RI     |
| Del_lysMI3'   | CCG <b><u>GAA TTC</u></b> GAT GCT ATT GAT CAT GTT AC              | <i>Eco</i> RI     |
| Del_lysME3'   | CCC CCT <b><u>GCA GTC</u></b> GAT ATG TAT TAG GAG TT              | <i>Pst</i> I      |
| <b>Construction of <i>S. thermophilus</i> cseΔchap mutant</b> |   |                   |
| Del_CHAPE5'   | CCC CCA <b><u>AGC TTG</u></b> ATG CTA TTG ATG CTA TTG ATC ATG TTA | <i>Hind</i> III   |
| Del_CHAPI5'   | CCG <b><u>GAA TTC</u></b> AGT TGA TGC ATA TTC GTA T               | <i>Eco</i> RI     |
| Del_CHAPI3'   | CCC CCC <b><u>GAA TTC</u></b> ATT TAT CCA TAA TGA GTA CA          | <i>Eco</i> RI     |
| Del_CHAPE3'   | CCC CCC <b><u>CTG CAG</u></b> CCT GTG GTT GAA G                   | <i>Pst</i> I      |
| <b>Construction of pET15b::lysM<sub>cse</sub></b>             |   |                   |
| surlysM5'   | CCC CCC CCA <b><u>TAT G</u></b> TTT ATC AAA ATC TAA AAC           | <i>Nde</i> I      |
| surlysM3'   | CCC <b><u>CGG ATC C</u></b> TT AAG CTG GAG TAT CTG ATG C          | <i>Bam</i> HI     |
| <b>Construction of pET15b::chap</b>                           |   |                   |
| surchap5'   | CCC CCC CCA <b><u>TAT G</u></b> CT AGC AGC TAC ATA CGA ATA        | <i>Nde</i> I      |
| surchap3'   | CCC <b><u>CGG ATC C</u></b> TT ATG GAT AAA TAT AAT ATA CAG AA     | <i>Bam</i> HI     |

<sup>a</sup> Restriction sites are in boldface and underlined.

**Table 3:** (A) Distribution of muuropeptides and (B) proposed structures and amounts of five muuropeptides generated by digestion of *B. subtilis* cell-wall peptidoglycan with mutanolysin only (a) or double digested with mutanolysin and the CHAP-His<sub>6</sub>-tagged protein (b).

| A)       | % of all peaks <sup>c</sup> |      |
|----------|-----------------------------|------|
|          | (a)                         | (b)  |
| Monomers | 33.7                        | 74.3 |
| Dimers   | 60.4                        | 23.9 |
| Trimers  | 5.9                         | 1.8  |

| B) | Peak <sup>a</sup> | [M+Na] <sup>+</sup><br>observed | Proposed structure <sup>b</sup>      | % of all peaks <sup>c</sup> |      |
|----|-------------------|---------------------------------|--------------------------------------|-----------------------------|------|
|    |                   |                                 |                                      | (a)                         | (b)  |
|    | 1                 | 852.3                           | DS-Tri with one amidation            | 14.3                        | 26.3 |
|    | 2                 | 1815.7                          | DS-Tri-DS-Tetra with one amidation   | 10.1                        | 6.5  |
|    | 3                 | 1814.7                          | DS-Tri -DS-Tetra with two amidations | 19.0                        | 4.7  |
|    | 4                 | 964.4                           | DS-Tetra                             | 0.3                         | 2.4  |
|    | 5                 | 963.4                           | DS-Tetra with one amidation          | 2.2                         | 16.7 |

<sup>a</sup> Peak numbers refer in Fig. 6A. <sup>b</sup> DS-Tri, disaccharide tripeptide (L-Ala-D-γGln-mDAP); DS-Tetra, disaccharide tetrapeptide (L-Ala-D-γGln-mDAP-D-Ala); the letter a indicate an amidation; disaccharide GlcNAc-MurNAc (GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid). <sup>c</sup> Percentage of each peak was calculated as the ratio of the peak area over the sum of areas of all the peaks identified by RP-HPLC of chromatogram. (a) and (b) indicated the chromatograms (a) and (b) observed in Fig. 6A and B.

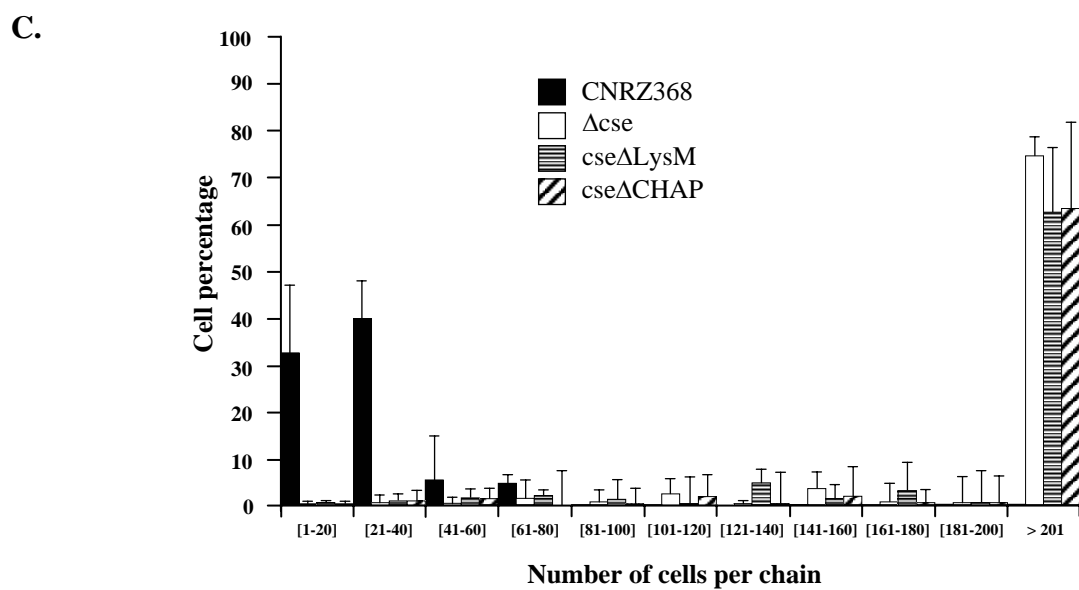
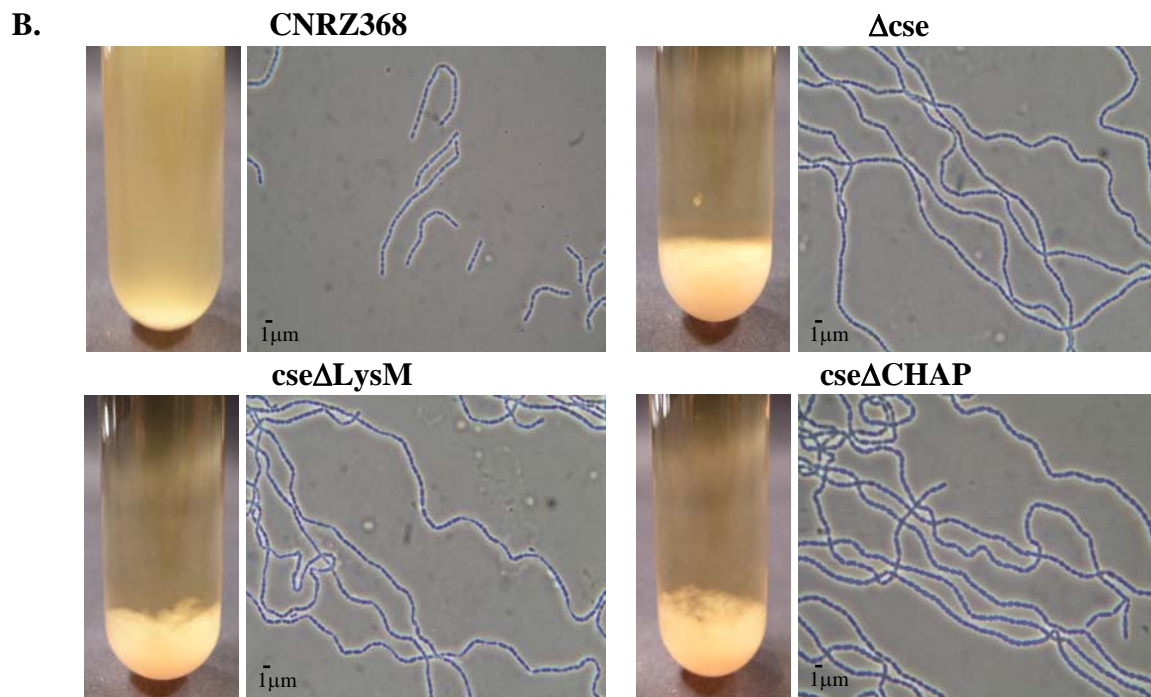
**Table 4:** Proposed structures for the *S. thermophilus* muuropeptides and the digestion products obtained with the CHAP-His<sub>6</sub>-tagged protein.

| Peak <sup>a</sup> | [M+Na] <sup>+</sup> | Proposed structure <sup>b</sup>   |
|-------------------|---------------------|---|
| 6                 | 1061.3              | DS-Tetra –Ala Ala with one amidation  |
| 7                 | 1132.4              | DS- Penta – Ala Ala with one amidation  |
| 8                 | 2081.9              | DS- Tetra – Ala Ala – DS Tetra – Ala Ala with two amidations                        |
| 9                 | 2153.0              | DS- Penta – Ala Ala – DS Tetra – Ala Ala with two amidations                        |
| 10                | 3173.3              | DS- Penta – Ala Ala – DS Tetra – Ala Ala – DS Tetra – Ala Ala with three amidations |

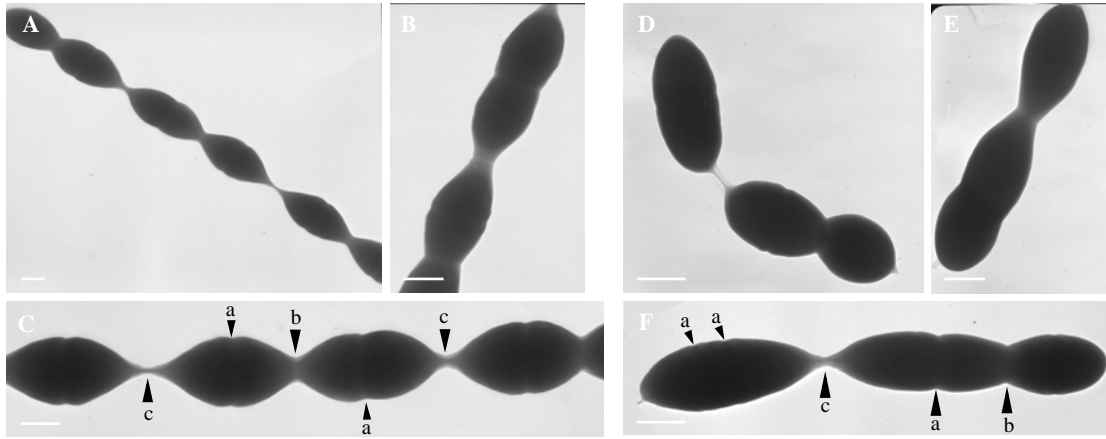
<sup>a</sup> Peak numbers refer to Fig. 6B. <sup>b</sup> DS-Tetra, disaccharide tetrapeptide (L-Ala-D-γGln-L-Lys-D-Ala), DS-Penta, disaccharide pentapeptide (L-Ala-D-γGln-L-Lys-D-Ala-D-Ala).

**A.**

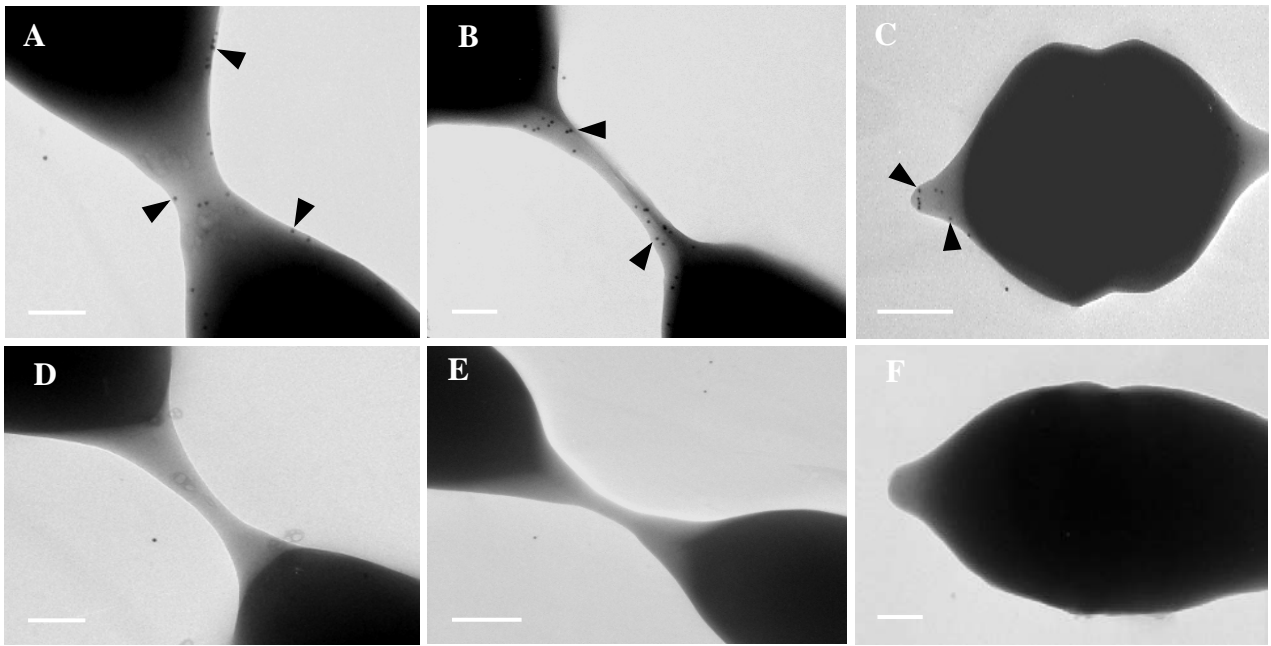
| Name                     | <i>cse</i> modular structure | Gene Size | Position of the deleted sequence |
|--------------------------|------------------------------|-----------|----------------------------------|
| CNRZ368 (Wild-type)      |                              | 1383 pb   | /                                |
| $\Delta cse$             |                              | 24 pb     | 12..1374                         |
| <i>cse</i> $\Delta$ LysM |                              | 924 pb    | 114..573                         |
| <i>cse</i> $\Delta$ CHAP |                              | 1065 pb   | 1053..1374                       |



**Figure 1**



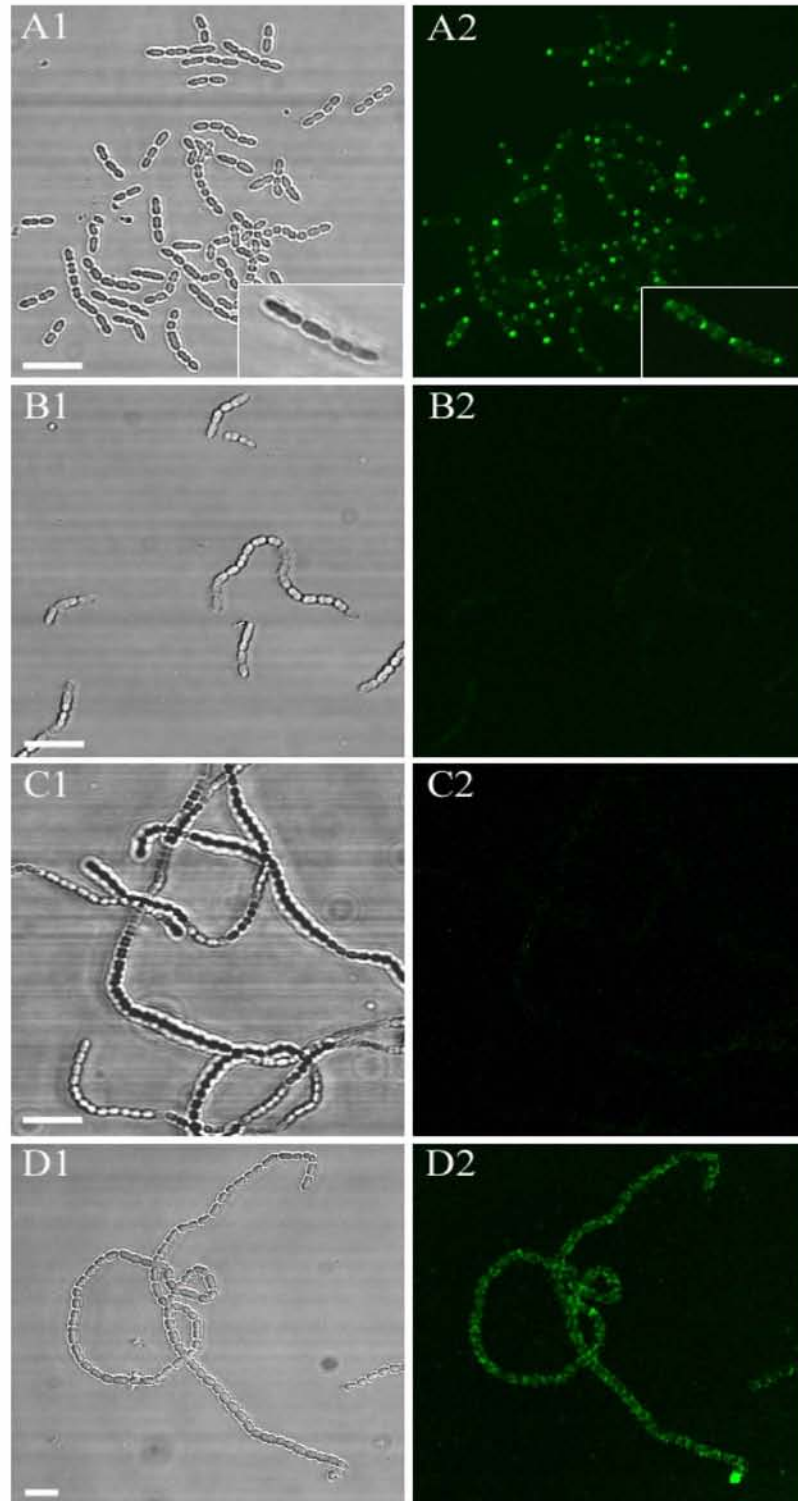
**Figure 2**



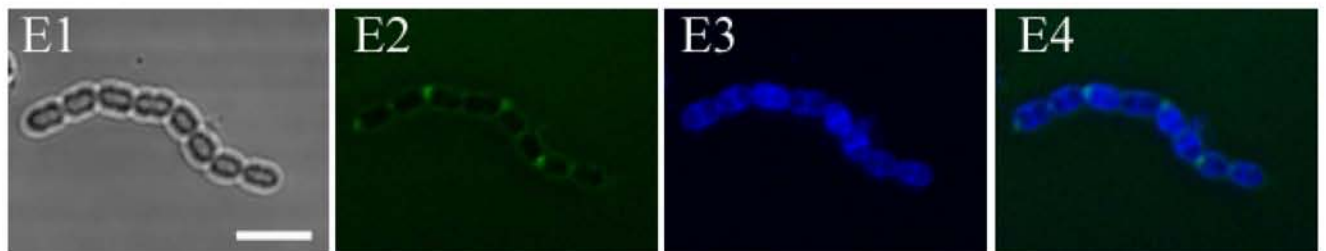
**Figure 3**



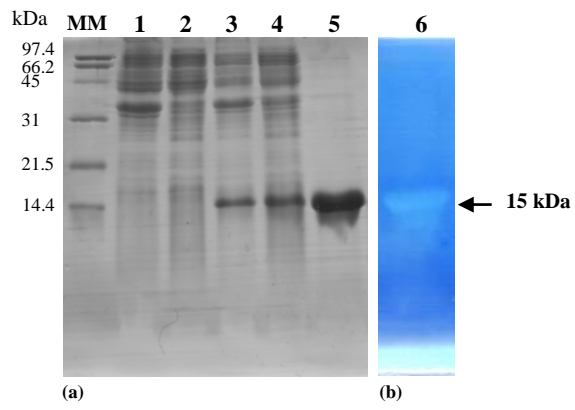
**A.**



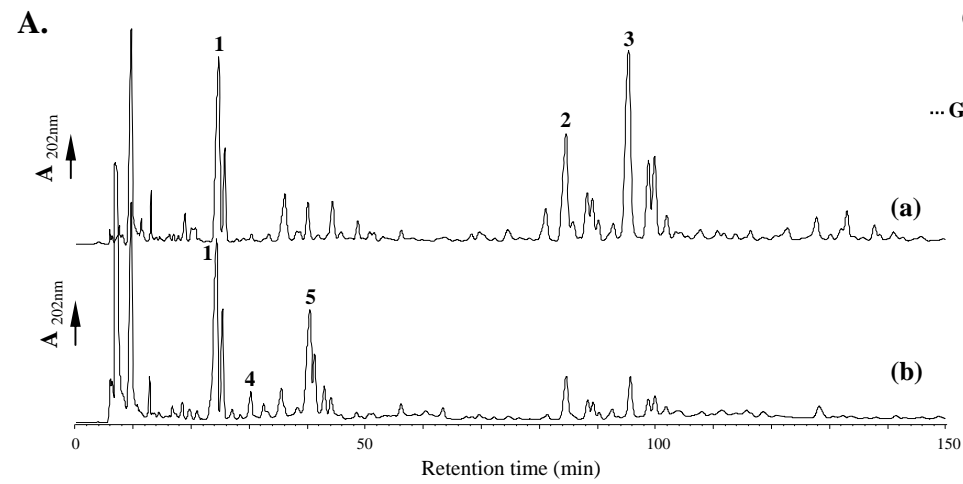
**B.**



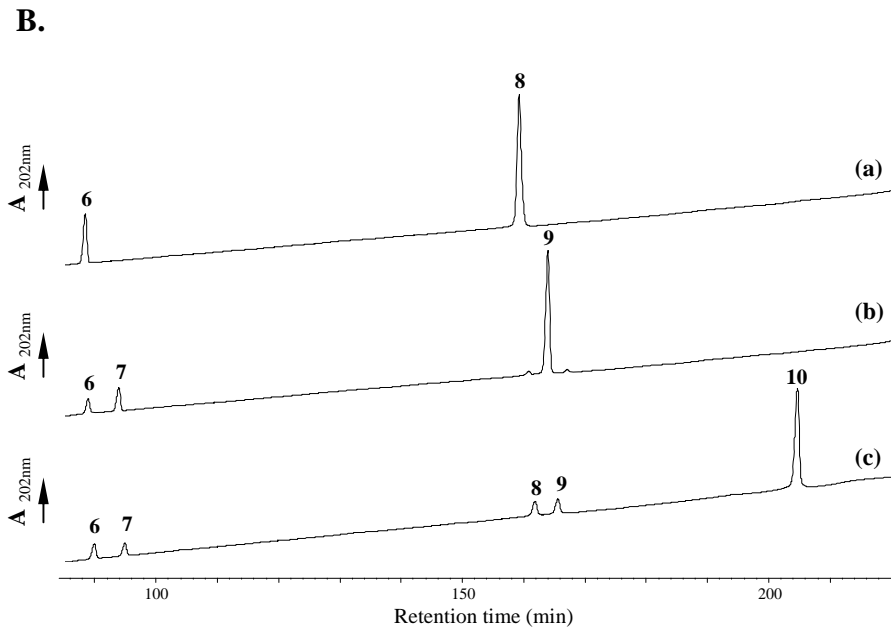
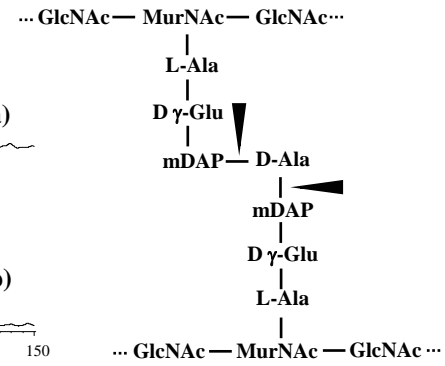
**Figure 4**



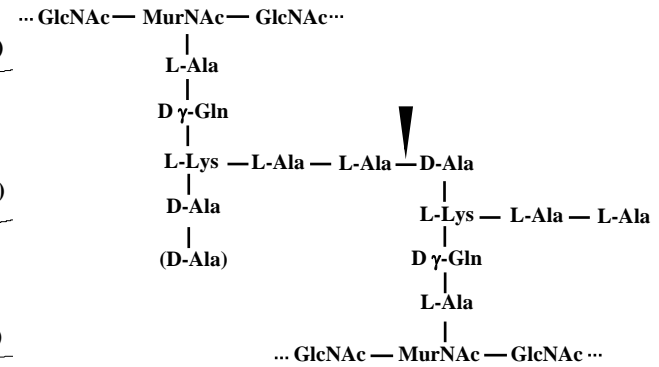
**Figure 5**



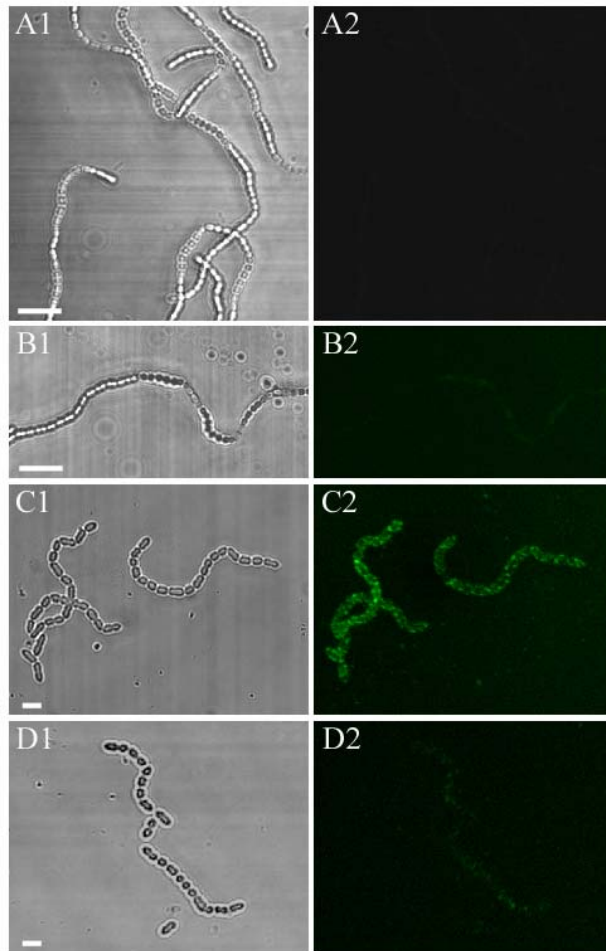
**C1.**



**C2.**

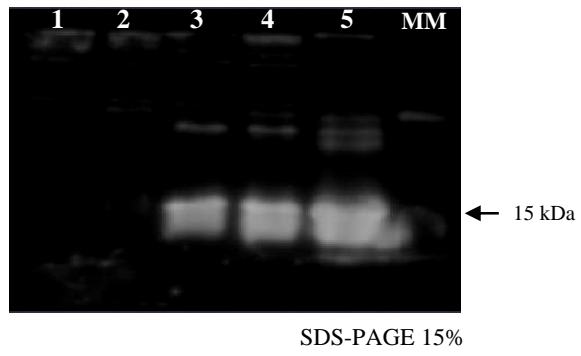


**Figure 6**

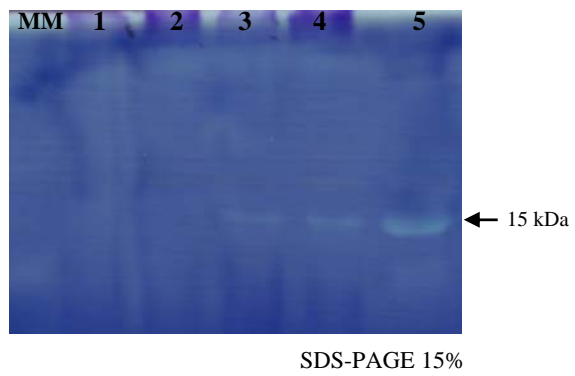


**Figure7 Supplementary data**

**A.**



**B.**



**Figure8 supplementary data**



A black and white photograph of a string of light-colored, oval-shaped beads or pearls. The beads are arranged in a slightly curved line, extending from the bottom left towards the top right. The background is a dark, textured surface, possibly a fabric or a wall with a pebbled texture. The lighting is soft, highlighting the smooth, rounded surfaces of the beads.

*Discussion - Perspectives*





Le peptidoglycane est le composant majeur de la paroi des bactéries à coloration Gram positive. Le maintien de son intégrité est essentiel à la survie cellulaire (Rogers et al., 1980). De nombreuses hydrolases du peptidoglycane sont synthétisées par les bactéries, afin d'assurer d'importants processus cellulaires tels que la croissance, la division bactérienne, la séparation des cellules, le recyclage du peptidoglycane, la lyse cellulaire ou encore la sporulation et la formation de flagelles (Smith et al., 1996).

La diversité des hydrolases du peptidoglycane des bactéries à coloration Gram positive est très grande. Cette diversité se traduit par l'existence d'enzymes attribuées à la quasi-totalité des liaisons peptidiques que l'on peut retrouver au sein de la paroi cellulaire. En fonction des spécificités enzymatiques, les hydrolases du peptidoglycane sont divisées en cinq classes : les *N*-acétylmuramoyl-L-alanine amidases, les *N*-acétylglucosaminidases, les *N*-acétylmuramidases, les endopeptidases et enfin les carboxypeptidases (Ghuysen 1968 ; Shockman et Barrett, 1983 ; Shockman et Höltje 1994) (Figure 44).

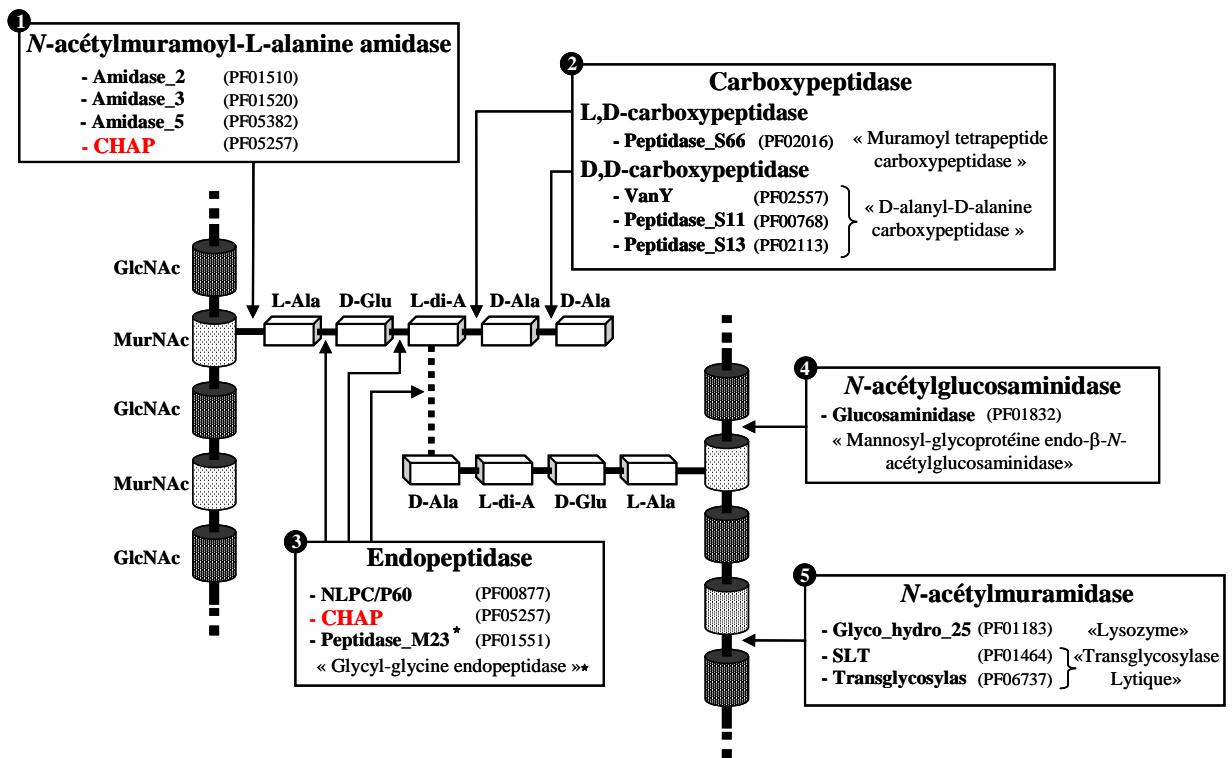


Figure 44 : Diversité des domaines catalytiques assurant la spécificité de clivage des cinq classes d'hydrolases du peptidoglycane (Layec et al., 2008b)

Les hydrolases du peptidoglycane sont classées en cinq catégories selon leur spécificité de clivage dans le peptidoglycane. Les domaines catalytiques associés à chacune des classes sont indiqués, ainsi que leur n° d'accèsion de la banque de données Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) (Finn et al., 2006). Sur cette figure est schématisée le peptidoglycane des bactéries à coloration Gram positive selon Scheilfer et Kandler (1972).

Ces différentes classes d'enzymes sont caractérisées par un ou plusieurs domaines catalytiques distincts (Figure 44). Parmi ces domaines, se trouve le domaine CHAP qui confère à Cse de *S. thermophilus*, son activité d'hydrolase du peptidoglycane et qui fera l'objet du chapitre I de la discussion.

## **I. La « Superfamille des hydrolases du peptidoglycane CHAP » : les domaines CHAP et NLPC/P60.**

Les données bibliographiques indiquent que la « superfamille CHAP » a été découverte récemment par des analyses bioinformatiques de génomes de procaryotes et également d'eucaryotes (Ananthraman et Aravind, 2003 ; Bateman et Rawlings, 2003 ; Ridgen et *al.*, 2003). Chez les bactéries, la présence du domaine CHAP est une caractéristique partagée par les protéines ayant toutes potentiellement, une activité d'hydrolyse du peptidoglycane (Bateman et Rawlings, 2003 ; Ridgen et *al.*, 2003). Toutefois, peu de protéines appartenant à la superfamille CHAP ont été caractérisées biochimiquement. Certaines d'entre elles pourraient être d'ailleurs totalement étrangères à la fonction de dégradation de la paroi cellulaire, puisque l'on retrouve également dans cette superfamille, des lipoprotéines putatives (Ananthraman et Aravind, 2003).

Les hydrolases du peptidoglycane de la superfamille CHAP sont caractérisées par la présence d'au moins deux résidus aminés hautement conservés : une cystéine et une histidine. Dans le modèle proposé par Ananthraman et Aravind (2003), l'hydrolyse de la liaison du peptidoglycane se déroule classiquement par une réaction chimique acide/base. La cystéine serait directement impliquée dans la catalyse en intervenant dans une attaque nucléophile mettant en jeu son groupement thiol, tandis que l'histidine agirait comme un donneur de protons (Ananthraman et Aravind, 2003). La présence d'un résidu aminé polaire additionnel de type asparagine ou encore acide aspartique serait aussi cruciale pour assurer la fonction enzymatique, puisqu'il servirait à l'orientation correcte du résidu catalytique histidine dans son site actif (Ananthraman et Aravind, 2003 ; Ridgen et *al.*, 2003). La mutagenèse ciblée de la cystéine et l'histidine conservées, ou la délétion complète du domaine CHAP, conduit à la perte d'activité des hydrolases du peptidoglycane des phages  $\phi 11$  des staphylocoques (Navarre et *al.*, 1999), B30 de *S. agalactiae* (Donovan et *al.*, 2006b ; Pritchard et *al.*, 2004) ou encore de PlyC des streptocoques (Nelson et *al.*, 2006). Ceci démontre ainsi que le domaine CHAP est un domaine catalytique et que les résidus aminés cystéine et histidine du site actif sont essentiels (Ananthraman et Aravind, 2003 ; Bateman et Rawlings, 2003 ; Ridgen et *al.*, 2003). D'ailleurs, l'une des principales conclusions de ce travail de recherche est que le domaine CHAP est un domaine catalytique qui confère à la protéine Cse de *S. thermophilus*, son activité d'hydrolase du peptidoglycane permettant la séparation des cellules.

Par ailleurs, l'analyse bioinformatique de l'ensemble des protéines de la superfamille CHAP au sein du phylum des *Firmicutes*, a permis d'élargir nos connaissances sur ces protéines. En effet, cette analyse révèle que la superfamille CHAP inclut deux domaines catalytiques distincts : CHAP et NLPC/P60 (Layec et *al.*, 2008a). Ces deux domaines qui présentent des similitudes : présence des résidus aminés conservés cystéine, histidine, et du résidu aminé polaire asparagine ou acide aspartique, ont été considérés au préalable comme identiques (Ananthraman et Aravind, 2003 ; Ridgen et *al.*, 2003). Toutefois, les résultats de notre analyse indiquent que les domaines CHAP et NPLC/P60 présentent des caractéristiques qui leur sont propres et par conséquent qu'il s'agit de deux domaines distincts (Layec et *al.*, 2008a).

# 1. CHAP et NLPC/P60, deux domaines distincts.

L'alignement des séquences protéiques du domaine CHAP, ainsi que celui des séquences protéiques du domaine NLPC/P60 de l'ensemble des protéines analysées au sein du phylum des *Firmicutes* ont permis de différencier ces deux domaines catalytiques (Figure 45).

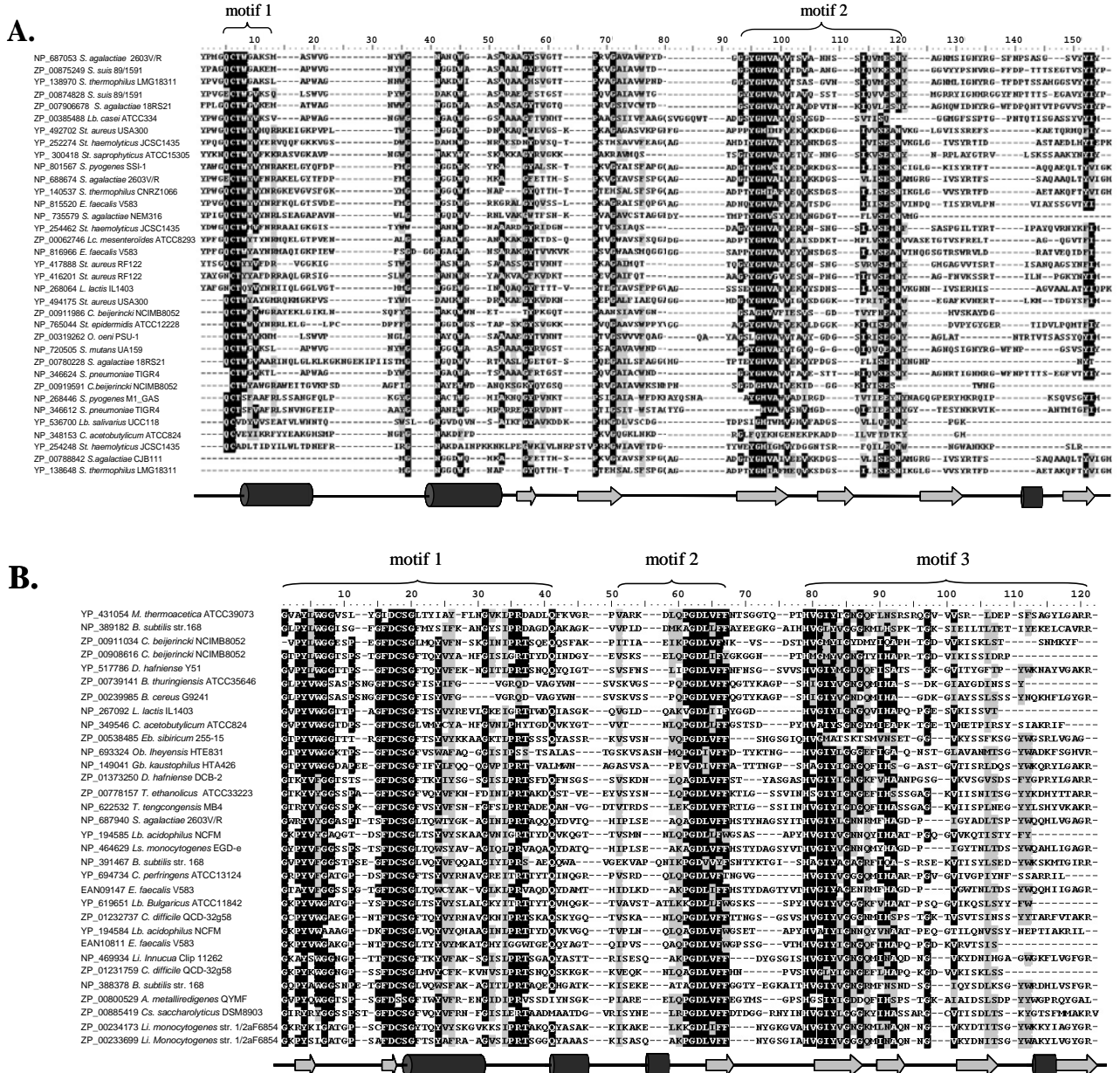


Figure 45 : Mise en évidence des différences entre les domaines CHAP et NLPC/P60 au sein du phylum des *Firmicutes* (Layec et al., 2008a).

A. Alignement de séquences protéiques du domaine CHAP de 35/230 protéines analysées. B. Alignement de séquences protéiques du domaine NLPC/P60 de 32/100 protéines analysées par ClustalW du logiciel BioEdit (Thompson et al., 1994). Les motifs en acides aminés conservés sont QCTWYV(n)<sub>2</sub>R et GnYGHVAnVenV(n)<sub>3</sub>InISEnNY pour le domaine catalytique CHAP et GnPYVWGG(n)<sub>2</sub>P(n)<sub>3</sub>GFDCSG(n)<sub>3</sub>VYV(n)<sub>4</sub>GInLPRT(n)<sub>3</sub>Q, V(n)<sub>x</sub>LQPGDLVFF et HVGIYIGNGQnIHA(n)<sub>4</sub>G(n)<sub>3</sub>V(n)<sub>4</sub>L(n)<sub>3</sub>YW(n)<sub>8</sub>R pour le domaine catalytique NLPC/P60 où la lettre n représente les résidus aminés non conservés et la lettre x correspond au nombre d'acides aminés pouvant aller de 2 à 7 résidus aminés.

Ainsi, le domaine CHAP est caractérisé par la conservation de deux motifs en acides aminés qui sont absents des séquences protéiques du domaine NLPC/P60, tandis que trois motifs propres au domaine NLPC/P60 ont été identifiés (Figure 45). Par ailleurs, le domaine CHAP présente une structure secondaire de type  $\alpha\alpha(\beta)_5\alpha\beta$  qui diffère de celle du domaine NLPC/P60 de type  $\beta\beta(\alpha)_3(\beta)_4\alpha\beta$ .

## **2. La nature des liens peptidiques hydrolysés par la superfamille CHAP est très variée.**

Peu de protéines à domaine NLPC/P60 ont été actuellement caractérisées. Les hydrolases du peptidoglycane NlpC/p60 qui ont été rapportées jusqu'à présent, sont principalement des enzymes de *B. subtilis* qui clivent exclusivement la liaison peptidique entre les résidus aminés acide  $\gamma$ -glutamique et acide *méso*-diaminopimélique de la chaîne latérale du peptidoglycane (Fukushima *et al.*, 2006 et 2008 ; Suzuki et Tahara, 2003 ; Yamaguchi *et al.*, 2004). Récemment, la spécificité de clivage de la protéine YjgB de *L. lactis* présentant un domaine catalytique NLPC/P60 a été identifiée. Cette dernière hydrolyse le lien peptidique de l'acide D- $\gamma$ -glutamyl-L-lysine (Redko *et al.*, 2007), liaison peptidique analogue à celle clivée par les enzymes à domaine NLPC/P60 chez *B. subtilis*.

Concernant les protéines à domaine CHAP, la majorité des hydrolases du peptidoglycane qui ont été à ce jour identifiées, sont des enzymes phagiques. Par ailleurs, il est intéressant de constater que les protéines CHAP phagiques sont dans la plupart des cas, des enzymes bifonctionnelles, c'est-à-dire comportant deux domaines catalytiques différents. C'est par exemple le cas des endopeptidases phagiques  $\phi 11$ ,  $\phi MR11$  et LysWMY des staphylocoques (Navarre *et al.*, 1999 ; Rashel *et al.*, 2007 ; Yokoi *et al.*, 2005) ou encore l'endopeptidase du phage B30 (Donovan *et al.*, 2006a ; Pritchard *et al.*, 2004). Néanmoins, cette caractéristique n'est pas une généralité puisque l'endopeptidase PlyC du phage C1 des streptocoques ne présente qu'un seul domaine catalytique : le domaine CHAP (Nelson *et al.*, 2006). De plus, lorsque le domaine CHAP de ces enzymes phagiques est délété, l'activité de lyse des cellules est fortement diminuée, voire inexistante (Donovan *et al.*, 2006a ; Navarre *et al.*, 1999 ; Pritchard *et al.*, 2004 ; Yokoi *et al.*, 2005) démontrant ainsi l'importance de ce domaine dans la lyse cellulaire.

La spécificité de clivage du domaine CHAP de certaines de ces enzymes a été déterminée, notamment, celui de l'endolysine du phage B30 de *S. agalactiae* qui hydrolyse le lien peptidique D-alanyl-L-alanine du pont interpeptidique présent dans la paroi des streptocoques (Baker *et al.*, 2006 ; Pritchard *et al.*, 2004). L'endopeptidase du phage  $\phi 11$  de *S. aureus* quant à elle, clive la liaison peptidique entre le résidu aminé D-alanine de la chaîne latérale et le premier résidu aminé glycine du pont interpeptidique dont la structure est composée de cinq résidus aminés glycines. Cette structure est d'ailleurs spécifique à la paroi des staphylocoques (Navarre *et al.*, 1999). L'enzyme Skl du phage SK137 de *Streptococcus mitis*, clive la liaison amide entre le groupe D-lactyl de l'acide N-acétylmuramique du polymère disaccharidique et le premier résidu aminé L-alanine de la chaîne latérale du peptidoglycane (Lull *et al.*, 2006). Ainsi, les protéines CHAP codées par les phages présentent soit une activité d'endopeptidase (B30 ;  $\phi 11$  ; LysWMY) ou une activité N-acétylmuramoyl-L-alanine amidase (Skl).

Contrairement aux enzymes phagiques, l'activité enzymatique des protéines CHAP codées par les génomes bactériens a été caractérisée pour seulement deux d'entre-elles : l'enzyme Sle1 de *S. aureus* qui présente une activité *N*-acétylmuramoyl-L-alanine amidase (Kajimura et al., 2005) et l'endopeptidase Cse de *S. thermophilus* qui a été déterminé au cours de ces travaux.

Ainsi, le domaine CHAP a la particularité d'appartenir à deux classes d'enzymes distinctes : les *N*-acétylmuramoyl-L-alanine amidases et les endopeptidases (Kajimura et al., 2005 ; Llull et al., 2006 ; Navarre et al., 1999 ; Pritchard et al., 2004), contrairement au domaine NLPC/P60 qui n'appartient qu'à la classe des endopeptidases (Figure 44). Par ailleurs, si le domaine catalytique NLPC/P60 est spécialisé dans le clivage d'un lien peptidique spécifique dans la chaîne latérale du peptidoglycane, le domaine catalytique CHAP est capable d'hydrolyser de multiples liaisons du peptidoglycane (Baker et al., 2006 ; Kajimura et al., 2005 ; Llull et al., 2006 ; Navarre et al., 1999). D'ailleurs, le domaine CHAP de la protéine Cse de *S. thermophilus* est capable d'hydrolyser de multiples liaisons du peptidoglycane telles que le lien peptidique entre le résidu aminé D-alanine de la chaîne latérale et le résidu aminé L-alanine du pont interpeptidique de la paroi de *S. thermophilus* ou encore le lien peptidique entre l'acide *méso*-diaminopimélique et le résidu aminé D-alanine de la paroi de *B. subtilis*.

### **3. La superfamille CHAP est associée à divers processus cellulaires.**

La plupart des protéines de la superfamille CHAP identifiées à ce jour, sont impliquées dans les processus de lyse cellulaire, de division cellulaire ou encore de séparation cellulaire.

Au sein du phylum des *Firmicutes*, toutes les protéines phagiques CHAP ont pour rôle de lyser les cellules bactériennes. Ainsi, les enzymes phagiques telles que l'endopeptidase du phage B30 de *S. agalactiae* (Donovan et al., 2006a ; Pritchard et al., 2004), les endopeptidases des phages  $\phi$ 11 et  $\phi$ 12 de *S. aureus* (Navarre et al., 1999 ; Sass et al., 2007), l'enzyme LysWMY du phage  $\phi$ WMY de *Staphylococcus warneri* (Yokoi et al., 2005) ou encore l'enzyme de Skl du phage SK137 de *S. mitis* (Llull et al., 2006) s'inscrivent dans le processus de lyse cellulaire. Concernant, les enzymes bactériennes, seule la protéine AcMB de *L. lactis* a été identifiée comme étant impliquée dans la lyse des cellules en phase stationnaire (Huard et al., 2003). Cette endo- $\beta$ -glucosaminidase présente deux domaines catalytiques : « Glucosaminidase » et « CHAP ». Seule l'activité du domaine glucosaminidase a été démontré, mais il n'est pas exclu que le domaine CHAP présente lui aussi une activité catalytique nécessaire à la fonction d'AcMB.

A ce jour, les données bibliographiques indiquent que les protéines bactériennes de la superfamille CHAP interviennent dans les processus de division et de séparation cellulaire. La seule protéine actuellement identifiée dans le processus de division cellulaire, est la protéine PcsB/GbpB à domaine CHAP des streptocoques impliquée dans l'orientation et le positionnement des plans de division (Chia et al., 2001 ; Mattos-Graner et al., 2006 ; Ng et al., 2004 ; Reinscheid et al., 2001). Toutes les autres sont impliquées dans le processus de séparation cellulaire. C'est le cas des enzymes LytE, LytF et CwlS (YojL) de *B. subtilis* (Fukushima et al., 2006 ; Yamamoto et al., 2003), de la protéine p60 de *Listeria monocytogenes* (Machata et al., 2005) ou encore de la protéine Sle1 de

*S. aureus* (Kajimura et al., 2005). C'est également le cas de Cse, une protéine CHAP qui agit spécifiquement dans la séparation des cellules de *S. thermophilus* (Borges et al., 2005).

Le fait que les protéines CHAP interviennent dans des processus cellulaires distincts, lyse et séparation cellulaire, pourrait suggérer que la séquence protéique du domaine CHAP des enzymes de lyse cellulaire puisse présenter des caractéristiques qui lui sont propres et bien distinctes de celles de la séquence protéique du domaine CHAP des enzymes de division et de séparation des cellules. Si tel est le cas, deux classes distinctes pourraient être établies : la classe des enzymes de lyse cellulaire et la classe des enzymes de division et séparation cellulaire. Afin d'étayer cette hypothèse, les séquences protéiques des domaines CHAP ont été comparées entre elles. Cependant, cette analyse *in silico* a révélé une importante divergence entre les séquences protéiques du domaine CHAP des protéines de lyse cellulaire allant de 60 à 80%, mais aussi entre les protéines de division et de séparation cellulaire. Seuls les domaines CHAP des protéines Cse et PcsB de *S. thermophilus* sont homologues témoignant de leur étroite apparentée (Borges et al., 2005). Par ailleurs, quelle que soit la fonction assurée par les protéines CHAP, les acides aminés conservés du domaine CHAP sont identiques et/ou similaires, ne permettant pas ainsi de distinguer les enzymes de lyse cellulaire de celles des enzymes de division et de séparation cellulaire.

Par conséquent, l'important polymorphisme du domaine CHAP n'est pas étroitement corrélé à la diversité des processus cellulaires dans lequel ce domaine est impliqué.

#### **4. Distribution hétérogène des domaines CHAP et NlpC/P60 au sein des génomes des *Firmicutes*.**

Au cours de l'analyse *in silico* des protéines de la superfamille CHAP, nous avons constaté que la présence des gènes codant des protéines CHAP diffère de celle des gènes codant des protéines NlpC/p60 au sein des génomes du phylum des *Firmicutes* (Layec et al., 2008a). Ainsi, la majorité des protéines CHAP est codée par les génomes de la famille des *Lactobacillales* (56,5%) et par les génomes de la famille des *Bacillales* (42%). Quant aux protéines NlpC/p60, contrairement aux protéines CHAP, elles sont codées par seulement 20% des génomes de la famille des *Lactobacillales*. La majorité de ces protéines sont principalement codées par les génomes de la famille des *Bacillales* (37%) et de la famille des *Clostridia* (34%). Ainsi, la distribution des protéines CHAP et NlpC/p60 est très différente au sein du phylum des *Firmicutes*. Par ailleurs, au cours de cette recherche bioinformatique, l'absence de gènes codant des protéines NlpC/p60 dans les génomes du genre *Staphylococcus* a été observée. Ces génomes codent exclusivement des protéines CHAP. De même, les génomes des genres *Bacillus* et *Listeria*, codant principalement des protéines NlpC/p60, sont dépourvus de gènes codant des protéines CHAP (Layec et al., 2008a). Dans le cas de la famille des *Lactobacillales*, les protéines CHAP sont majoritairement codées par les génomes du genre *Streptococcus* (~ 88,5%). Il est également intéressant de noter que la majorité des génomes séquencés du genre *Streptococcus* sont dépourvus de gènes codant des protéines NlpC/p60. Cependant, un gène codant une protéine NlpC/p60 a été révélé pour sept génomes séquencés, correspondant à une lipoprotéine putative retrouvée sur un élément transposable suggérant une acquisition du gène de celle-ci par transfert horizontal. La présence du gène codant cette lipoprotéine dans le génome de la souche *S. pyogenes* MGAS10750 et son absence pour les 12 autres génomes séquencés de *S. pyogenes*, serait un argument en faveur de cette hypothèse.

La distribution hétérogène des domaines catalytiques CHAP et NLPC/P60 au sein du phylum de *Firmicutes*, pourrait s'expliquer par le fait que la paroi cellulaire des genres *Bacillus* et *Listeria* est pourvue de l'acide *méso*-diaminopimélique (Schleifer et Kandler, 1972). En effet, les hydrolases du peptidoglycane à domaine NLPC/P60 qui ont été rapportées jusqu'à présent, hydrolysent seulement la liaison peptidique entre les résidus aminés acide  $\gamma$ -glutamique et acide *méso*-diaminopimélique de la chaîne latérale du peptidoglycane (Fukushima et al., 2006 ; Suzuki et Tahara, 2003 ; Yamaguchi et al., 2004).

En conclusion, au sein du phylum des *Firmicutes*, la superfamille CHAP est constituée d'au moins deux domaines catalytiques distincts : CHAP et NPLC/P60. Les protéines CHAP et NlpC/p60 qui ont été caractérisées jusqu'à ce jour, sont toutes des hydrolases du peptidoglycane impliquées dans divers processus cellulaires. Ces enzymes, et tout particulièrement celles qui ont un domaine CHAP, sont capables d'hydrolyser de multiples liaisons du peptidoglycane. Afin d'élargir nos connaissances, d'un point de vue moléculaire, sur le mode d'action du domaine catalytique CHAP, une des perspectives de ce travail serait d'élucider la structure tridimensionnelle de celui-ci par la cristallisation du domaine CHAP de la protéine Cse et d'étudier les interactions moléculaires entre le domaine CHAP et divers substrats. De même, une étude analogue sur le domaine NLPC/P60 pourrait être envisagée. De plus, ces analyses pourraient contribuer à déterminer pourquoi une telle diversité de spécificité de clivage du lien peptidique est observée pour le domaine catalytique CHAP.

Par ailleurs, les genres bactériens possédant le domaine CHAP et/ou NLPC/P60 ne constituent pas de groupes homogènes dans l'arbre phylogénétique des *Firmicutes*. Ainsi, quant à l'origine de la distribution hétérogène de ces deux domaines, CHAP et NLPC/P60, aucune hypothèse phylogénétique évidente ne peut être formulée.

## **II. La séparation cellulaire, un processus postérieur ou ultime de la division cellulaire ?**

La division cellulaire est un processus essentiel consistant à transmettre le patrimoine génétique d'une bactérie à ses descendants. Ce processus complexe aboutit la formation de deux cellules filles identiques à partir d'une cellule mère.

Au cours de la division cellulaire, les notions de « scission » et de « séparation » cellulaire sont souvent confondues, dans la mesure où la division cellulaire conduit à la disjonction de cellules filles individuelles. Dans ce travail, ces deux notions ont été différenciées. La scission cellulaire est le mécanisme aboutissant à la formation de deux cellules filles à partir de la cellule mère grâce à la mise en place d'un septum, tandis que la séparation cellulaire est le mécanisme permettant de disjoindre les deux cellules filles en cellules individuelles par clivage du peptidoglycane qui les relie. Toutefois, la séparation cellulaire est étroitement liée à la scission cellulaire, puisque celle-ci ne peut avoir lieu que si la septation de la cellule mère en deux cellules filles est aboutie.

### **1. La division cellulaire chez les streptocoques.**

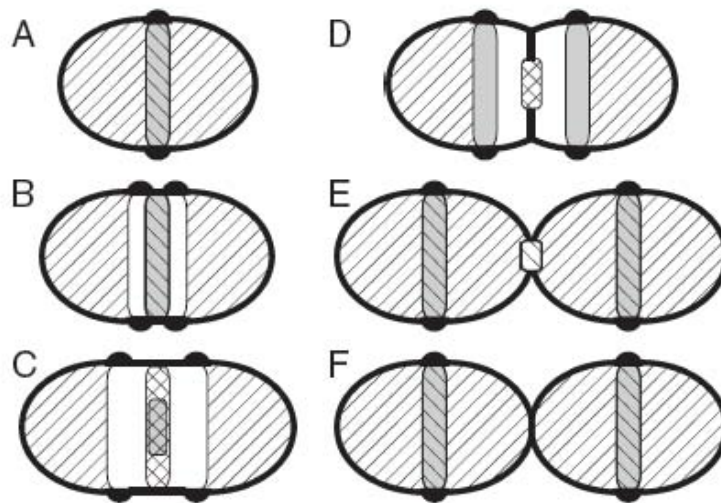
La division cellulaire est définie comme un processus complexe dans lequel quatre grands mécanismes ont lieu : (i) la réplication de l'ADN ; (ii) la ségrégation chromosomique, (iii) la scission de la cellule mère en deux cellules filles non disjointes et (iv) la séparation des cellules filles en cellules individuelles. Seul ces deux derniers points seront abordés dans ce manuscrit.

Au cours de la division cellulaire, un septum de division est formé au futur site de séparation. Ce processus nécessite le recrutement de diverses protéines de la division pour promouvoir la formation du septum, et tout particulièrement la protéine FtsZ, une protéine de division indispensable à l'initiation de la septation (pour revue Rothfield et al., 1999 ; Errington et al. 2003 ; Buddelmeijer et Beckwith, 2002). Les protéines de division interviennent plus précisément, dans l'invagination de la membrane plasmique, dans la synthèse du peptidoglycane permettant l'accroissement de la taille cellulaire et dans la synthèse du peptidoglycane septal permettant la scission de la cellule mère en deux cellules filles. Par ailleurs, une ou plusieurs protéines de division joueraient un rôle dans la coordination de la formation du septum avec d'autres événements du cycle cellulaire incluant la réplication chromosomique et la partition des deux chromosomes. Et au final, la division cellulaire s'achèvera par la séparation des deux nouvelles cellules filles en cellules individuelles.

Chez *S. pneumoniae*, neuf protéines de division sont essentielles : FtsZ, FtsA, FtsK, DivIB/FtsQ, FtsL, FtsW, DivIC/FtsB, PBP3/FtsI et DivIVA. Pendant que la protéine FtsK est impliquée dans la ségrégation chromosomique (Aussel et al., 2002), la protéine FtsZ polymérise une structure en anneau dit « anneau de constriction-Z » au site de



septation nécessaire à l'invagination de la membrane plasmique et qui va permettre d'initier la scission (Figure 46).



**Figure 46 : Illustration schématique de la mise en place du site de séparation des cellules lors de la division cellulaire de *S. pneumoniae* (Noirclerc-Savoie et al., 2005).**

Chez *S. pneumoniae*, les protéines de division FtsZ, DivIC, FtsW et toutes les PBP de hauts poids moléculaires sont localisées sur le plan équatorial de la cellule mère (A). L'anneau équatorial est dupliqué (B) et ces 2 nouveaux anneaux correspondront aux futurs plans équatoriaux des 2 cellules filles (C). L'éloignement de ces 2 anneaux résulte de la synthèse longitudinale du peptidoglycane par les PBP non-septal. Au niveau du plan équatorial de la cellule mère, futur site de séparation, la membrane plasmique s'invagine (D), les protéines de division DivIB et FtsL vont être recrutées (E), de même que les PBP septales pour la synthèse du peptidoglycane permettant la scission de la cellule mère en 2 cellules filles (F).

La protéine FtsZ est aidée de la protéine FtsA qui assure l'interaction de cette dernière avec la membrane plasmique (Lara et al., 2005). Les protéines DivIB, FtsL et DivIC dont le mode d'action reste inconnu, sont également recrutées au niveau du plan équatorial, sur le futur site de séparation (Noirclerc-Savoie et al., 2005 ; Le Gouëllec et al., 2008). Il en est de même pour les protéines FtsW et PBP3 qui participent toutes deux à la biosynthèse de la paroi. En effet, la protéine FtsW est nécessaire à la translocation des précurseurs du peptidoglycane au travers de la membrane plasmique, précurseurs qui sont ensuite incorporés dans le peptidoglycane par l'intermédiaire des PBP (Gérard et al., 2002). La protéine PBP3 assure quant à elle, la synthèse du peptidoglycane septal permettant la scission de la cellule mère en deux cellules filles (Morlot et al., 2003). La synthèse du peptidoglycane non-septal et longitudinal est réalisée notamment par PBP2x, PBP1a, PBP2b et RodA (Noirclerc-Savoie et al., 2003 ; Ishino et al., 1986). Récemment, la protéine DivIVA a été identifiée comme étant une protéine de division. DivIVA est recrutée dans une étape tardive de la scission, puisque celle-ci est impliquée dans la formation et la maturation des pôles cellulaires avant séparation des cellules (Fadda et al., 2007). Par conséquent, la participation d'un grand nombre de protéines est nécessaire pour la mise en place du site de séparation des cellules lors de la division cellulaire chez *S. pneumoniae*. Cependant, à l'heure actuelle, les données biochimiques et structurales sur ces molécules et sur leurs interactions sont peu connues.

Les protéines de la division cellulaire sont conservées dans la plupart des espèces bactériennes. L'ensemble de ces protéines de division est notamment retrouvé dans le génome de *S. thermophilus*. De même que l'implication des protéines RodA et PBP2b dans la synthèse longitudinale du peptidoglycane a également été étudiée chez *S. thermophilus* (Thibessard et al.,

2002b), suggérant ainsi que lors de la division cellulaire, la scission des cellules de cette espèce pourrait être semblable à celle de *S. pneumoniae*.

Ainsi, la scission de la cellule mère en deux cellules filles identiques résulte de la formation d'un septum entre celles-ci. Si la plupart des étapes de la scission sont assez bien décrites, notamment chez *S. pneumoniae* (Morlot et al., 2004 ; Noirclerc-Savoie et al., 2005), à l'heure actuelle, les mécanismes de régulation dans la séparation des deux cellules filles en cellules individuelles lors de la division cellulaire sont peu connus.

Le fait que l'aspect morphologique d'une bactérie résulte de son patrimoine génétique et que la séparation cellulaire soit un phénomène ubiquitaire, quel que soit leur mode de croissance, sous forme unicellulaire isolée, associée en paire ou en chaîne, ou encore sous forme de filament mycélien suggère que la séparation cellulaire soit un mécanisme actif contrôlé par le génome, plutôt qu'un mécanisme passif influé par les conditions physico-chimiques. D'ailleurs, l'identification d'hydrolases du peptidoglycane impliquées dans le processus de séparation des cellules qui ont été rapportées dans la littérature, mais aussi au cours de ce travail démontre bien qu'il s'agit d'un processus actif contrôlé génétiquement. Il reste néanmoins à savoir si l'expression des gènes impliqués dans l'ultime étape de la division cellulaire est inductible ou constitutive.

## **2. La séparation cellulaire implique l'action d'hydrolases du peptidoglycane très diverses.**

Le fait qu'un phénotype « longue chaîne de cellules » est observé pour le mutant *S. thermophilus* délété du gène *cse*, nous a amené à considérer le processus de séparation cellulaire comme un processus actif, génétiquement contrôlé.

### **2.1 La séparation des cellules, un processus actif dépendant d'hydrolases du peptidoglycane.**

Dans la littérature, l'inactivation de gènes conduisant à un phénotype « longue chaîne de cellules », sans affecter la morphologie des cellules a également été rapportée. Ainsi, *acmA* de *L. lactis* (Buist et al., 1995), *lytA* et *lytB* de *S. pneumoniae* (De Las Rivas et al., 2002 ; Garcia et al., 1999a ; Sanchez-Puelles et al. 1986), *aml* de *S. mutans* (Yoshimura et al., 2006), *lytE*, *lytF* et *cwIS* de *B. subtilis* (Fukushima et al., 2006 ; Yamamoto et al., 2003), *murA* et *p60* de *Listeria monocytogenes* (Carroll et al., 2003 ; Machata et al., 2005), *sleI* et *atl* de *S. aureus* (Kajimura et al., 2005 ; Yamada et al., 1996) ou encore *atlA* d'*Enterococcus faecalis* (Eckert et al., 2006 ; Mesnage et al., 2008) par exemple, sont tous des gènes dont l'inactivation a conduit à une abolition de la séparation des cellules filles et par conséquent, sont tous impliqués dans le mécanisme de séparation cellulaire. Tout comme le gène *cse* de *S. thermophilus*, ces gènes codent des hydrolases du peptidoglycane.

Les données bibliographiques indiquent que l'intervention d'une ou plusieurs protéines peut être nécessaire pour assurer le processus de séparation cellulaire. Ainsi, chez *E. coli*, AmiC, ayant une activité *N*-acétylmuramoyl-L-alanine amidase a été identifiée comme étant l'enzyme clé de clivage du lien peptidique permettant la séparation cellulaire (Heidrich et al., 2001). Néanmoins, son

action peut être accrue par l'intervention d'autres hydrolases du peptidoglycane, notamment les *N*-acétylmuramoyl-L-alanine amidases AmiA et AmiB, qui ont été identifiées comme étant des enzymes mineures de la séparation des cellules d'*E. coli* (Bernhardt et al., 2003 ; Heidrich et al., 2001 ; Heidrich et al., 2002 ; Höltje et al., 1998 ; Priyadarshini et al., 2006). Chez *B. subtilis*, les enzymes majeures de la séparation des cellules sont LytE et LytF (Ishikawa et al., 1998 ; Margot et al., 1998 ; Margot et al., 1999 ; Oshini et al., 1999). Récemment, une troisième enzyme a été identifiée chez *B. subtilis* : CwlS (YojL) (Fukushima et al., 2006). La construction du double mutant *lytElytF* et du triple mutant *lytElytFCwlS* s'est traduit par la formation de très longues chaînes de cellules, suggérant l'action simultanée des endopeptidases LytE, LytF et CwlS dans la séparation des cellules de *B. subtilis* (Fukushima et al., 2006 ; Yamamoto et al., 2003). Chez *S. pneumoniae*, la protéine majeure du processus de séparation des cellules est LytB, une endo- $\beta$ -glucosaminidase (De Las Rivas et al., 2002 ; Garcia et al., 1999a) présentant un phénotype « longue chaîne de cellules » lorsque le gène *lytB* est inactivé. Ce phénotype est accru chez le double mutant *lytAlytB*, suggérant que la protéine LytA, une *N*-acétylmuramoyl-L-alanine amidase (Sanchez-Puelles et al. 1986) pourrait elle aussi jouer un rôle, bien que mineur, dans la séparation des cellules de *S. pneumoniae* (Garcia et al., 1999b).

Dans le cas de *S. thermophilus*, l'originalité du gène *cse* est que sa simple délétion se traduit par des chaînes de cellules extrêmement longues pouvant contenir plus de 3000 cellules, indiquant que la protéine Cse est l'enzyme clé de la séparation des cellules de *S. thermophilus* (Borges et al., 2005). Afin d'étayer cette hypothèse, la recherche *in silico* d'enzymes mineures de séparation cellulaire dans les génomes séquencés de *S. thermophilus* a été réalisée, mais aucun candidat possible n'a été trouvé. Ainsi, chez *S. thermophilus*, la séparation cellulaire semble être assurée uniquement par Cse.

Ainsi, l'ensemble de ces données a permis d'identifier la nature du contrôle génétique du processus de séparation cellulaire qui est médié par l'action d'hydrolases du peptidoglycane nécessaires à la rupture des liens du peptidoglycane reliant les deux cellules filles. Toutefois, la modulation de l'expression des gènes codant ces enzymes au cours de la division cellulaire à savoir s'il s'agit d'une expression constitutive ou inductible reste à déterminer.

## **2.2 Diversité des enzymes de séparation cellulaire.**

Bien que toutes les enzymes de séparation cellulaire soient des hydrolases du peptidoglycane, la nature et la spécificité de clivage de ces enzymes s'avèrent très diverses (Figure 47). Ainsi, la spécificité enzymatique des enzymes assurant la libération des cellules filles intactes serait variable d'une espèce à l'autre.

Les enzymes de séparation cellulaire appartiennent à quatre des cinq classes d'hydrolases du peptidoglycane (Figure 47). Elles peuvent être des *N*-acétylmuramoyl-L-alanine amidases. Tel est le cas des protéines Sle1 de *S. aureus* et LytA de *S. pneumoniae* (Kajimura et al., 2005 ; Sanchez-Puelles et al. 1986). Elles peuvent être également des *N*-acétylglucosaminidases. C'est par exemple le cas, des protéines AcmA de *L. lactis* (Buist et al., 1995), AtlA d'*E. faecalis* (Eckert et al., 2006), LytB de *S. pneumoniae* (De Las Rivas et al., 2002 ; Garcia et al., 1999a) et MurA de *L. monocytogenes* (Carroll et al., 2003). Mais aussi, appartenir à la classe des *N*-acétylmuramidases telle que Aml de *Streptococcus mutans* (Yoshimura et al., 2006) ou encore à celle des endopeptidases.

● **N-acétylmuramoyl-L-alanine amidase**

*Staphylococcus aureus*



● **N-acétylglucosaminidase**

*Lactococcus lactis*



*Listeria monocytogenes*



*Enterococcus faecalis*



*Streptococcus pneumoniae*



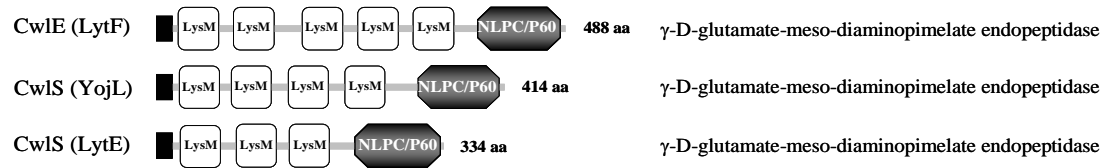
● **N-acétylmuramidase**

*Streptococcus mutans*



● **Endopeptidase**

*Bacillus subtilis*



*Listeria monocytogenes*



*Streptococcus thermophilus*



Figure 47 : Diversité des hydrolases du peptidoglycane de séparation de cellules au sein des Firmicutes (d'après Layec et al., 2008b).

Les différentes hydrolases du peptidoglycane sont classées en fonction de leur spécificité de clivage du peptidoglycane. Seules les enzymes majeures de séparation cellulaires sont illustrées. L'astérisque fait remarquer que l'enzyme Atl de *S. aureus* est une enzyme bifonctionnelle ayant à la fois une activité de N-acétylmuramoyl-L-alanine amidase et de N-acétylglucosaminidase. La spécificité de clivage est précisée pour les endopeptidases. L'architecture modulaire des différentes enzymes est représentée. De même, le nom des domaines catalytiques et des domaines de liaison est indiqué. Le rectangle noir figuré sur les schémas des protéines correspond au peptide signal.

C'est par exemple le cas, des protéines LytE, LytF, CwlS (YojL) de *B. subtilis* (Ishikawa et al., 1998 ; Margot et al., 1998 ; Margot et al., 1999 ; Oshini et al., 1999 ; Fukushima et al., 2006), p60 de *L. monocytogenes* (Machata et al., 2005) et Cse de *S. thermophilus*.

A ce jour, aucune des enzymes de séparation cellulaire n'appartient à la classe des carboxypeptidases. Cette classe comprend uniquement les enzymes impliquées dans les étapes d'assemblage du peptidoglycane lors de la croissance et la division cellulaire (Morlot et al., 2004 ; Schuster et al., 1990).

Certaines enzymes de séparation cellulaire présentent plusieurs domaines catalytiques qui peuvent appartenir à des classes d'hydrolases du peptidoglycane différentes. C'est le cas d'Atl de *S. aureus* qui possède à la fois les activités *N*-acétylmuramoyl-L-alanine amidase et endo- $\beta$ -glucosaminidase (Yamada et al., 1996).

### **2.3 Les enzymes de séparation cellulaires seraient spécifiques d'espèces.**

L'analyse bibliographique réalisée au sein du phylum des *Firmicutes* a révélé que les enzymes de séparation cellulaire identifiées jusqu'alors, sont spécifiques d'espèces. Ainsi, la protéine Cse n'est codée que par le génome de *S. thermophilus*. De même que LytB est spécifique de *S. pneumoniae* et Aml de *S. mutans*. Par conséquent, il semblerait qu'au cours de l'évolution, une enzyme de séparation cellulaire aurait été sélectionnée pour chacune des espèces bactériennes. Afin d'étayer cette hypothèse, la recherche d'enzymes de séparation cellulaire potentielles dans les génomes séquencés des espèces bactériennes appartenant au genre *Streptococcus* a été réalisée. Au moins un candidat par espèce a été retenu (Layec et al., 2008b). De même que LytB, Aml et Cse, les candidats retenus sont spécifiques d'espèces. Toutefois, la fonction biologique exacte de ces candidats reste à déterminer. La recherche des acteurs de séparation cellulaire pourrait s'étendre chez *S. salivarius* et *S. vestibularis*, deux espèces phylogénétiquement proches de *S. thermophilus* (Kawamura et al., 1995 ; Poyart et al., 1998).

Ainsi, contrairement aux protéines de division cellulaire qui sont très conservées, les enzymes de séparation cellulaire sont spécifiques d'espèces. Néanmoins, quelle peut être la nature des pressions de sélection qui s'exerce sur la spécificité de ces enzymes ?

### **2.4 La séparation cellulaire, un processus postérieur à la division cellulaire ?**

Il est communément admis que la séparation cellulaire est l'ultime processus de la division cellulaire. Néanmoins, les données bibliographiques ainsi que ce travail nous amènent à considérer que la séparation cellulaire pourrait être un mécanisme à part entière, postérieur à la division cellulaire. Un argument en faveur de cette hypothèse est que contrairement aux protéines de division cellulaire, l'inactivation des protéines de séparation cellulaire n'affecte pas la morphologie des cellules et seule la formation de cellules en chaîne est observée. Par conséquent, son inhibition n'empêche pas les cellules de se diviser mais seulement de se disjoindre. Par ailleurs, les enzymes de séparation cellulaire sont très diverses et spécifiques d'espèces, contrairement aux protéines de division cellulaire.

### **3. La séparation cellulaire nécessite une localisation spécifique des hydrolases du peptidoglycane.**

La liaison peptidique entre le résidu aminé D-alanine de la chaîne latérale du peptidoglycane et le résidu L-alanine du pont interpeptidique, clivée par Cse est omniprésente dans la paroi cellulaire de *S. thermophilus* (Schleifer et Kandler, 1972). Ainsi, se pose la question de savoir comment les cellules filles sont-elles séparées les unes des autres sans être lysées ?

#### **3.1 Les enzymes de séparation cellulaires sont localisées dans des régions subcellulaires spécifiques : aux pôles et/ou aux septa des cellules.**

Le processus de séparation cellulaire implique un clivage rigoureux et précis du peptidoglycane afin de préserver son intégrité et éviter la lyse cellulaire. Ainsi, l'hypothèse selon laquelle les enzymes de séparation cellulaire n'induisent pas la lyse des cellules filles au cours de leur séparation pourrait s'expliquer, tout au moins en partie, par leur localisation ciblée. Un argument en faveur de cette hypothèse, sont les résultats des études de la localisation d'enzymes de séparation cellulaire de bactéries à coloration Gram positive rapportés dans la littérature. Mais aussi, l'étude de la localisation *in situ* de Cse à la surface des cellules de *S. thermophilus* qui a été réalisée au cours de ce travail.

Ainsi, par exemple, la protéine LytB de *S. pneumoniae* (De Las Rivas et al., 2002), ainsi que les endopeptidases LytE, LytF et CwlS (YojL) de *B. subtilis* se lient préférentiellement aux pôles des cellules. L'autolysine Atl de *S. aureus*, est retrouvée spécifiquement aux régions septales (Yamada et al., 1996), tandis que l'autolysine AcmA de *L. lactis* est localisée à la fois aux pôles des cellules et aux septa (Steen et al., 2005). Comme toutes hydrolases du peptidoglycane intervenant dans la séparation des cellules, nous avons montré que la localisation de Cse est ciblée, cette protéine étant retrouvée aux septa des cellules. Toutefois, la particularité de Cse par rapport à ces autres enzymes est que celle-ci ne se lie qu'aux septa matures où le lien de peptidoglycane entre les cellules filles est très fin (Figure 34).

Ainsi, bien que le peptidoglycane recouvre toute la surface des cellules bactériennes et que les liaisons peptidiques clivées par les enzymes de séparation cellulaire soient ubiquitaires, l'action de ces enzymes ciblées préférentiellement voire spécifiquement aux pôles et/ou aux septa des cellules permettrait d'expliquer qu'elles ne conduisent pas à la lyse cellulaire.

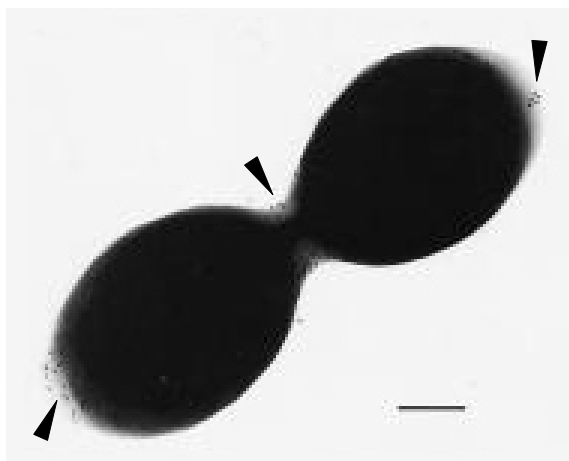
#### **3.2 Quel est le mécanisme conduisant à la localisation spécifique de Cse aux septa matures des cellules de *S. thermophilus* ?**

Plusieurs hypothèses non mutuellement exclusives peuvent être émises afin d'expliquer la localisation exclusive de Cse aux septa matures des cellules de *S. thermophilus*.

**a. Le peptide signal permet l'adressage des protéines de surface.**

Une des explications possible de la quasi-exclusivité de la fixation de la protéine Cse aux septa matures est de considérer un mécanisme spécifique d'adressage de Cse à son site d'action. Toutes les cellules vivantes présentent des mécanismes spécifiques qui permettent de cibler les protéines de surface sur leur site d'action. Chez les bactéries, le peptide signal joue un rôle primordial dans le transport des protéines à travers la membrane plasmique, en association avec les systèmes de translocation Sec et Tat (Pugsley *et al.*, 2004). Récemment, Carlsson *et al.* (2006) ont démontré que le peptide signal influe directement sur la localisation des deux protéines de surface de *S. pyogenes* : la protéine M et la protéine F (PrtF). Ces deux protéines diffèrent par leur localisation, la protéine M étant sécrétée au niveau des septa, tandis que la protéine PrtF se localise préférentiellement aux pôles cellulaires. Les protéines M et PrtF ont toutes deux, un peptide signal qui leur est propre. L'échange de ces peptides signaux distincts entre les protéines M et PrtF a conduit à un changement de leur localisation habituelle, suggérant que les peptides signaux pourraient contenir les informations nécessaires à l'adressage spécifique des protéines à la surface des cellules bactériennes, en plus de leur rôle classique dans la translocation des protéines au travers de la membrane plasmique (Carlsson *et al.*, 2006). Par conséquent, la possibilité que le peptide signal de Cse gouverne son exportation aux septa matures n'est pas à exclure.

Les résultats obtenus au cours de ce travail de recherche indiquent que l'extrémité 5' du gène *cse* contenant le peptide signal et le module LysM est probablement issue du gène ancestral *sip* de *S. thermophilus*. Ce dernier code une protéine de surface immunogène de fonction inconnue chez les streptocoques. Les seules données concernant la protéine Sip ont été obtenues chez *S. agalactiae* où celle-ci est localisée à la fois aux pôles cellulaires et aux régions septales (Figure 48) (Rioux *et al.*, 2001). Ainsi, la spécificité de localisation de la protéine Sip diffère de celle de la protéine Cse de *S. thermophilus*.



**Figure 48 : Localisation de la protéine Sip à la surface des cellules de *S. agalactiae* (Rioux *et al.*, 2001).**

La localisation de la protéine Sip est indiquée par des flèches pointées sur les grains d'or couplés aux pôles des cellules et au niveau du septum. Cette photographie a été prise en microscopie électronique à transmission.

La barre correspond à 0,2  $\mu\text{m}$ .

Le peptide signal de la protéine Cse de *S. thermophilus* (de la position 1 à 28 en résidu aminé) et celui de la protéine Sip de *S. agalactiae* (de la position 1 à 26 en résidu aminé) a révélé un important taux de divergence (60%). Cette divergence de séquence pourrait expliquer la localisation distincte de ces deux protéines. Par ailleurs, le site de clivage des peptides signaux de ces deux protéines prédits par le logiciel SignalP3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) est différent. Il s'agit respectivement des séquences aminées VHA-DE et VQA-QE pour les protéines Cse et Sip. Par conséquent, les données

actuelles nous permettent d'émettre l'hypothèse selon laquelle le peptide signal de Cse et de Sip serait directement lié à la sécrétion de ces protéines à différentes régions subcellulaires.

Afin de confirmer cette hypothèse, il serait donc intéressant d'étudier de nouveau la localisation de la protéine Cse à la surface des cellules de *S. thermophilus* après échange de son peptide signal avec celui de la protéine Sip de *S. agalactiae*.

***b. Le domaine LysM est impliqué dans l'attachement des protéines à la paroi.***

Une autre explication possible est de considérer que la spécificité de clivage de Cse pourrait être assurée par l'accès restreint de l'enzyme à son site d'action, via son domaine d'attachement à la paroi, le domaine LysM. Un argument en faveur de cette hypothèse est que de nombreuses protéines de surface, et notamment les hydrolases du peptidoglycane, à domaine LysM présentent une localisation ciblée. Par exemple, la protéine AcmA de *L. lactis* (Steen *et al.*, 2003), les protéines LytE, LytF et CwlS (YojL) de *B. subtilis* (Fukushima *et al.*, 2006 ; Yamamoto *et al.*, 2003) ou encore la protéine de surface Sip de *S. agalactiae* (Rioux *et al.*, 2001) sont retrouvées à des sites d'action spécifiques : aux pôles des cellules et/ou aux régions septales.

Par ailleurs, le domaine LysM serait nécessaire à l'activité des hydrolases du peptidoglycane sur leur site d'action. Il a été démontré que l'activité des protéines AcmA de *L. lactis* et AtlA d'*E. faecalis* qui possèdent respectivement 3 et 6 domaines LysM dans leur extrémité C-terminale était diminuée ou perdue par délétion d'un ou plusieurs de ces domaines (Steen *et al.*, 2005 ; Eckert *et al.*, 2006). Cette diminution ou perte d'activité résulterait de l'incapacité de la protéine à se lier à la paroi, et par conséquent à agir sur son site d'action. Cette hypothèse a été confirmée par l'étude de la localisation des dérivés de la protéine AcmA par immunofluorescence. Les auteurs ont montré que la fixation d'AcmA était perdue lorsque 2 des 3 domaines LysM étaient délétés (Steen *et al.*, 2005). De même, au cours de ce travail, nous avons montré que le domaine LysM est essentiel pour la séparation des cellules de *S. thermophilus*. Les résultats de la localisation par immunofluorescence ont révélé une incapacité de fixation de la protéine Cse délété du domaine LysM sur les cellules de *S. thermophilus* confirmant que celui-ci est le domaine d'attachement de Cse à la paroi.

La comparaison de la séquence protéique du domaine LysM de Cse (de la position 53 à 96 en résidu aminé) de *S. thermophilus* avec celle du domaine LysM de Sip (de la position 51 à 96 en résidu aminé) de *S. agalactiae* a révélé 39,1% de divergence qui pourrait être éventuellement responsable des localisations distinctes de ces deux protéines. Afin d'étayer cette hypothèse, il serait donc intéressant d'échanger le domaine LysM de Cse avec celui de la protéine Sip de *S. agalactiae*. Toutefois, la séquence du peptide signal de ces deux protéines est également très différente, et il n'est pas exclu que la localisation distincte de Cse et Sip soit due à la fois au peptide signal et au domaine LysM.

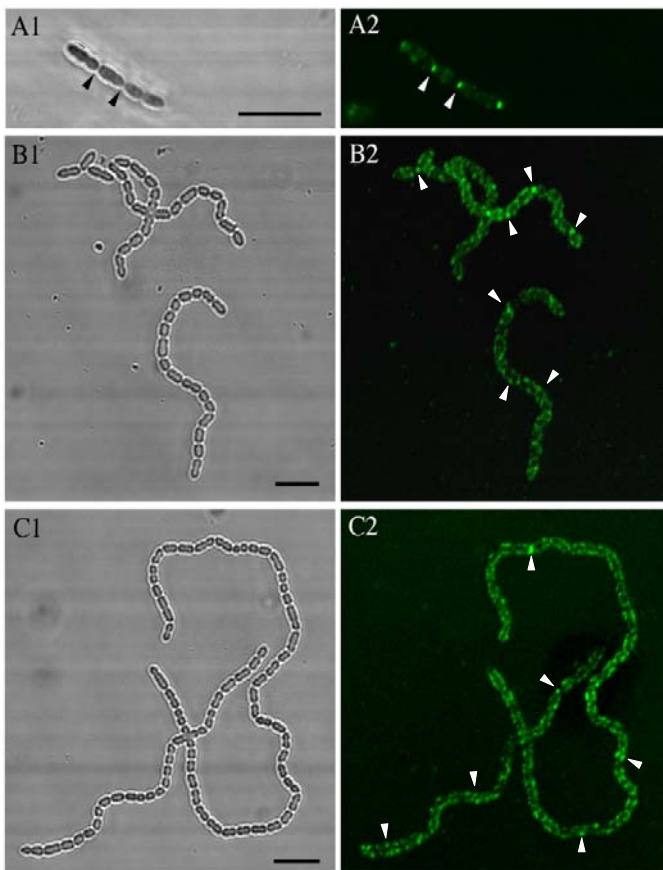
Or, les données obtenues au cours de ce travail révèlent que le peptide signal de la protéine Cse de *S. thermophilus* est identique à celui des protéines Sip de *S. salivarius* et *S. vestibularis*. Par ailleurs, son domaine LysM présente 11,4% de divergence avec ceux de ces mêmes protéines. Afin de déterminer l'impact de la divergence du domaine LysM sur la localisation ciblée de Cse aux septa matures, l'immunolocalisation de la protéine Sip à la surface des cellules de *S. salivarius* à l'aide des



anticorps polyclonaux anti-LysM<sub>Cse</sub> (reconnaissant l'extrémité N-terminale de Cse du résidu aminé 1 à 183) a été entreprise. Les résultats préliminaires obtenus n'ont pas permis de conclure à une localisation ciblée de la protéine Sip. En effet, le sérum pré-immun (témoin de contrôle) s'est avéré non valide à cause d'interaction de celui-ci avec des composants de la paroi de *S. salivarius*. En perspective, il serait donc intéressant de produire des anticorps anti-Sip spécifiques afin de déterminer la localisation de la protéine Sip à la surface des cellules de *S. salivarius*. Si celle-ci est différente de celle de la protéine Cse, une expérience d'échange du domaine LysM entre Cse de *S. thermophilus* et Sip de *S. salivarius* pourrait être envisagée. Elle permettrait de savoir si le domaine LysM est ou non le seul acteur permettant la localisation spécifique de Cse aux septa matures.

**c. La localisation spécifique de Cse peut-elle être influencée par les régions « Var » et « CHAP » ?**

Il est probable que le peptide signal et le domaine LysM agissent de manière coopérative : le peptide signal permettrait l'adressage spécifique de Cse aux septa matures, tandis que le domaine LysM assurerait son attachement à la paroi de *S. thermophilus*. Ainsi, tous deux seraient suffisants pour consentir à un clivage rigoureux et précis du peptidoglycane afin de préserver l'intégrité des cellules de *S. thermophilus* et éviter toute lyse cellulaire. Néanmoins, d'autres régions de la protéine Cse ont été jugées importantes pour le positionnement correct de cette protéine au niveau des septa matures. En effet, la délétion de la région centrale « Var » ainsi que celle du domaine catalytique CHAP a conduit à une modification de la localisation de Cse à la surface des cellules de *S. thermophilus* (Figure 49).



**Figure 49 : Immunolocalisation de la protéine Cse et de ses dérivées à la surface des cellules de *S. thermophilus*.**

A. Souche sauvage, B. Mutant *cse*ΔVar et C. Mutant *cse*ΔCHAP. Les photographies des chaînes de cellules de *S. thermophilus* ont été prises en phase de contraste au microscope confocal à balayage laser (A1 à C1).

La protéine Cse (A2) et ses dérivées *Cse*ΔVar (B2) et *Cse*ΔCHAP (C2) ont été révélées par marquage en immunofluorescence avec les anticorps Alexa-fluor488 qui s'hybrident sur les anticorps primaire, anti-LysM<sub>Cse</sub>. Sur la photographie A2, la localisation ciblée de la protéine Cse est indiquée par des flèches. Sur les photographies B2 et C2, quelques flèches sont représentées pour indiquer la localisation dispersée des protéines *Cse*ΔVar et *Cse*ΔCHAP. La barre correspond à 25 μm.

Les signaux de fluorescence (Figure 49) montrent que les protéines tronquées (Cse $\Delta$ Var et Cse $\Delta$ CHAP) sont des protéines de surface de *S. thermophilus* localisées sur tout le contour des cellules, contrairement à celle de la protéine Cse qui présente une localisation ciblée. Bien que nous ne puissions pas exclure un rôle direct de ces deux régions (Var et CHAP) sur la fixation à la paroi et/ou au site de reconnaissance du substrat, ces données suggèrent que la conformation tridimensionnelle pourrait être impliquée dans l'adressage de Cse à son site d'action. Toutefois, la localisation des protéines Cse $\Delta$ Var et Cse $\Delta$ CHAP pourrait différer de celle de Cse à cause de modifications de la conformation tridimensionnelle de ces protéines.

Néanmoins, plusieurs hypothèses concernant l'implication de la région « Var » et du domaine « CHAP » dans la localisation ciblée de Cse peuvent être émises :

- La délétion de la région « Var » suggère que celle-ci a peu ou pas d'influence sur l'activité de Cse dans le processus de séparation des cellules de *S. thermophilus*. Néanmoins, la région « Var » n'est pas dépourvue de rôle, puisque sa délétion laisse à supposer que celle-ci serait nécessaire au positionnement correct de Cse à son site d'action.

L'hypothèse privilégiée est que cette région soit un « linker » reliant les deux domaines fonctionnels LysM et CHAP entre eux. Elle pourrait avoir pour rôle d'espacer physiquement le domaine catalytique CHAP du domaine de liaison à la paroi LysM, afin que ce dernier puisse être à proximité de son site de clivage. Ceci pourrait être une explication à la localisation restreinte de Cse à la surface des cellules, et par conséquent, la délétion de la région « Var » abolirait cette restriction. Si tel est le cas, on peut supposer que le changement de la région « Var » par une autre région de faible complexité ne devrait pas influencer sur la localisation de Cse. Ainsi, l'étude de la localisation de Cse à la surface des cellules d'un mutant de *S. thermophilus* LMG18311 pourrait être réalisée afin de confirmer cette hypothèse. La particularité de ce mutant est que celui-ci présente un allèle *cse* modifié dans lequel la région « Var » a été remplacée par celle de l'allèle *cse* de la souche *S. thermophilus* CNRZ368 qui diffère en séquence et en taille (Borges *et al.*, 2005).

Par ailleurs, l'hypothèse selon laquelle la région « Var » serait un domaine de liaison à la paroi n'est pas à exclure. En effet, cette région est particulièrement riche en résidus aminés proline, sérine et thréonine. Or, le domaine de liaison à la paroi de la protéine Ftf de *S. salivarius* est également riche en ces résidus aminés (Rathsam et Jacques, 1998). Les fonctions hydroxyles des résidus aminés sérines et thréonines pourraient interagir avec des composants de la paroi en formant des liaisons hydrogènes (Rathsam et Jacques, 1998). Un rôle d'attachement à la paroi pour un domaine d'une autre protéine de *S. thermophilus* : l'hydrolase du peptidoglycane Mur-1, a également été proposé. Cette protéine associée à la paroi ne présente aucune similarité avec les domaines d'attachement connus. En revanche, elle présente une région particulière riche en résidus aminés glycine, sérine et thréonine. Les auteurs ont par conséquent fait l'hypothèse que cette région permettrait la liaison de la protéine Mur-1 à la paroi cellulaire de *S. thermophilus* (Husson-Kao *et al.*, 2000).

Il a également été démontré par des expériences de fusion avec des protéines rapporteurs, que des régions dont l'interrogation dans les banques de données n'a révélé aucun domaine connu, ni aucune homologie de séquence, sont suffisantes pour diriger spécifiquement la localisation de la protéine de fusion (Baba et Schneewind, 1998). Il s'agit des trois séquences répétées de la protéine Atl de *S. aureus* (Oshida *et al.*, 1995 ; Yamada *et al.*, 1996). La fusion de chacune de ces séquences avec une protéine rapporteur a révélé une localisation de ces protéines de fusion au niveau de l'anneau équatorial à la surface cellulaire, (c'est-à-dire au septum naissant, futur site de scission cellulaire)

(Baba et Schneewind, 1998), ce qui est en accord avec les résultats de microscopie électronique de la protéine Atl à la surface des cellules de *S. aureus* (Yamada et al., 1996). Ainsi, Baba et Schneewind (1998) ont émis l'hypothèse selon laquelle Atl se fixerait sur un récepteur de paroi spécifique responsable de sa localisation ciblée (Baba et Schneewind, 1998). De même, De Las Rivas et al., (2002) ont postulé que des récepteurs spécifiques seraient responsables du positionnement correct de la protéine LytB aux régions polaires des cellules de *S. pneumoniae*, permettant ainsi une hydrolyse localisée du peptidoglycane conduisant à la séparation des cellules (De Las Rivas et al., 2002). Cependant, dans le cas de LytB, il a été démontré que celle-ci interagissait avec les résidus cholines de la paroi cellulaire (Garcia et al., 1999b ; López et al., 2000), pouvant faire office de récepteurs spécifiques de cette protéine.

Le fait que la région « Var » ne correspond à aucun domaine connu et ne présente aucune homologie de séquence et que celle-ci soit nécessaire au positionnement correct de Cse, suggère donc que cette région, riche en motifs répétés ferait office de ligand spécifique de récepteurs responsables de la localisation ciblée de Cse. Cette région pourrait aussi interagir avec des composés de la paroi de *S. thermophilus* tels que les acides téichoïques, les acides lipotéichoïques, les polysaccharides ou encore les résidus cholines dont la présence permettrait de réguler la fixation des protéines à la surface des cellules.

En perspective, il serait donc intéressant d'élucider la fonction de la région « Var » et son implication dans la localisation spécifique de Cse. Des expériences analogues à celles des séquences répétées Atl de *S. aureus* par Baba et Schneewind (1998) pourraient être envisagées. Ainsi, la fusion de la région « Var » avec la protéine GST (pour Glutathione S-transferase) pourra être réalisée. La localisation de la protéine de fusion « GST-Var » par microscopie électronique à transmission et/ou à fluorescence, après incubation de celle-ci avec les cellules de *S. thermophilus* sera étudiée. Si la région « Var » est apte à interagir avec la paroi, la protéine hydride aurait par conséquent, une localisation pariétale qui pourrait être en accord ou non avec la localisation de la protéine Cse aux septa matures. Si tel est le cas, l'étude de l'impact du polymorphisme de la région « Var » sur le positionnement correct de Cse pourra être entrepris par des expériences de fusion entre la protéine GST et diverses combinaisons de motifs répétés constituant cette région.

- Dans le cas du domaine CHAP, il a été postulé pour l'endolysine du phage B30 de *S. agalactiae* que celui-ci pourrait jouer un rôle dans la reconnaissance au substrat, en plus de sa fonction d'hydrolyse du peptidoglycane (Donovan et al., 2006b). Afin d'étayer cette hypothèse, le changement du domaine CHAP de Cse avec celui de l'endolysine B30 pourrait être envisagé, d'autant plus que ces deux domaines ne présentent que seulement 12,8% d'identité. De même, le domaine CHAP de Cse pourrait être remplacé par celui de l'enzyme de séparation cellulaire Sle1 de *S. aureus* (Kajimura et al., 2005) présentant 29,4% d'identité et dont le substrat reconnu est différent de celui de la protéine Cse. Ainsi, si le domaine CHAP est réellement impliqué dans la reconnaissance au substrat, dans les deux cas, cet échange conduirait sans doute à une localisation et/ou une activité différente de celle de la protéine Cse.

Par ailleurs, de même que pour la région « Var », on pourrait imaginer que le domaine CHAP soit également la cible de récepteurs spécifiques responsable du positionnement correct de Cse. Néanmoins, lorsque l'on délète l'une des deux régions (Var ou CHAP), la localisation des protéines tronquées est dispersée (Figure 49). Ainsi, il n'est pas exclu que l'association du domaine CHAP avec

la région « Var » soit nécessaire pour que la protéine Cse puisse interagir spécifiquement avec des récepteurs. Si tel est le cas, en perspective, il conviendra d'identifier ses récepteurs.

***d. Les composants de la paroi pourraient être responsables de la localisation spécifique des hydrolases du peptidoglycane.***

Une hypothèse alternative de la localisation quasi-exclusive de Cse aux septa matures est de supposer que cette protéine interagit spécifiquement avec des composants de la paroi. C'est le cas par exemple, de la protéine LytA de *S. pneumoniae* qui se lie préférentiellement aux régions septales, riches en résidus cholines (De Las Rivas *et al.*, 2002). Toutefois, Cse étant dépourvue de domaine de liaison à la choline, il n'est donc pas envisageable qu'elle puisse se lier spécifiquement aux septa matures grâce à ce type de résidus.

En revanche, il est possible que les septa matures soient dépourvus de certains composants de la paroi tels que les polysaccharides ou les acides lipotéichoïques. Contrairement à ce que l'on pourrait supposer, les acides lipotéichoïques ne sont pas répartis uniformément dans la paroi cellulaire. En effet, il a été démontré que les acides lipotéichoïques qui sont des composés polyglycérolphosphates ou polyribitolphosphates associés aux molécules de glycane de la paroi, sont responsables de la fixation spécifique d'AcmA de *L. lactis*, puisque celle-ci ne se lie que dans les régions qui en sont dépourvues (Steen *et al.*, 2003). A l'inverse, la protéine AcmA peut se fixer sur l'intégralité de la surface des cellules de *L. lactis* traitées à l'acide trichloroacétique, éliminant les acides lipotéichoïques (Steen *et al.*, 2003). Par ailleurs, les polysaccharides associés à la paroi empêcherait également la fixation d'AcmA sur les cellules de *Lactobacillus helveticus* (Steen *et al.*, 2003). Ainsi certaines caractéristiques propres à l'espèce comme la synthèse de polysaccharides pourraient influencer sur la fixation des protéines à la surface des cellules bactériennes.

Afin d'étayer l'hypothèse selon laquelle les acides lipotéichoïques pourrait influencer sur la localisation ciblée de Cse, il conviendra d'étudier le positionnement de Cse sur des cellules de *S. thermophilus* traitées à l'acide trichloroacétique. Par ailleurs, si la localisation est spécifique d'espèces, une expérience consistant à étudier la localisation de Cse à la surface des cellules de différentes bactéries à coloration Gram positive par immunolocalisation pourrait être envisagée. Ceci pourrait être réalisée soit par transformation génétique de diverses bactéries en introduisant un plasmide porteur d'une copie fonctionnelle du gène *cse*, soit en incubant la protéine Cse avec des cellules de diverses bactéries.

Récemment, il a été démontré que la modification structurale du peptidoglycane par des *O*-acétylations diminue la liaison de la protéine AcmA sur la paroi modifiée de *L. lactis* (Veiga *et al.*, 2007), tandis que celle-ci n'est pas affectée lorsque la paroi de *L. lactis* est modifiée par des dé-*N*-acétylations (Meyrand *et al.*, 2007). Dans le génome de *S. thermophilus*, le gène responsable de l'acétylation du peptidoglycane (*pdgA*) est tronqué. Ainsi, il n'est pas exclu que l'acétylation ou toute autre modification de la paroi de *S. thermophilus* puisse également influencer sur le positionnement ciblé de Cse. Afin d'étayer cette hypothèse, il serait intéressant d'introduire une copie fonctionnelle du gène *pdgA* dans le génome de *S. thermophilus* et de voir son impact sur la localisation de Cse. Ainsi, si la structure chimique de la paroi de *S. thermophilus* est réellement impliquée dans la localisation de Cse, sa modification conduira sans doute à une modification de la localisation de Cse. A plus long terme, l'établissement de la cartographie de la structure du peptidoglycane de *S. thermophilus* pourrait

apporter des éléments de réponse à la localisation spécifique de Cse par une étude comparative de celle-ci avec celle d'autres espèces bactériennes.

En conclusion, lors de la division cellulaire, le processus de séparation cellulaire des bactéries, est un processus actif contrôlé génétiquement dont la finalité est de disjoindre les deux cellules filles par hydrolyse du lien du peptidoglycane qui les relie. Tous les gènes impliqués dans ce processus, codent des hydrolases du peptidoglycane localisées aux pôles et/ou septa cellulaires. Cela n'exclut pas que des paramètres physico-chimiques environnementaux puissent influencer sur la séparation physique des cellules par cassure aléatoire. Par ailleurs, le processus de séparation cellulaire n'interfère pas avec la division cellulaire, puisqu'au cours de ces travaux, nous avons montré que la morphologie des cellules du mutant *S. thermophilus* LMG18311 $\Delta$ cse n'est pas affectée, contrairement au mutant *S. thermophilus* LMG18311 $\Delta$ pcsB qui présente des cellules difformes résultant d'un positionnement et d'une orientation incorrects des plans de division (Figure 27). Ainsi, la séparation cellulaire pourrait être un processus postérieur ou ultime de la division cellulaire.

Ce travail nous a amené à émettre de nouvelles hypothèses quant au modèle de séparation cellulaire de *S. thermophilus* et soulève de nouveaux questionnements : A quel moment peut-on différencier d'un point de vue moléculaire la séparation cellulaire de la scission ? A quel moment les gènes de séparation cellulaire sont-ils exprimés au cours de la division cellulaire ? Ont-ils une expression inductible ou constitutive ? Existe-il des récepteurs spécifiques responsables de la localisation ciblée des enzymes de séparation cellulaire ? Est-ce que Cse peut interagir avec la protéine DivIVA impliquée dans la formation et la maturation des pôles cellulaires avant la séparation ? De plus, le fait que Cse soit codée uniquement par le génome de *S. thermophilus* nous amène à la question de savoir si le modèle de séparation cellulaire de *S. thermophilus* peut être transposé aux deux autres espèces du groupe *salivarius*. Les réponses à ces questions nous permettront d'élargir nos connaissances sur la séparation cellulaire et notamment de celle de *S. thermophilus*.

En perspective, il serait donc intéressant d'une part d'étudier l'expression du gène *cse* par la technique de RT-PCR à temps réel à différentes phases de la croissance de *S. thermophilus*, et d'autre part, de suivre l'expression de la protéine Cse tout au long de la croissance et la division cellulaire de *S. thermophilus* par immunolocalisation afin de savoir à quel moment précisément intervient Cse. Parallèlement, pour répondre à la question : Est-ce que Cse interagit avec d'autres protéines ? Le suivi de l'expression de protéines de division cellulaire telles que PcsB, impliquée dans le positionnement et l'orientation des plans de division ou encore DivIVA, agissant avant la rupture du lien de peptidoglycane entre les cellules filles, pourrait être réalisé. Il conviendra également de rechercher les acteurs de la séparation cellulaire de *S. salivarius* et de *S. vestibularis* afin de les comparer avec ceux de *S. thermophilus*. Ceci permettra de savoir si le modèle de séparation des cellules de *S. thermophilus* peut être ou non transposé aux deux autres espèces du groupe *salivarius*.

### **III. Mécanisme de création de la diversité des hydrolases du peptidoglycane : implication du réassortiment de modules et des régions de faible complexité.**

Au cours de ce travail, nous avons mis en évidence que les hydrolases du peptidoglycane impliquées dans le processus de séparation cellulaire sont très diverses. Cette diversité peut s'expliquer d'une part par la nature variée des domaines catalytiques et des domaines d'attachement au peptidoglycane, et d'autre part par les différentes associations de ces domaines entre eux (Figure 47). Par ailleurs, les données bibliographiques suggèrent que le mécanisme majeur responsable de la diversité des hydrolases du peptidoglycane en général serait le réassortiment de modules. La création du gène *cse* par réassortiment de modules, codant une hydrolase du peptidoglycane est un exemple supplémentaire.

#### **1. L'architecture modulaire des hydrolases du peptidoglycane contribue à leur diversité.**

La plupart des hydrolases du peptidoglycane se caractérise par une structure modulaire composée d'au moins deux domaines fonctionnels : l'un contenant le site catalytique de l'enzyme et l'autre responsable de la reconnaissance et/ou fixation à des composants de la paroi cellulaire. C'est le cas des enzymes de séparation cellulaire qui disposent toutes d'un domaine d'hydrolyse du peptidoglycane et d'un ou plusieurs domaines d'attachement à la paroi (Figure 47). *Cse* n'échappe pas à cette règle puisqu'elle présente à son extrémité C-terminale, le domaine catalytique CHAP et à son extrémité N-terminale, un domaine d'attachement à la paroi cellulaire, le domaine LysM.

Au sein du phylum des *Firmicutes*, les hydrolases du peptidoglycane présentent des domaines catalytiques et des domaines de liaison et/ou reconnaissance du substrat très variés. En effet, l'analyse bioinformatique de l'ensemble des hydrolases du peptidoglycane des *Firmicutes* a révélé 14 domaines catalytiques distincts qui sont répartis en fonction de leur spécificité enzymatique sur la figure 50. Certaines classes d'enzymes présentent plusieurs domaines catalytiques distincts capables de cliver la même liaison peptidique du peptidoglycane. C'est le cas par exemple, de la classe des *N*-acétylmuramoyl-L-alanine amidases dont l'activité catalytique peut être assurée par quatre domaines distincts : « Amidase\_2 » ; « Amidase\_3 » ; « Amidase\_5 » et « CHAP (Figure 50).

Les hydrolases du peptidoglycane peuvent posséder plusieurs domaines catalytiques identiques telle que l'endopeptidase YwtD de *B. subtilis* présentant trois domaines NLPC/P60 (Suzuki et Tahara, 2003), ou différents. Dans ce dernier cas, la présence de domaines catalytiques distincts confère aux hydrolases du peptidoglycane des activités enzymatiques multiples. Ainsi, l'autolysine Atl de *S. aureus* est une enzyme bifonctionnelle qui possède à son extrémité N-terminale, un domaine « Amidase\_2 » conférant une activité *N*-acétylmuramoyl-L-alanine amidase et à son extrémité C-

terminale, un domaine « Glucosaminidase » responsable de son activité  $\beta$ -endoglucosaminidase (Biswas *et al.*, 2006 ; Oshida *et al.*, 1995 ; Yamada *et al.*, 1996). L'endolysine du phage B30 de *S. agalactiae* présente quant à elle, à la fois les activités *N*-acétylmuramidase et endopeptidase résultant respectivement de l'action indépendante de ses domaines « Glyco\_hydro\_25 » et « CHAP » (Donovan *et al.*, 2006a ; Pritchard *et al.*, 2004). Un autre exemple, est celui de l'enzyme LysWMY du phage  $\phi$ WMY de *Staphylococcus warneri* dont l'activité *N*-acétylmuramoyl-L-alanine amidase et l'activité endopeptidase résultent respectivement de la présence des domaines catalytiques « Amidase\_2 » et « CHAP » (Yokoi *et al.*, 2005). L'enzyme  $\phi$ MR11 de *S. aureus* présente quant à elle, un domaine catalytique « CHAP » et un domaine catalytique « Glucosaminidase » (Rashel *et al.*, 2008).

La figure 50 indique ainsi pour chacun des domaines catalytiques distincts de chacune des classes d'hydrolases du peptidoglycane, les domaines pouvant y être associés. Ces domaines peuvent être, des domaines catalytiques additionnels et/ou des domaines de surface de nature diverse, tels que des domaines de liaison à la paroi, des domaines de liaisons à la couche S et des domaines d'adhésion à la matrice extracellulaire. Au total, 27 domaines de surface ont été répertoriés (Layec *et al.*, 2008b). Les domaines d'attachement à la paroi les plus couramment observés sont LysM, ChBD (pour Choline binding protein domain) et SH3 (indiqués dans les octogones en blanc, figure 50). Ces domaines sont responsables de la spécificité de liaison à la paroi cellulaire des hydrolases du peptidoglycane. Par exemple, la liaison de la protéine LytA aux résidus cholines de la paroi cellulaire de *S. pneumoniae* est promue par le domaine ChBD de cette dernière (Garcia *et al.*, 1994 ; De Las Rivas *et al.*, 2002). De même, la protéine ALE-1 de *Staphylococcus capitis* se lie spécifiquement au peptidoglycane des staphylocoques grâce à son domaine SH3 (Fujiwara *et al.*, 2005). Les domaines de liaison peuvent être présents en copie unique ou au contraire en multiple copie. Par exemple, Cse est caractérisée par un seul domaine LysM alors que AtlA d'*E. faecalis* et AcmA de *L. lactis* présentent respectivement 6 et 3 domaines LysM à leur extrémité C-terminale (Eckert *et al.*, 2006 ; Steen *et al.*, 2005). Les protéines LytA et LytB de *S. pneumoniae* possèdent quant à elles, respectivement 6 et 12 domaines ChBD (Sanchez-Puelles *et al.*, 1990 ; De Las Rivas *et al.*, 2002).

Par ailleurs, les hydrolases du peptidoglycane peuvent présenter divers domaines de surfaces associés à un domaine catalytique unique. Tel est le cas de la protéine p60 de *L. monocytogenes* qui présente deux domaines de liaison distincts : LysM et SH3 associés à un domaine catalytique NLPC/P60 (Kohler *et al.*, 1991 ; Machata *et al.*, 2005). De même, une hydrolase du peptidoglycane peut posséder plusieurs domaines catalytiques associés à un domaine de surface unique. C'est par exemple le cas des enzymes B30 de *S. agalactiae* et LysWMY de *S. warnerii* qui présentent deux domaines catalytiques associés à l'unique domaine de liaison SH3 (Donovan *et al.*, 2006a ; Pritchard *et al.*, 2004 ; Yokoi *et al.*, 2005).

Par conséquent, la diversité des hydrolases du peptidoglycane est non seulement reflétée par l'importante variété des domaines catalytiques et des domaines de liaison et/ou reconnaissance de substrat, mais aussi par la multiplicité combinatoire de leurs associations.

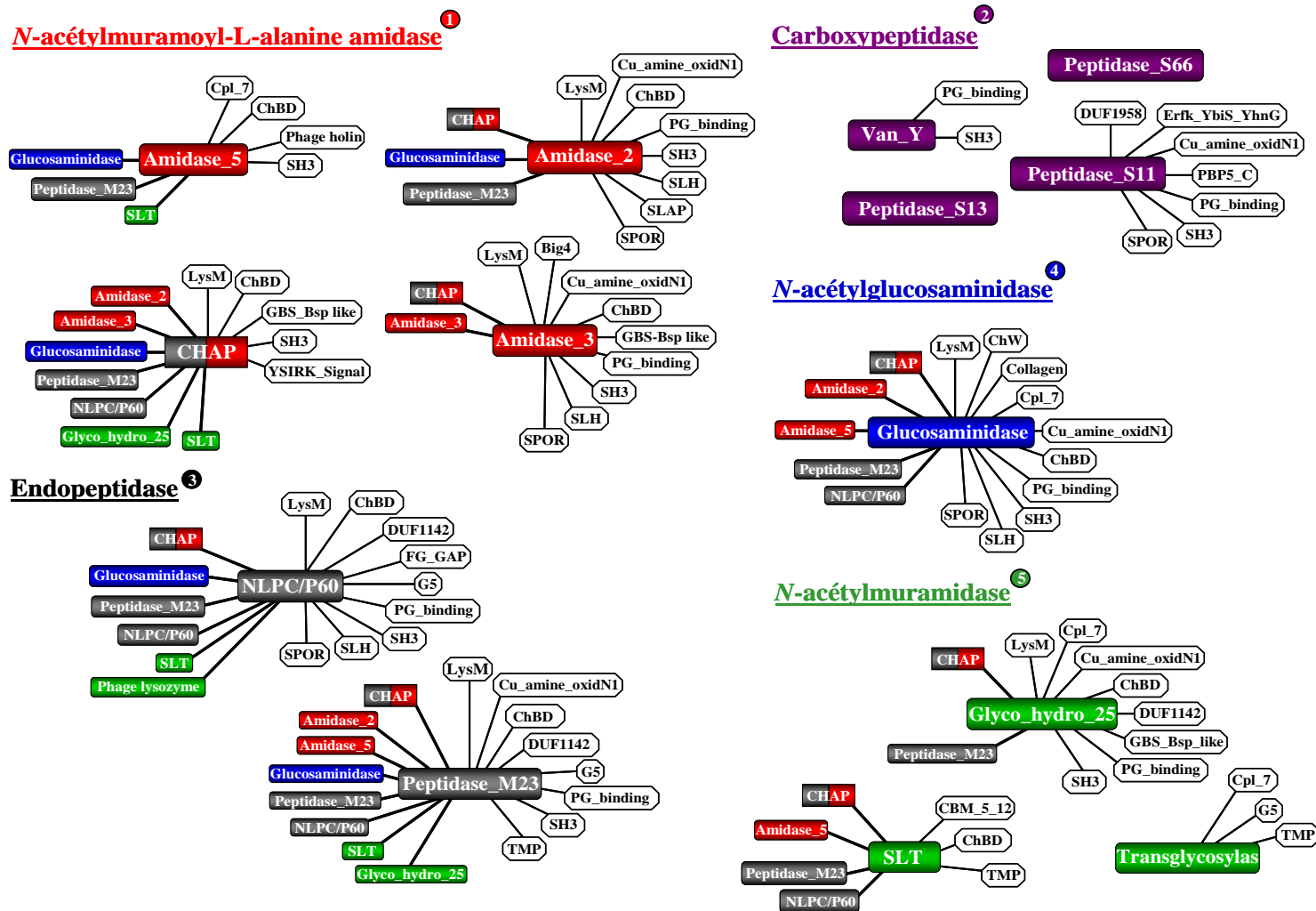


Figure 50 : Diversité des architectures modulaires des 5 classes d'hydrolases du peptidoglycane au sein des *Firmicutes* (d'après Layec et al., 2008b).

Chaque classe est représentée par un ou plusieurs domaines catalytiques distincts dont la couleur est en association avec leur spécificité de clivage. Au sein des hydrolases du peptidoglycane des *Firmicutes*, ces domaines catalytiques peuvent être associés à des domaines catalytiques identiques ou différentes et/ou à des domaines de substrat indiqués par des octogones en blanc. Les différentes possibilités d'association à partir d'un domaine catalytique distinct sont représentées schématiquement par des traits noirs. La séquence protéique de chacune des hydrolases du peptidoglycane a été soumise à la banque de données Pfam pour la recherche des domaines additionnels (<http://www.sanger.ac.uk/Software/Pfam/>) (Finn et al., 2006). Parmi les domaines de substrat, on retrouve des domaines de liaison à la paroi (LysM, ChBD, SH3, PG\_binding, GBS\_Bsp like, Cpl-7, SPOR) ; des domaines d'adhésion (FG-GAP, G5) ; des domaines de liaison à la couche S (SLH, SLAP) ; le domaine de liaison aux carbohydrates (CBM\_5\_12) ; le domaine de liaison à la pénicilline (PBP5\_C) ; le domaine de liaison à la matrice extracellulaire (Collagen) ; le domaine d'oxydation (Cu\_amine\_oxidN1) ; des domaines phagiques (TMP, Phage\_holin) ; un motif transmembranaire (YSIRK\_signal) ; un domaine de liaison aux immunoglobulines (Big4) et des domaines de fonction inconnue (DUF1142, DUF1958, Erfk\_YbiS\_YhnG).



## **2. Le réassortiment de modules est à l'origine de la diversité des hydrolases du peptidoglycane.**

Pour expliquer la diversité des hydrolases du peptidoglycane, l'hypothèse d'une évolution de ces enzymes par réassortiment de modules a été émise et confirmée par la construction *in vitro* de multiples hydrolases du peptidoglycane chimériques fonctionnelles, notamment chez *S. pneumoniae* (Croux *et al.*, 1993a et 1993b ; Diaz *et al.*, 1990 et 1991 ; Sanz *et al.*, 1996 ; Sheenan *et al.* 1996). Ce mécanisme d'évolution est basé sur l'interchangeabilité des modules fonctionnels. Il a l'avantage d'apparaître très efficace pour la création de la diversité, puisqu'un seul événement de recombinaison entre deux modules peut être suffisant pour créer une protéine chimérique fonctionnelle présentant de nouvelles propriétés (Cramer *et al.*, 1998 ; López *et al.*, 2000).

Dans la littérature ainsi qu'au cours de ce travail, des gènes codant des hydrolases du peptidoglycane créés spontanément par réassortiment de modules ont été rapportés, et notamment chez les streptocoques. L'ensemble de ces données suggère que la structure modulaire des hydrolases du peptidoglycane pourrait faciliter la formation d'enzymes chimériques fonctionnelles.

Un exemple de chimère spontanée dans le genre *Streptococcus* est l'enzyme Pal du phage Dp-1 de *S. pneumoniae*, dont la nature mosaïque résulterait de réarrangement de modules provenant des génomes bactériens et phagiques (Sheehan *et al.*, 1997). Cette protéine se caractérise par la présence, à son extrémité N-terminale, d'un domaine d'hydrolyse du peptidoglycane qui proviendrait du phage BK5-T de *L. lactis*, tandis que son extrémité C-terminale porte un domaine ChBD homologue à celui porté par les protéines lytiques de *S. pneumoniae*. La structure chimérique de cette enzyme confère une activité d'hydrolase du peptidoglycane chez *S. pneumoniae*. Par conséquent, cette enzyme fonctionnelle illustre l'émergence de nouvelles protéines mosaïques dans le génome bactérien par réassortiment de modules intergéniques. Un autre exemple, est celui de la protéine Skl du phage SK137 de *Streptococcus mitis*. Cette protéine présente à son extrémité N-terminale, le domaine catalytique CHAP et à son extrémité C-terminale, le domaine de la liaison à la choline, ChBD, semblable également à celui de la plupart des hydrolases du peptidoglycane de *S. pneumoniae* (Llull *et al.*, 2006). Les auteurs ont montré que Skl est le premier exemple de protéine ChBD à domaine CHAP ayant une activité *N*-acétylmuramoyl-L-alanine amidase. Un mécanisme d'évolution de Skl par réassortiment de modules a été postulé par les auteurs (Llull *et al.*, 2006 ; López *et al.*, 2000).

Ainsi, la protéine Cse de *S. thermophilus* constitue un nouvel exemple de protéine chimérique provenant d'un réarrangement spontané de domaines du genre *Streptococcus*. Les résultats actuels laissent supposer qu'elle résulte de la fusion du domaine de liaison à la paroi, LysM et du domaine catalytique, CHAP.

Néanmoins, si le réassortiment de module semble être impliqué dans la création des hydrolases du peptidoglycane, et par conséquent, contribue à leur diversification, la pression de sélection s'exerçant sur le maintien de ces enzymes reste à déterminer. Il est intéressant de noter que la plupart des hydrolases du peptidoglycane de *S. pneumoniae*, ainsi que celles des phages qui infectent cette même bactérie sont caractérisées par la présence du domaine ChBD. Contrairement à la paroi cellulaire des autres bactéries, celle de *S. pneumoniae* est riche en résidus choline, et par conséquent, il a été postulé que cet enrichissement particulier en résidus cholines de la paroi cellulaire de

*S. pneumoniae* agirait comme un élément de pression de sélection sur le maintien du domaine ChBD (López et al., 2000). Ce domaine de liaison est retrouvé associé à divers domaines catalytiques tels que « Amidase\_2 » pour la protéine LytA, « Glucosaminidase » pour la protéine LytB ou encore « Glyco\_hydro\_25 » pour la protéine LytC de *S. pneumoniae* (De Las Rivas et al., 2002 ; Garcia et al., 1999b ; Sanchez-Puelles et al., 1986). Si tel est le cas, néanmoins quelle pourrait être la pression de sélection s'exerçant sur les autres hydrolases du peptidoglycane qui ne se lient pas aux résidus cholines ?

### **3. La création de nouvelles hydrolases du peptidoglycane pourrait être favorisée par la présence de régions de faible complexité.**

Le succès de protéines fonctionnelles par réassortiment de modules résulte du fait que la fonctionnalité de chacun des domaines reste intacte. Par conséquent, le site de fusion doit être localisé dans des régions dites « interdomaines » dont les séquences sont de faible complexité (De Château et Björck, 1994). Ces séquences sont qualifiées de faible complexité puisqu'elles sont quasi-exclusivement constituées de répétitions dont la composition en nucléotides ou en acides aminés de la séquence est biaisée (Wootton et Ferderhen, 1996).

La présence de la région « Var » entre les deux domaines fonctionnels, LysM et CHAP de la protéine Cse d'une part et la faible complexité de celle-ci d'autre part suggèrent qu'il s'agisse d'une région interdomaine. Un argument en faveur de cette hypothèse est que le remplacement de la région « Var » de l'allèle *cse* de la souche *S. thermophilus* LMG18311 au locus chromosomique par celle de l'allèle *cse* de la souche *S. thermophilus* CNRZ368 a peu ou pas d'influence sur l'activité de séparation cellulaire de la protéine Cse, bien que ces régions présentent une divergence de 40% et une différence de 60 pb (Borges et al., 2005). Un résultat analogue a été obtenu lors de l'étude de l'impact d'une variation de la longueur de la charnière interdomaine des protéines NifA et NtrC sur l'activité régulatrice de ces deux protéines (Wootton et Drummond, 1989).

Par ailleurs, l'analyse de la séquence de la région « Var » de 17 allèles *cse* supplémentaires a montré que celle-ci est variable en séquence et en taille. Cette région a la particularité d'être exclusivement constituée de motifs répétés en tandem ou dispersés (Borges et al., 2006). Cette caractéristique est partagée par de nombreuses protéines de surface, notamment chez les streptocoques. La plus connue est la protéine M des streptocoques pathogènes dont la variabilité en séquence et en nombre de motifs répétés est un important facteur de la virulence car il permet entre autre, d'échapper au système immunitaire de l'hôte (Fischetti, 1989 ; Hollingshead et al., 1987). Les séquences répétées, puisqu'elles sont la cible d'événement de recombinaison, participent à l'évolution rapide des protéines de surface des bactéries à coloration Gram positive (Kehoe, 1994). Par analogie, on pourrait supposer que la présence de motifs répétés dans la région interdomaine ferait du réassortiment de module, un mécanisme d'évolution rapide permettant la création de nouvelles protéines fonctionnelles.

Ainsi, ces régions variables et de faible complexité car riches en motifs répétés pourraient être à l'origine du succès du réassortiment de modules dans l'évolution des protéines de surfaces, et notamment des hydrolases du peptidoglycane des bactéries à coloration Gram positive. Les régions de faible complexité et variables observées au sein du gène *cse* ont été définies au cours de ce travail,

comme étant le siège des bornes du réassortiment de modules ayant conduit à la création de *cse* chez *S. thermophilus* (Borges et al., 2005 et 2006).

#### **4. Conséquence fonctionnelle du réassortiment de module à l'origine de la création de *cse*.**

Le mécanisme de réassortiment de modules, puisqu'il crée de la diversité, contribue potentiellement à la création de gènes codant des protéines de fonction nouvelle.

Au cours de ce travail, nous avons montré que le gène *cse*, est un gène chimérique résultant de l'association du module LysM de *sip* avec le module CHAP de *pcsB* (Borges et al., 2005). Le gène *cse* est fonctionnel et code l'hydrolase du peptidoglycane majeure de séparation cellulaire de *S. thermophilus*. Le fait que le gène *cse* soit présent uniquement dans le génome de *S. thermophilus*, nous a amené à supposer que le réassortiment de modules aurait pu contribué à l'émergence d'une fonction nouvelle chez *S. thermophilus*.

Ainsi, l'extrémité 5' du gène *cse* est homologue à celle du gène *sip*. Contrairement, au gène *cse*, le gène *sip* est présent dans tous les génomes des streptocoques séquencés, à l'exception du génome de *S. thermophilus*. De même que *cse*, le gène *sip* code une protéine de surface liée à la paroi cellulaire. Cependant, contrairement à Cse, la protéine Sip ne présente aucun domaine d'activité d'hydrolyse du peptidoglycane. Le réassortiment de modules aurait donc permis de créer une nouvelle hydrolase du peptidoglycane dans le génome de *S. thermophilus*.

La fonction de Sip reste inconnue, il n'est pas exclu qu'elle puisse jouer un rôle directement ou indirectement dans la séparation cellulaire en interagissant avec une protéine inconnue porteuse d'un domaine de clivage du peptidoglycane. Si sa localisation est compatible avec cette hypothèse, alors la protéine Sip pourrait être un récepteur d'enzymes de séparation. Ainsi, l'analyse des conséquences phénotypiques de la délétion du gène *sip* chez *S. salivarius* ou *S. vestibularis* devra être réalisée.

Par ailleurs, l'extrémité 3' du gène *cse* est homologue à celle du gène *pcsB*. Les gènes *cse* et *pcsB* codent tous deux des hydrolases du peptidoglycane ayant en commun le domaine CHAP, porteur de l'activité catalytique de la protéine. Le fort pourcentage entre le domaine CHAP de ces deux protéines laisse supposer qu'ils présentent la même spécificité d'hydrolyse du peptidoglycane de *S. thermophilus*. Pourtant, contrairement au mutant  $\Delta pcsB$  de *S. thermophilus* dont la morphologie cellulaire est fortement affectée par le positionnement et l'orientation aléatoire des septa, le mutant  $\Delta cse$  de *S. thermophilus* ne présente pas de modification de sa morphologie cellulaire (Figure 27). Ainsi, Cse et PcsB ont des fonctions différentes : la première est une enzyme de séparation cellulaire et la deuxième, une enzyme de division cellulaire. Il est probable que la divergence fonctionnelle entre Cse et PcsB résulte en partie à leur domaine N-terminal associé au domaine catalytique CHAP ayant des fonctions différentes (LysM, dans le cas de Cse et la région COG3883 de fonction inconnue dans le cas de PcsB).

Ainsi, le gène *cse* constitue un bon exemple de gène chimérique créé par réassortiment de modules ayant conduit à l'émergence d'une protéine de fonction nouvelle chez *S. thermophilus*.

En conclusion, la création de nouvelles protéines par réassortiment de modules pourrait s'expliquer de deux façons non mutuellement exclusives. L'une, par le fait que chacun des domaines est indépendamment actif (Diaz *et al.*, 1990 et 1991 ; Buist *et al.*, 1995 ; Mesnage et Fouet, 2002 ; Huard *et al.*, 2003 ; Steen *et al.*, 2003). Ainsi, une nouvelle combinaison de domaine peut être fonctionnelle (Diaz *et al.*, 1990 et 1991 ; Croux *et al.*, 1993a et 1993b). Et l'autre, par la présence de régions de faible complexité situées à l'extérieur des modules, qui favoriseraient le réassortiment de modules en étant le siège des réarrangements. Toutefois, le succès évolutif du réassortiment de modules n'est que le résultat des pressions de sélection qui s'exercent à maintenir ces nouveaux gènes codant des protéines fonctionnelles dans le génome bactérien.

D'ailleurs, on pourrait supposer que plus la région de faible complexité est grande, plus les chances qu'un événement de réassortiment de modules puisse avoir lieu sont importantes. Ainsi, il n'est pas exclu qu'une corrélation positive entre la taille des régions de faible complexité et la fréquence de création de nouvelles protéines fonctionnelles par réassortiment de modules puisse exister. Afin d'étayer cette hypothèse, il serait intéressant de déterminer l'impact de longueur de la région « Var » de Cse sur le réassortiment de modules. Ainsi, la construction expérimentale de plusieurs plasmides porteurs de région « Var » de tailles différentes pourrait être réalisée. Les plasmides seront de deux types : l'un porteur d'une fusion transcriptionnelle entre un promoteur et la séquence complète ou partielle de « Var » et l'autre, porteur d'une séquence complète ou partielle de « Var » fusionné à un gène rapporteur, par exemple à un gène de résistance à un antibiotique. L'événement de réassortiment entre ces deux types de plasmides s'il a lieu, sera sélectionné par l'expression du gène de résistance chez un organisme bactérien tel que *E. coli* ou *B. subtilis*. De plus, la nature des motifs répétés de la région de faible complexité de Cse pourrait également influencer sur le réassortiment de modules. Si tel est le cas, il serait donc intéressant de savoir si les deux groupes de « Var » établis par F. Borges (Borges *et al.*, 2006) ont un impact sur le réassortiment de modules. Pour cela, la séquence « var » du groupe I et la séquence « var » du groupe II pourront être fusionnées soit à un promoteur, soit à un gène de résistance afin de savoir si la fréquence de réassortiment est dépendante de la nature de la région « Var ».

Néanmoins, si le mécanisme majeur de l'évolution des hydrolases du peptidoglycane semble être le réassortiment de modules, les forces évolutives permettant la sélection de ces enzymes, et tout particulièrement, les enzymes de séparation cellulaire spécifiques d'espèces, restent à déterminer.

## **IV. Impact de la longueur des chaînes sur l'adaptation de *Streptococcus thermophilus* à son environnement.**

Le fait que le gène *cse* soit présent dans toutes les souches de *S. thermophilus* et qu'il soit spécifique de cette espèce suggèrent que sa sélection confère un avantage adaptatif à *S. thermophilus*. Par ailleurs, ce gène impliqué dans le processus de séparation cellulaire régule probablement la longueur des chaînes de cellules de *S. thermophilus*. Ainsi, la question de l'impact de la longueur des chaînes de cellules sur la valeur adaptative de *S. thermophilus* se pose.

### **1. Facteurs majeurs influençant la croissance des cellules en chaîne.**

Les données bibliographiques ainsi que celles obtenues au cours de ce travail, indiquent que l'agencement des cellules est contrôlé par l'action d'hydrolases du peptidoglycane amenant à des cellules bactériennes sous forme unicellulaire isolée, associées en paire ou en chaîne. Les paramètres environnementaux peuvent aussi influencer sur le nombre de cellules par chaînes bactériennes.

#### **1.1 Les facteurs génétiques déterminant le caractère chaînes de cellules.**

##### *a. La croissance des cellules en chaîne serait spécifique d'espèces.*

Au cours de ce travail, nous avons constaté que les enzymes de séparation cellulaire sont spécifiques d'espèces. Chez les streptocoques, trois enzymes de séparation cellulaire ont été caractérisées. Ainsi, la séparation cellulaire de *S. pneumoniae* est assurée par l'endo- $\beta$ -glucosaminidase, LytB (De Las Rivas et al., 2002 ; Garcia et al., 1999a). Chez *S. mutans*, l'enzyme de séparation cellulaire est la protéine Aml, une *N*-acétylmuramidase (Yoshimura et al., 2006). Chez *S. thermophilus*, l'enzyme de séparation cellulaire est l'endopeptidase Cse.

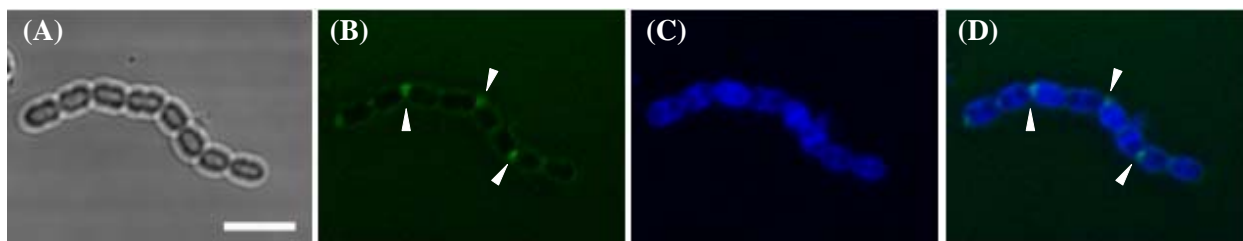
Ainsi, ces enzymes spécifiques d'espèces pourraient permettre de contrôler la longueur des chaînes de cellules qui est un critère propre de l'espèce, sans pour autant interférer avec celles des autres bactéries. Un argument en faveur de cette hypothèse est que la protéine Aml de *S. mutans* est capable de lyser les cellules de *S. mutans* et *Streptococcus sobrinus*, deux espèces phylogénétiquement proches (Facklam, 2002) mais d'aucun autre streptocoque (Yoshimura et al., 2006). De même, Cse pourrait contrôler spécifiquement la longueur des chaînes de cellules de *S. thermophilus*, et/ou de *S. salivarius* et *S. vestibularis*, les deux espèces les plus proches phylogénétiquement de *S. thermophilus*. Afin d'étayer cette hypothèse, il serait intéressant de tester l'activité de Cse sur les cellules des mutants délétés du gène *lytB* de *S. pneumoniae* et du gène *aml* de *S. mutans*, ainsi que sur les cellules de tous les autres streptocoques. Cette étude pourrait être élargir aux espèces bactériennes partageant la même niche écologique que *S. thermophilus*, telles que *Lb. bulgaricus* et *L. lactis*.

**b. La croissance des cellules en chaîne résulterait d'une déficience en hydrolase du peptidoglycane.**

Les données bibliographiques ainsi que ce travail indiquent que l'efficacité du processus de séparation des cellules n'a aucune incidence sur la morphologie cellulaire, et par conséquent sur la division cellulaire. Par contre, elle a un impact direct sur la taille des chaînes de cellules.

L'inactivation de gènes codant les enzymes de séparation cellulaire conduit à la formation de très longues chaînes de cellules suggérant que la croissance en chaîne pourrait résulter d'une déficience en ces enzymes et/ou d'une action partielle.

*S. thermophilus* est une espèce bactérienne où les chaînes de cellules peuvent être constituées de plus de 100 cellules, tel est le cas pour la souche LMG18311, par exemple. Ceci suggère que le nombre de liens peptidiques non rompus peut être très élevé. Les résultats de microscopie à fluorescence montrent que Cse ne se fixe pas à tous les septa matures (Figure 51). Ainsi, Cse assure partiellement la séparation des cellules. Une insuffisance de l'expression du gène *cse* et/ou une inefficacité de la protéine Cse ou encore une déficience en hydrolases du peptidoglycane dans le génome de *S. thermophilus* pourraient expliquer que tous les liens peptidiques aux septa matures ne soient pas clivés.



**Figure 51 : Localisation de Cse à la surface des cellules de *S. thermophilus* CNRZ368.**

(A). La photographie de la chaîne de cellules de *S. thermophilus* a été prise en phase de contraste au microscope confocal à balayage laser. (B). Localisation de la protéine Cse par immunofluorescence. (C). Marquage de la paroi de *S. thermophilus* avec le calcofluor se fixant sur les molécules de glycanes. (D). Résultat du double marquage. La barre correspond à 5  $\mu$ m.

Il a été démontré pour une souche de *S. thermophilus* que le nombre de cellules par chaîne peut être fortement réduit sous l'action d'enzymes. En effet, les chaînes de cellules de cette souche ont été raccourcies jusqu'à deux cellules par chaînes en moyenne sous l'action de lysozyme d'œuf de poule ou de l'enzyme AcmA de *L. lactis* ajoutée dans le milieu de culture à une concentration non limitante pour la croissance de *S. thermophilus* (Mercier et al., 2000). Par ailleurs, il est intéressant de constater que les chaînes de cellules de *L. lactis* sont plus courtes (moins de 10 cellules) que celle de *S. thermophilus*. L'hypothèse privilégiée à cette constatation serait que les chaînes de cellules de *S. thermophilus* sont plus longues à cause d'une déficience en enzyme de séparation cellulaire ou d'autres hydrolases du peptidoglycane. Un argument en faveur de cette hypothèse est que les bactéries à courte chaîne de cellules présentent plusieurs enzymes pouvant contribuer directement ou indirectement à la séparation des cellules. Ainsi chez *L. lactis*, où l'enzyme de séparation cellulaire est AcmA, quatre autres hydrolases du peptidoglycane ont été identifiées. Il s'agit des enzymes AcmB, AcmC, AcmD et YjgB (Buist et al., 1995 ; Huard et al., 2003 et 2004 ; Redko et al., 2007) qui sont toutes absentes du génome de *S. thermophilus*. De même, l'enzyme majeure LytB assurant la séparation cellulaire et les deux hydrolases du peptidoglycane LytA et LytC du diplocoque *S. pneumoniae* sont absentes du génome de *S. thermophilus* (Garcia et al., 1999a et 1999b ; Sanchez-

Puelles et al., 1986). Une analyse bioinformatique du génome de *S. thermophilus* n'a pas révélé d'enzymes de séparation cellulaire autre que Cse ni d'autre enzyme telle que des autolysines ou lysozymes, à l'exception de l'hydrolase du peptidoglycane Mur-1 dont la fonction reste inconnue (Husson-Kao et al., 2000).

*c. Lien entre la division cellulaire et la croissance des cellules en chaîne.*

Des protéines de division cellulaire peuvent également influencer la longueur des chaînes de cellules. Un phénotype « longue chaîne de cellules » notamment chez *S. thermophilus* a également été observé pour les mutants délétés du gène *rodA*, *pbp2b* et *mreD* (Thibessard et al., 2002b et 2004). Ces gènes codent des protéines essentielles à la synthèse du peptidoglycane non septal et longitudinal permettant l'accroissement de la taille cellulaire au cours de la division bactérienne (Noirclerc-Savoie et al., 2003 ; Ishino et al., 1986). Néanmoins, le phénotype est beaucoup moins drastique que celui du mutant  $\Delta cse$  et il présente des modifications de la morphologie cellulaire.

A l'inverse, un phénotype « courte chaîne de cellules » a été observé pour le double mutant  $\Delta\Delta acmA-ponA$ . L'inactivation du gène *ponA* chez le mutant *L. lactis* MG1363 $\Delta acmA$  présentant un phénotype « longue chaîne de cellules », s'est traduite par une diminution importante du nombre de cellules par chaînes (Mercier et al., 2002). Le gène *ponA* code la protéine PBP1a, une enzyme bifonctionnelle essentielle dans la biosynthèse du peptidoglycane lors de la division cellulaire (pour revue Goffin, 1998). Cette enzyme présente une activité de transglycosylation permettant l'élongation des chaînes de glycane et une activité de transpeptidation, permettant l'insertion des chaînes de glycane néoformée dans le peptidoglycane. Par conséquent, le raccourcissement des chaînes de cellules résulte probablement de la fragilisation des liaisons peptidiques du peptidoglycane entraînant des cassures (Mercier et al., 2002).

Ces enzymes de division cellulaire n'intervenant que dans le métabolisme du peptidoglycane, il est peu probable qu'elles agissent directement dans le processus de séparation cellulaire. Toutefois, une action indirecte de ces enzymes n'est pas exclue. Le processus de séparation cellulaire étant l'ultime étape de la division cellulaire, il est possible qu'il ne puisse se dérouler correctement que si le mécanisme de scission des cellules est à son terme. Dans ce cas, une inhibition ou un ralentissement de la division cellulaire à une étape antérieure à la séparation cellulaire pourrait avoir un impact négatif sur le déroulement de celle-ci.

## **1.2 Les facteurs environnementaux influent sur la longueur des chaînes de cellules.**

Le nombre de cellules par chaînes peut aussi être influencée par des paramètres environnementaux. Dans certains cas, il a été démontré que la variation de la longueur des chaînes de cellules était corrélée à des modifications physico-chimiques de l'environnement. Ainsi, la variation de la longueur des chaînes de cellules constitue dans certains cas une réponse adaptative. Par exemple, chez *S. mutans* et *S. sobrinus*, la longueur des chaînes de cellules augmente avec l'acidification du milieu (Nakano et al., 2005 ; Thibodeau et Ford, 1991). Inversement, la taille des chaînes de cellules est diminuée en milieu alcalin (Thibodeau et Ford, 1991). Par ailleurs, une augmentation de la longueur des chaînes de cellules de *S. agalactiae* a été observée lors de l'inactivation des gènes *stk1* et *stp1*, codant respectivement une sérine/thréonine kinase et une phosphatase qui sont impliquées dans

la transduction des signaux environnementaux (Rajagopal et al., 2003), suggérant que les signaux perçus par la bactérie dans l'environnement peut influencer sur l'agencement de ses cellules. Chez *L. lactis*, la longueur des chaînes de cellules peut être diminuée en condition de stress. Il a été démontré que l'inactivation du gène *htrA* codant une protéase impliquée dans la maturation d'AcMA de *L. lactis* se traduit par une augmentation de la longueur des chaînes de cellules (Poquet et al., 2000). Or le taux de transcrite du gène *htrA* est augmenté en condition de stress conduisant à une diminution des chaînes de cellules de *L. lactis* (Foucaud-Scheunemann et Poquet, 2003).

Ainsi, la longueur des chaînes de cellules est un caractère morphologique qui est modulé suite à des modifications des paramètres environnementaux, suggérant que la croissance des cellules en chaîne peut conférer un avantage adaptatif.

## **2. Influence de la longueur des chaînes de cellules sur l'adaptation des streptocoques à leur environnement.**

### **2.1 La longueur des chaînes de cellules module la virulence des streptocoques pathogènes.**

Dans le cas des streptocoques pathogènes, la longueur des chaînes de cellules peut être un paramètre déterminant de la virulence. La réussite de la colonisation d'un hôte par les streptocoques pathogènes peut être attribuée à trois facteurs essentiels : (i) la capacité de ces microorganismes à adhérer à la surface des cellules et/ou des tissus de l'hôte ; (ii) à leur capacité à envahir les cellules et à utiliser rapidement les éléments nutritifs sous la fluctuation des conditions environnementales ; et (iii) leur capacité à survivre au sein des cellules et/ou tissus hôtes par échappement au système immunitaire et de résister à l'action d'éventuels antibiotiques (Jenkinson, 1997).

Si la longueur des chaînes de cellules n'est pas un facteur de virulence, à proprement dit, elle y contribue largement. En effet, une corrélation entre la taille des chaînes de cellules et l'efficacité d'adhésion à une surface a été démontrée pour *S. mutans*, responsable des caries dentaires (Nakano et al., 2005). L'adhésion de *S. mutans* sur les dents est une étape cruciale de l'infection qui est fortement diminuée chez une mutant inactivé du gène *brpA* présentant un phénotype « longue chaîne de cellules » (Nakano et al., 2005). Néanmoins, Nakano et al., (2005) ont rapporté que dans certains cas, *S. mutans* serait capable de se propager dans le système sanguin, et les longues chaînes de cellules se fixeraient sur les cellules cardiaques provoquant ainsi une endocardite. Ceci pourrait être expliquer par le fait que les longues chaînes de cellules survivent plus longtemps dans le système sanguin et sont plus résistantes à la phagocytose (Nakano et al., 2005).

Ainsi, la croissance en longue chaîne de cellules aurait un effet négatif sur l'adhésion de *S. mutans* aux surfaces empêchant ainsi leur colonisation et un effet positif sur sa dissémination au sein des tissus hôtes. Par conséquent, il n'est pas exclu qu'au cours du cycle d'infection des bactéries pathogènes, il y ait une régulation de la taille des chaînes de cellules influencée par le stade de virulence où les courtes chaînes de cellules permettraient de réaliser les premières étapes de l'infection, tandis que les longues chaînes de cellules poursuivraient l'infection.



Toutefois, le phénotype « longue chaîne de cellules » est également observé pour les espèces non pathogènes telle que *S. thermophilus*, mais quel peut-être l'avantage adaptatif d'une croissance des cellules en chaîne pour l'espèce *S. thermophilus* ?

## **2.2 Causes et conséquences de la longueur des chaînes de cellules sur la croissance de *S. thermophilus*.**

L'espèce *S. thermophilus*, cultivée en co-culture avec *Lb. bulgaricus* ou *L. lactis*, ou encore dans un flore plus complexe, trouve toutes les conditions nécessaires à sa croissance dans le milieu lait. Le fait que la croissance des cellules en chaîne de *S. thermophilus* soit régulée par le gène *cse* et qu'il soit présent chez toutes les souches de *S. thermophilus* suggère que son maintien confère un avantage adaptatif à cette espèce. Ainsi, le phénotype du mutant  $\Delta cse$  où les chaînes de cellules sont extrêmement longues est contre sélectionné.

### ***a. Le polymorphisme intraspécifique de la longueur des chaînes de cellules pourrait être corrélé à un polymorphisme génétique.***

L'analyse du nombre de cellules par chaînes de différentes souches de *S. thermophilus* a révélé que la longueur des chaînes peut être très variable d'une souche à l'autre, quelle que soit l'origine de ces souches : lait fermenté, yaourt ou fromage. Ainsi, se pose la question des paramètres contribuant au polymorphisme intraspécifique de la taille des chaînes de cellules au sein de l'espèce *S. thermophilus* se pose. Le polymorphisme intraspécifique de la taille des chaînes de cellules pourrait être corrélée à un polymorphisme génétique. L'expression génétique de *cse* pourrait varier d'une souche à l'autre et par conséquent aurait des incidences sur la longueur des chaînes de cellules. Il est aussi envisagé que l'efficacité de *Cse* soit différente d'une souche à l'autre. En perspective, il sera donc indispensable d'une part d'étudier l'expression de ce gène dans différents contextes génétiques (par l'étude des allèles *cse* de plusieurs souches) par la technique de RT-PCR en temps réel. Et d'autre part, de tester l'activité de *Cse* sur chacune des souches. Une hypothèse alternative, serait que le nombre de gènes codant les hydrolases du peptidoglycane diffère d'une souche à l'autre. Ainsi, les gènes candidats potentiels pourraient être recherchés dans chacune des souches de *S. thermophilus*, et leur impact sur la valeur adaptative de *S. thermophilus* pourrait être déterminé par la construction de mutants.

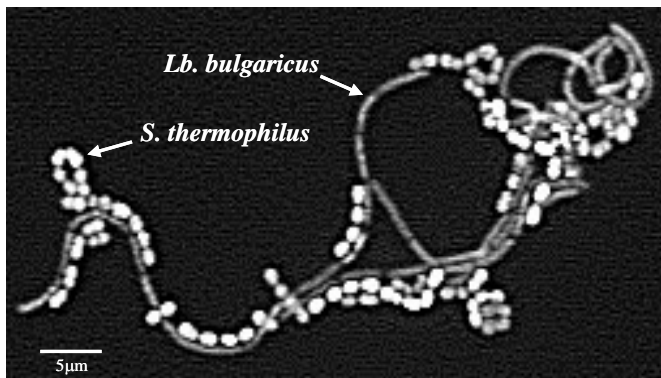
### ***b. La longueur des chaînes de cellules de *S. thermophilus* pourrait être une réponse adaptative aux stress.***

Il n'est pas exclu que comme pour *S. mutans*, la croissance des cellules en chaîne de *S. thermophilus* soit influencée par les paramètres environnementaux. Au cours des procédés industriels de fabrication des yaourts et de fromages, les fluctuations de l'environnement de *S. thermophilus* sont importantes. Ainsi, la modification du pH, les variations de température, les fortes pressions osmotiques, ou encore la modification de l'environnement par la précipitation des protéines du lait pourraient avoir des répercussions sur la taille des chaînes de cellules de cette espèce. L'existence d'une corrélation entre la modification du pH et la longueur des chaînes de cellules pour *S. mutans* et *S. sobrinus* (Nakano et al., 2005 ; Thibodeau et Ford, 1991) est un argument en faveur de cette hypothèse. Si tel est le cas, il conviendra de déterminer parmi ces paramètres lesquels peuvent

influer sur la variabilité de la longueur des chaînes de cellules de cette bactérie en soumettant *S. thermophilus* à ces diverses conditions de stress.

***c. La croissance des cellules en chaîne de *S. thermophilus* pourrait faciliter les relations intracellulaires avec ses partenaires dans la production des produits laitiers.***

Par ailleurs, *S. thermophilus* et *Lb. bulgaricus* sont connus pour leur protocoopération dans la fabrication des yaourts lors de la fermentation du lait. Cette coopération se traduit par un effet bénéfique mutuel entre ces deux espèces, influençant d'une part leur croissance et d'autre part, leur production d'acide lactique et d'arômes (Garault et al., 2000). De plus, il est probable que les co-cultures de ces deux bactéries favoriseraient les échanges génétiques constatés entre elles (Bolotin et al., 2004). Ainsi, il est possible que la communication intracellulaire entre *S. thermophilus* et *Lb. bulgaricus* soit favorisée par la croissance des cellules en chaîne de *S. thermophilus* (Figure 52).



**Figure 52 : Protocoopération entre *S. thermophilus* et *Lb. bulgaricus* (Bolotin et al., 2004).**

Adhésion de *S. thermophilus* avec *Lb. bulgaricus*. Les 2 espèces ont été cultivées ensemble en milieu liquide à la mi-phase exponentielle et observées au microscope optique.

***d. Lien entre la croissance des cellules en chaîne et le caractère probiotique de *S. thermophilus*.***

*S. thermophilus* est une bactérie lactique étudiée pour son caractère probiotique. Il a été établi que les produits laitiers, en particulier les yaourts contenant les bactéries vivantes *S. thermophilus* et *Lb. bulgaricus*, ont des effets bénéfiques sur la santé humaine (Guarner et al., 2005). Notamment, *S. thermophilus* qui présenterait des bienfaits au niveau de la tolérance au lactose, réduisant les symptômes gastro-intestinaux (Guarner et al., 2005). Cependant, le mode d'action de *S. thermophilus* dans le tractus digestif reste méconnu. De plus, ces effets ne sont que transitoires, puisque cette espèce est incapable de coloniser durablement cet environnement. Ainsi, il est probable qu'une amélioration des capacités d'adhésion de *S. thermophilus* dans le tractus digestif augmenterait son action bénéfique sur la santé humaine. Or, la taille des chaînes de cellules serait un facteur important dans le pouvoir adhésif. Par conséquent, il serait intéressant de savoir si la longueur des chaînes de cellules peut influencer sur l'aptitude de *S. thermophilus* à adhérer à la surface des cellules intestinales. En perspective, une étude de la variabilité intraspécifique de la taille des chaînes de cellules de *S. thermophilus* à adhérer sur différentes surfaces abiotiques et biotiques pourrait être réalisée. De même que, l'implantation de souches ayant des chaînes de cellules de tailles différentes pourrait être testée dans un modèle animal adéquat.

A black and white photograph of a string of light-colored, oval-shaped beads or pearls. The beads are arranged in a slightly curved line, extending from the bottom left towards the top right. The background is a dark, textured surface, possibly a fabric or paper with a fine, repeating pattern. The lighting is soft, highlighting the smooth, rounded surfaces of the beads.

*Conclusion générale*



## **Conclusion générale**

---

Cette étude a permis d'élargir notre compréhension sur le processus de séparation cellulaire, encore peu connu chez les streptocoques. L'étude approfondie de *Cse* de *S. thermophilus* a révélé que la séparation cellulaire est un processus actif génétiquement contrôlé. Les gènes impliqués dans ce processus codent tous, des hydrolases du peptidoglycane ayant pour action de rompre le lien de peptidoglycane reliant les deux cellules filles. Les enzymes de séparation cellulaire sont caractérisées d'une part, par le phénotype « longue chaîne de cellules » lors de leur inactivation sans en affecter la morphologie cellulaire et d'autre part, par leur localisation spécifique aux pôles et/ou septa cellulaires. Ces enzymes sont également très diverses et variables d'une espèce à l'autre. Ainsi, cette diversité de spécificité de clivage du peptidoglycane pourrait suggérer que la composition de la paroi cellulaire au niveau des sites de séparation (pôles et septa cellulaires) soit différente d'une espèce à l'autre. Et enfin, les enzymes de séparation cellulaire sont spécifiques d'espèces, tout au moins au sein du phylum des *Firmicutes*. Ainsi, la spécificité de ces enzymes pourrait permettre de contrôler la longueur des chaînes de cellules, qui est un critère propre à l'espèce, sans pourtant interférer avec celles des autres bactéries de la même niche écologique.

La nature des pressions de sélection s'exerçant sur la diversité et la spécificité des enzymes de séparation cellulaire reste à déterminer. Néanmoins, il est établi que la variabilité des composants de la paroi cellulaire peut permettre aux bactéries d'échapper aux systèmes immunitaires et à d'autres agressions environnementales telles que les antibiotiques. Ainsi, la sélection naturelle favoriserait la diversification de la paroi bactérienne. Chez les bactéries, la diversification du peptidoglycane s'accompagne d'une diversification des enzymes de séparation cellulaire et pose les questions de savoir : Est-ce que les modifications structurales du peptidoglycane sont à l'origine de la diversification des enzymes de séparation cellulaire ou à l'inverse, est-ce que ce sont les modifications de ces enzymes qui entraînent la diversification du peptidoglycane ? Ou encore, sur quel acteur du métabolisme du peptidoglycane, les pressions de sélection s'exercent-elles ?

Par ailleurs, cette étude ainsi que les travaux réalisés précédemment en collaboration avec F. Borges sur le caractère chimérique du gène *cse*, a permis de proposer que le succès évolutif de création de nouvelles protéines fonctionnelles par réassortiment de modules pourrait résulter d'une part de la fonctionnalité des modules indépendamment actifs, et d'autre part, de la présence de régions de faible complexité situées à l'extérieur des modules, définies en tant que siège des bornes du réassortiment de modules.

Finalement, le gène *cse* constitue un bon exemple de gène chimérique créé par réassortiment de modules ayant conduit à l'émergence d'une protéine de fonction nouvelle chez *S. thermophilus*.





*Références bibliographiques*





## Références bibliographiques

---

- Achari, A., Hale, S.P., Howard, A.J., Clore, G.M., Gronenborn, A.M., Hardman, K.D., and Whitlow, M. (1992) 1.67-Å X-ray structure of the B2 immunoglobulin-binding domain of streptococcal protein G and comparison to the NMR structure of the B1 domain. *Biochemistry* **31**: 10449-10457.
- Achaz, G., Rocha, E.P., Netter, P., and Coissac, E. (2002) Origin and fate of repeats in bacteria. *Nucleic Acids Res* **30**: 2987-2994.
- Anantharaman, V., and Aravind, L. (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* **4**: R11.
- Atrih, A., Bacher, G., Allmaier, G., Williamson, M.P., and Foster, S.J. (1999) Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *J Bacteriol* **181**: 3956-3966.
- Aussel, L., Barre, F.X., Aroyo, M., Stasiak, A., Stasiak, A.Z., and Sherratt, D. (2002) FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* **108**: 195-205.
- Baba, T., and Schneewind, O. (1998) Targeting of muralytic enzymes to the cell division site of Gram-positive bacteria: repeat domains direct autolysin to the equatorial surface ring of *Staphylococcus aureus*. *Embo J* **17**: 4639-4646.
- Baker, J.R., Liu, C., Dong, S., and Pritchard, D.G. (2006) Endopeptidase and glycosidase activities of the bacteriophage B30 lysin. *Appl Environ Microbiol* **72**: 6825-6828.
- Bateman, A., and Bycroft, M. (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* **299**: 1113-1119.
- Bateman, A., and Rawlings, N.D. (2003) The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem Sci* **28**: 234-237.
- Bellanger, X. (2007) Mise en évidence du transfert conjugatif des îlots génomiques ICES<sub>1</sub> et ICES<sub>3</sub> de *Streptococcus thermophilus*. 15<sup>ème</sup> Colloque du Club des Bactéries Lactiques, Rennes.
- Bernhardt, T.G., and de Boer, P.A. (2003) The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol Microbiol* **48**: 1171-1182.
- Biswas, R., Voggu, L., Simon, U.K., Hentschel, P., Thumm, G., and Gotz, F. (2006) Activity of the major staphylococcal autolysin Atl. *FEMS Microbiol. Lett.* **259**: 260-268.
- Bjorck, L., Kastern, W., Lindahl, G., and Wideback, K. (1987) Streptococcal protein G, expressed by streptococci or by *Escherichia coli*, has separate binding sites for human albumin and IgG. *Mol Immunol* **24**: 1113-1122.
- Bollinger, J.M., Jr., Kwon, D.S., Huisman, G.W., Kolter, R., and Walsh, C.T. (1995) Glutathionylspermidine metabolism in *Escherichia coli*. Purification, cloning, overproduction, and characterization of a bifunctional glutathionylspermidine synthetase/amidase. *J Biol Chem* **270**: 14031-14041.
- Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S.D., and Sorokin, A. (2001) The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**: 731-753.

- Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S.D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G.D., Fonstein, M., Overbeek, R., Kyprides, N., Purnelle, B., Prozzi, D., Ngui, K., Masuy, D., Hancy, F., Burteau, S., Boutry, M., Delcour, J., Goffeau, A., and Hols, P. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**: 1554-1558.
- Borges, F. (2005) 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*. In *Laboratoire de Génétique et Microbiologie, UMR UHP/INRA 1128, IFR 110 Vandoeuvre-lès-Nancy*: Université Henri Poincaré, Nancy 1.
- Borges, F., Layec, S., Thibessard, A., Fernandez, A., Gintz, B., Hols, P., Decaris, B., and Leblond-Bourget, N. (2005) *cse*, a Chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J Bacteriol* **187**: 2737-2746.
- Borges, F., Layec, S., Fernandez, A., Decaris, B., and Leblond-Bourget, N. (2006) High genetic variability of the *Streptococcus thermophilus cse* central part, a repeat rich region required for full cell segregation activity. *Antonie Van Leeuwenhoek* **90**: 245-255.
- Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B., and Guedon, G. (1999) Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* **233**: 151-161.
- Broadbent, J.R., McMahon, D.J., Welker, D.L., Oberg, C.J., and Moineau, S. (2003) Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J Dairy Sci* **86**: 407-423.
- Brodeur, B.R., Boyer, M., Charlebois, I., Hamel, J., Couture, F., Rioux, C.R., and Martin, D. (2000) Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. *Infect Immun* **68**: 5610-5618.
- Buddelmeijer, N., and Beckwith, J. (2002) Assembly of cell division proteins at the *E. coli* cell center. *Curr Opin Microbiol* **5**: 553-557.
- Buist, G., Kok, J., Leenhouts, K.J., Dabrowska, M., Venema, G., and Haandrikman, A.J. (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.
- Buist, G., Steen, A., Kok, J., and Kuipers, O.P. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* **68**: 838-847.
- Burrus, V., Roussel, Y., Decaris, B., and Guedon, G. (2000) Characterization of a novel integrative element, ICES<sub>t1</sub>, in the lactic acid bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* **66**: 1749-1753.
- Burrus, V., Bontemps, C., Decaris, B., and Guedon, G. (2001) Characterization of a novel type II restriction-modification system, Sth368I, encoded by the integrative element ICES<sub>t1</sub> of *Streptococcus thermophilus* CNRZ368. *Appl Environ Microbiol* **67**: 1522-1528.
- Carlomagno, M.S., Chiariotti, L., Alifano, P., Nappo, A.G., and Bruni, C.B. (1988) Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. *J Mol Biol* **203**: 585-606.
- Carlsson, F., Stalhammar-Carlemalm, M., Flardh, K., Sandin, C., Carlemalm, E., and Lindahl, G. (2006) Signal sequence directs localized secretion of bacterial surface proteins. *Nature* **442**: 943-946.
- Carroll, S.A., Hain, T., Technow, U., Darji, A., Pashalidis, P., Joseph, S.W., and Chakraborty, T. (2003) Identification and characterization of a peptidoglycan hydrolase, MurA, of *Listeria monocytogenes*, a muramidase needed for cell separation. *J Bacteriol* **185**: 6801-6808.

- Chi, F., Nolte, O., Bergmann, C., Ip, M., and Hakenbeck, R. (2007) Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. *Int J Med Microbiol* **297**: 503-512.
- Chia, J.S., Chang, L.Y., Shun, C.T., Chang, Y.Y., Tsay, Y.G., and Chen, J.Y. (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.
- Clewell, D.B., Jaworski, D.D., Flannagan, S.E., Zitzow, L.A., and Su, Y.A. (1995) The conjugative transposon Tn916 of *Enterococcus faecalis*: structural analysis and some key factors involved in movement. *Dev Biol Stand* **85**: 11-17.
- Cramer, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**: 288-291.
- Croux, C., Ronda, C., Lopez, R., and Garcia, J.L. (1993a) Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Mol Microbiol* **9**: 1019-1025.
- Croux, C., Ronda, C., Lopez, R., and Garcia, J.L. (1993b) Role of the C-terminal domain of the lysozyme of *Clostridium acetobutylicum* ATCC 824 in a chimeric pneumococcal-clostridial cell wall lytic enzyme. *FEBS Lett* **336**: 111-114.
- de Chateau, M., and Bjorck, L. (1994) Protein PAB, a mosaic albumin-binding bacterial protein representing the first contemporary example of module shuffling. *J Biol Chem* **269**: 12147-12151.
- de Chateau, M., and Bjorck, L. (1996) Identification of interdomain sequences promoting the intronless evolution of a bacterial protein family. *Proc Natl Acad Sci U S A* **93**: 8490-8495.
- De Las Rivas, B., Garcia, J.L., Lopez, R., and Garcia, P. (2002) Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* **184**: 4988-5000.
- Delcour, J., Ferain, T., and Hols, P. (2000) Advances in the genetics of thermophilic lactic acid bacteria. *Curr Opin Biotechnol* **11**: 497-504.
- Delorme, C. (2007) Transfert de gènes dans le groupe *salivarius* des streptocoques. Biodiversité et plasticité des génomes microbiens, Journée d'animation scientifique MICALIS INRA Jouy-en-Josas.
- Delorme, C. (2007) Safety assessment of dairy microorganisms: *Streptococcus thermophilus*. *Int J Food Microbiol*.
- Delorme, C., Poyart, C., Ehrlich, S.D., and Renault, P. (2007) Extent of horizontal gene transfer in evolution of *Streptococci* of the *salivarius* group. *J Bacteriol* **189**: 1330-1341.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**: 1390-1400.
- Diaz, E., Lopez, R., and Garcia, J.L. (1990) Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. *Proc Natl Acad Sci U S A* **87**: 8125-8129.
- Diaz, E., Lopez, R., and Garcia, J.L. (1991) Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J Biol Chem* **266**: 5464-5471.
- Donovan, D.M., Dong, S., Garrett, W., Rousseau, G.M., Moineau, S., and Pritchard, D.G. (2006a) Peptidoglycan hydrolase fusions maintain their parental specificities. *Appl Environ Microbiol* **72**: 2988-2996.

- Donovan, D.M., Foster-Frey, J., Dong, S., Rousseau, G.M., Moineau, S., and Pritchard, D.G. (2006b) The cell lysis activity of the *Streptococcus agalactiae* bacteriophage B30 endolysin relies on the cysteine, histidine-dependent amidohydrolase/peptidase domain. *Appl Environ Microbiol* **72**: 5108-5112.
- Dowson, C.G., Coffey, T.J., Kell, C., and Whiley, R.A. (1993) Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. *Mol Microbiol* **9**: 635-643.
- Dramsi, S., Dehoux, P., Lebrun, M., Goossens, P.L., and Cossart, P. (1997) Identification of four new members of the internalin multigene family of *Listeria monocytogenes* EGD. *Infect Immun* **65**: 1615-1625.
- Dybvig, K., Sitaraman, R., and French, C.T. (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**: 13923-13928.
- Eckert, C., Lecerf, M., Dubost, L., Arthur, M., and Mesnage, S. (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. *J Bacteriol* **188**: 8513-8519.
- Eriksen, T.A., Kadziola, A., and Larsen, S. (2002) Binding of cations in *Bacillus subtilis* phosphoribosylidiphosphate synthetase and their role in catalysis. *Protein Sci* **11**: 271-279.
- Errington, J. (2003) Dynamic proteins and a cytoskeleton in bacteria. *Nat Cell Biol* **5**: 175-178.
- Facklam, R. (2002) What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**: 613-630.
- Fadda, D., Santona, A., D'Ulisse, V., Ghelardini, P., Ennas, M.G., Whalen, M.B., and Massidda, O. (2007) *Streptococcus pneumoniae* DivIVA: localization and interactions in a MinCD-free context. *J Bacteriol* **189**: 1288-1298.
- Fani, R., Lio, P., and Lazcano, A. (1995) Molecular evolution of the histidine biosynthetic pathway. *J Mol Evol* **41**: 760-774.
- Fani, R., Brilli, M., Fondi, M., and Lio, P. (2007) The role of gene fusions in the evolution of metabolic pathways: the histidine biosynthesis case. *BMC Evol Biol* **7 Suppl 2**: S4.
- Farrow, J.A., and Collins, M.D. (1984) DNA base composition, DNA-DNA homology and long-chain fatty acid studies on *streptococcus thermophilus* and *Streptococcus salivarius*. *J Gen Microbiol* **130**: 357-362.
- Fernandez-Espla, M.D., Garault, P., Monnet, V., and Rul, F. (2000) *Streptococcus thermophilus* cell wall-anchored proteinase: release, purification, and biochemical and genetic characterization. *Appl Environ Microbiol* **66**: 4772-4778.
- Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R. (2001) Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**: 4658-4663.
- Filer, D., and Furano, A.V. (1981) Duplication of the *tuf* gene, which encodes peptide chain elongation factor Tu, is widespread in Gram-negative bacteria. *J Bacteriol* **148**: 1006-1011.
- Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res*. **34**: D247-251.
- Fischetti, V.A. (1989) Streptococcal M protein: molecular design and biological behavior. *Clin Microbiol Rev* **2**: 285-314.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.

- Foucaud-Scheunemann, C., and Poquet, I. (2003) HtrA is a key factor in the response to specific stress conditions in *Lactococcus lactis*. *FEMS Microbiol Lett* **224**: 53-59.
- Fox, P.F. (1993) *Cheese: Chemistry, Physics and Microbiology*. London: Chapman and Hall.
- Fujiwara, T., Aoki, S., Komatsuzawa, H., Nishida, T., Ohara, M., Suginaka, H., and Sugai, M. (2005) Mutation analysis of the histidine residues in the glycylglycine endopeptidase ALE-1. *J. Bacteriol.* **187**: 480-487.
- Fukushima, T., Afkham, A., Kurosawa, S., Tanabe, T., Yamamoto, H., and Sekiguchi, J. (2006) A new D,L-endopeptidase gene product, YojL (renamed CwlS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis*. *J Bacteriol* **188**: 5541-5550.
- Fukushima, T., Kitajima, T., Yamaguchi, H., Ouyang, Q., Furuhashi, K., Yamamoto, H., Shida, T., and Sekiguchi, J. (2008) Identification and characterization of novel cell wall hydrolase CwlT: a two-domain autolysin exhibiting N-acetylmuramidase and DL-endopeptidase activities. *J Biol Chem* **283**: 11117-11125.
- Garault, P., Letort, C., Juillard, V., and Monnet, V. (2000) Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Appl Environ Microbiol* **66**: 5128-5133.
- Garcia, P., Garcia, J.L., Garcia, E., Sanchez-Puelles, J.M., and Lopez, R. (1990) Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* **86**: 81-88.
- Garcia, J.L., Diaz, E., Romero, A., and Garcia, P. (1994) Carboxy-terminal deletion analysis of the major pneumococcal autolysin. *J. Bacteriol.* **176**: 4066-4072.
- Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R., and Garcia, J.L. (1999a) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol* **31**: 1275-1281.
- Garcia, P., Paz Gonzalez, M., Garcia, E., Garcia, J.L., and Lopez, R. (1999b) The molecular characterization of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Mol Microbiol* **33**: 128-138.
- Gerard, P., Vernet, T., and Zapun, A. (2002) Membrane topology of the *Streptococcus pneumoniae* FtsW division protein. *J Bacteriol* **184**: 1925-1931.
- Ghuysen, J.M. (1968) Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriol. Rev.* **32**: 425-464.
- Giesbrecht, P., Kersten, T., Maidhof, H., and Wecke, J. (1998) Staphylococcal cell wall: morphogenesis and fatal variations in the presence of penicillin. *Microbiol Mol Biol Rev* **62**: 1371-1414.
- Goehring, N.W., and Beckwith, J. (2005) Diverse paths to midcell: assembly of the bacterial cell division machinery. *Curr Biol* **15**: R514-526.
- Goffin, C., and Ghuysen, J.M. (1998) Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. *Microbiol Mol Biol Rev* **62**: 1079-1093.
- Guarner, F., Perdigon, G., Corthier, G., Salminen, S., Koletzko, B., and Morelli, L. (2005) Should yoghurt cultures be considered probiotic? *British Journal of Nutrition* **93**: 783-786.
- Guedon, G., Bourgoin, F., Pebay, M., Roussel, Y., Colmin, C., Simonet, J.M., and Decaris, B. (1995) Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergeneric transfer of insertion sequences occur in lactic acid bacteria co-cultures? *Mol Microbiol* **16**: 69-78.
- Heidrich, C., Templin, M.F., Ursinus, A., Merdanovic, M., Berger, J., Schwarz, H., de Pedro, M.A., and Holtje, J.V. (2001) Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**: 167-178.

- Heidrich, C., Ursinus, A., Berger, J., Schwarz, H., and Holtje, J.V. (2002) Effects of multiple deletions of murein hydrolases on viability, septum cleavage, and sensitivity to large toxic molecules in *Escherichia coli*. *J Bacteriol* **184**: 6093-6099.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., and Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* **24**: 4420-4449.
- Hollingshead, S.K., Fischetti, V.A., and Scott, J.R. (1987) Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* **207**: 196-203.
- Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Dusko Ehrlich, S., Guedon, E., Monnet, V., Renault, P., and Kleerebezem, M. (2005) New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol Rev* **29**: 435-463.
- Holtje, J.V. (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol Mol Biol Rev* **62**: 181-203.
- Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* **190**: 1401-1412.
- Huard, C., Miranda, G., Wessner, F., Bolotin, A., Hansen, J., Foster, S.J., and Chapot-Chartier, M.P. (2003) Characterization of AcmB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*. *Microbiology* **149**: 695-705.
- Huard, C., Miranda, G., Redko, Y., Wessner, F., Foster, S.J., and Chapot-Chartier, M.P. (2004) Analysis of the peptidoglycan hydrolase complement of *Lactococcus lactis*: identification of a third N-acetylglucosaminidase, AcmC. *Appl Environ Microbiol* **70**: 3493-3499.
- Husson-Kao, C., Mengaud, J., Benbadis, L., and Chapot-Chartier, M.P. (2000) Mur1, a *Streptococcus thermophilus* peptidoglycan hydrolase devoid of a specific cell wall binding domain. *FEMS Microbiol Lett* **187**: 69-76.
- Idigoras, P., Valiente, A., Iglesias, L., Trieu-Cout, P., and Poyart, C. (2001) Meningitis due to *Streptococcus salivarius*. *J Clin Microbiol* **39**: 3017.
- Ishikawa, S., Hara, Y., Ohnishi, R., and Sekiguchi, J. (1998) Regulation of a new cell wall hydrolase gene, cw1F, which affects cell separation in *Bacillus subtilis*. *J. Bacteriol.* **180**: 2549-2555.
- Ishino, F., Park, W., Tomioka, S., Tamaki, S., Takase, I., Kunugita, K., Matsuzawa, H., Asoh, S., Ohta, T., Spratt, B.G., 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**: 7024-7031.
- Jahreis, K., Bentler, L., Bockmann, J., Hans, S., Meyer, A., Siepelmeyer, J., and Lengeler, J.W. (2002) Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. *J Bacteriol* **184**: 5307-5316.
- Jenkinson, H.F., and Lamont, R.J. (1997) Streptococcal adhesion and colonization. *Crit Rev Oral Biol Med* **8**: 175-200.
- Joris, B., Englebort, S., Chu, C.P., Kariyama, R., Daneo-Moore, L., Shockman, G.D., and Ghuysen, J.M. (1992) Modular design of the *Enterococcus hirae* muramidase-2 and *Streptococcus faecalis* autolysin. *FEMS Microbiol Lett* **70**: 257-264.

- Kajimura, J., Fujiwara, T., Yamada, S., Suzawa, Y., Nishida, T., Oyamada, Y., Hayashi, I., Yamagishi, J., Komatsuzawa, H., and Sugai, M. (2005) Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of *Staphylococcus aureus*. *Mol Microbiol* **58**: 1087-1101.
- Kawamura, Y., Hou, X.G., Sultana, F., Miura, H., and Ezaki, T. (1995) Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int J Syst Bacteriol* **45**: 406-408.
- Kehoe, M.A. (1994) Cell-wall-associated proteins in Gram-positive bacteria. In *Bacterial Cell Wall*. Hakenbeck, J.-M.G.a.R. (ed): Elsevier Science B.V, pp. 217-261.
- Kohler, S., Bubert, A., Vogel, M., and Goebel, W. (1991) Expression of the *iap* gene coding for protein p60 of *Listeria monocytogenes* is controlled on the posttranscriptional level. *J. Bacteriol.* **173**: 4668-4674.
- Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* **39**: 309-338.
- Lachica, R.V., Genigeorgis, C., and Hoeprich, P.D. (1971) Metachromatic agar-diffusion methods for detecting staphylococcal nuclease activity. *Appl Microbiol* **21**: 585-587.
- Lara, B., Rico, A.I., Petruzzelli, S., Santona, A., Dumas, J., Biton, J., Vicente, M., Mingorance, J., and Massidda, O. (2005) Cell division in cocci: localization and properties of the *Streptococcus pneumoniae* FtsA protein. *Mol Microbiol* **55**: 699-711.
- Lawrence, J.G., and Roth, J.R. (1999) Genomic flux: genome evolution by gene loss and acquisition. In *In organisation of the prokaryotic genome*. Charlebois, R.L. (ed). Washington, D.C.: American Society for Microbiology, pp. 263-288.
- Layec, S., Decaris, B., and Leblond-Bourget, N. (2008a) Characterization of proteins belonging to the CHAP-related superfamily within the Firmicutes. *J Mol Microbiol Biotechnol* **14**: 31-40.
- Layec, S., Decaris, B., and Leblond-Bourget, N. (2008b) Diversity of Firmicutes peptidoglycan hydrolases and specificities of those involved in daughter cell separation. *Res Microbiol* (in press).
- Le Gouellec, A., Roux, L., Fadda, D., Massidda, O., Vernet, T., and Zapun, A. (2008) Roles of pneumococcal DivIB in cell division. *J Bacteriol* **190**: 4501-4511.
- Le Loir, Y., Gruss, A., Ehrlich, S.D., and Langella, P. (1994) Direct screening of recombinants in gram-positive bacteria using the secreted staphylococcal nuclease as a reporter. *J Bacteriol* **176**: 5135-5139.
- Lefebure, T., and Stanhope, M.J. (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol* **8**: R71.
- Lesic, B., and Carniel, E. (2005) Horizontal transfer of the high-pathogenicity island of *Yersinia pseudotuberculosis*. *J Bacteriol* **187**: 3352-3358.
- Llull, D., Lopez, R., and Garcia, E. (2006) Skl, a novel choline-binding N-acetylmuramoyl-L-alanine amidase of *Streptococcus mitis* SK137 containing a CHAP domain. *FEBS Lett* **580**: 1959-1964.
- Lopez, R., Garcia, E., Garcia, P., and Garcia, J.L. (1997) The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb Drug Resist* **3**: 199-211.
- Lopez, R., Gonzalez, M.P., Garcia, E., Garcia, J.L., and Garcia, P. (2000) Biological roles of two new murein hydrolases of *Streptococcus pneumoniae* representing examples of module shuffling. *Res Microbiol* **151**: 437-443.
- Lopez, R., and Garcia, E. (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiol Rev* **28**: 553-580.

- Machata, S., Hain, T., Rohde, M., and Chakraborty, T. (2005) Simultaneous deficiency of both MurA and p60 proteins generates a rough phenotype in *Listeria monocytogenes*. *J. Bacteriol.* **187**: 8385-8394.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D.M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J.H., Diaz-Muniz, I., Dosti, B., Smeianov, V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B., and Mills, D. (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* **103**: 15611-15616.
- Margot, P., Wahlen, M., Gholamhoseinian, A., Piggot, P., and Karamata, D. (1998) The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. *J. Bacteriol.* **180**: 749-752.
- Margot, P., Pagni, M., and Karamata, D. (1999) *Bacillus subtilis* 168 gene *lytF* encodes a gamma-D-glutamate-meso-diaminopimelate murepeptidase expressed by the alternative vegetative sigma factor, sigmaD. *Microbiology* **145** : 57-65.
- Mathot, A.G., Beliard, E., and Thuault, D. (2003) *Streptococcus thermophilus* 580 produces a bacteriocin potentially suitable for inhibition of *Clostridium tyrobutyricum* in hard cheese. *J Dairy Sci* **86**: 3068-3074.
- Mattos-Graner, R.O., Porter, K.A., Smith, D.J., Hosogi, Y., and Duncan, M.J. (2006) Functional analysis of glucan binding protein B from *Streptococcus mutans*. *J Bacteriol* **188**: 3813-3825.
- McLaughlan, A.M., and Foster, S.J. (1998) Molecular characterization of an autolytic amidase of *Listeria monocytogenes* EGD. *Microbiology* **144** ( Pt 5): 1359-1367.
- Mercier, C., Domakova, E., Tremblay, J., and Kulakauskas, S. (2000) Effects of a muramidase on a mixed bacterial community. *FEMS Microbiol Lett* **187**: 47-52.
- Mercier, C., Durrieu, C., Briandet, R., Domakova, E., Tremblay, J., Buist, G., and Kulakauskas, S. (2002) Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol Microbiol* **46**: 235-243.
- Mesnager, S., and Fouet, A. (2002) Plasmid-encoded autolysin in *Bacillus anthracis*: modular structure and catalytic properties. *J Bacteriol* **184**: 331-334.
- Mesnager, S., Chau, F., Dubost, L., and Arthur, M. (2008) Role of N-acetylglucosaminidase and N-acetylmuramidase activities in *Enterococcus faecalis* peptidoglycan metabolism. *J Biol Chem* **283**: 19845-19853.
- Meyrand, M., Boughammoura, A., Courtin, P., Mezange, C., Guillot, A., and Chapot-Chartier, M.P. (2007) Peptidoglycan N-acetylglucosamine deacetylation decreases autolysis in *Lactococcus lactis*. *Microbiology* **153**: 3275-3285.
- Miclo, L., Perrin, E., Driou, A., Mellet, M., and Linden, G. (1995) Determination of the ratios of the aromatic amino acid residues by first- or second-derivative UV spectrometry for a simple characterization of peptides. *Int J Pept Protein Res* **46**: 186-192.
- Morlot, C., Zapun, A., Dideberg, O., and Vernet, T. (2003) Growth and division of *Streptococcus pneumoniae*: localization of the high molecular weight penicillin-binding proteins during the cell cycle. *Mol Microbiol* **50**: 845-855.
- Morlot, C., Noirclerc-Savoie, M., Zapun, A., Dideberg, O., and Vernet, T. (2004) The D,D-carboxypeptidase PBP3 organizes the division process of *Streptococcus pneumoniae*. *Mol. Microbiol.* **51**: 1641-1648.
- Mouw, A.R., Beachey, E.H., and Burdett, V. (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**: 676-684.



- Nakano, K., Fujita, K., Nishimura, K., Nomura, R., and Ooshima, T. (2005) Contribution of biofilm regulatory protein A of *Streptococcus mutans*, to systemic virulence. *Microbes Infect* **7**: 1246-1255.
- Navarre, W.W., Ton-That, H., Faull, K.F., and Schneewind, O. (1999) Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity. *J Biol Chem* **274**: 15847-15856.
- Nelson, D., Schuch, R., Chahales, P., Zhu, S., and Fischetti, V.A. (2006) PlyC: a multimeric bacteriophage lysin. *Proc Natl Acad Sci U S A* **103**: 10765-10770.
- Ng, W.L., Kazmierczak, K.M., and Winkler, M.E. (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.
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1-6.
- Noirclerc-Savoie, M., Morlot, C., Gerard, P., Vernet, T., and Zapun, A. (2003) Expression and purification of FtsW and RodA from *Streptococcus pneumoniae*, two membrane proteins involved in cell division and cell growth, respectively. *Protein Expr Purif* **30**: 18-25.
- Noirclerc-Savoie, M., Le Gouellec, A., Morlot, C., Dideberg, O., Vernet, T., and Zapun, A. (2005) In vitro reconstitution of a trimeric complex of DivIB, DivIC and FtsL, and their transient co-localization at the division site in *Streptococcus pneumoniae*. *Mol Microbiol* **55**: 413-424.
- Ohnishi, R., Ishikawa, S., and Sekiguchi, J. (1999) Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J Bacteriol* **181**: 3178-3184.
- Ohno, S. (1970) *Evolution by gene duplication*. Berlin: Springer-Verlag.
- Onyenwoke, R.U., Brill, J.A., Farahi, K., and Wiegel, J. (2004) Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch ( Firmicutes). *Arch Microbiol* **182**: 182-192.
- Oshida, T., Sugai, M., Komatsuzawa, H., Hong, Y.M., Suginaka, H., and Tomasz, A. (1995) A *Staphylococcus aureus* autolysin that has an N-acetylmuramoyl-L-alanine amidase domain and an endo-beta-N-acetylglucosaminidase domain: cloning, sequence analysis, and characterization. *Proc. Natl. Acad. Sci. U S A* **92**: 285-289.
- O'Sullivan, D., Twomey, D.P., Coffey, A., Hill, C., Fitzgerald, G.F., and Ross, R.P. (2000) Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. *Mol Microbiol* **36**: 866-875.
- Pace, H.C., and Brenner, C. (2001) The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**: REVIEWS0001.
- Partridge, S.M. (2000) Prosthetic valve endocarditis due to *Streptococcus vestibularis*. *J Infect* **41**: 284-285.
- Pavlovic, G., Burrus, V., Gintz, B., Decaris, B., and Guedon, G. (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICESt1-related elements from *Streptococcus thermophilus*. *Microbiology* **150**: 759-774.
- Perrin, E., Miclo, L., Driou, A., and Linden, G. (1995) Rapid determination of the ratios of three aromatic residues in peptides by reversed-phase high-performance liquid chromatography with a high-resolution photodiode-array detector. *J Chromatogr B Biomed Appl* **664**: 267-276.
- Podar, M., Eads, J.R., and Richardson, T.H. (2005) Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC Evol Biol* **5**: 42.

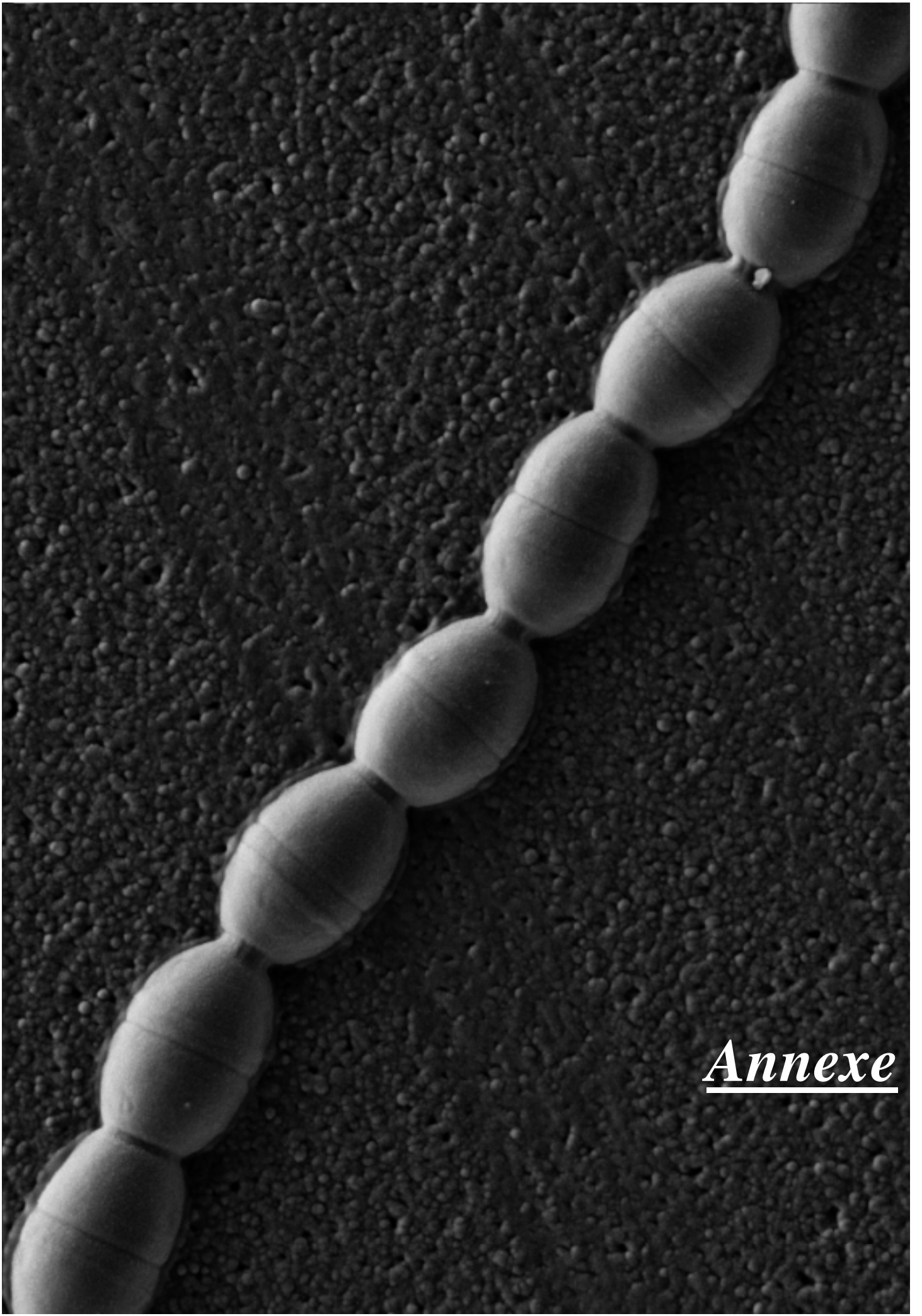
- Poquet, I., Ehrlich, S.D., and Gruss, A. (1998) An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *J Bacteriol* **180**: 1904-1912.
- Poquet, I., Saint, V., Sez nec, E., Simoes, N., Bolotin, A., and Gruss, A. (2000) HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Mol Microbiol* **35**: 1042-1051.
- Poyart, C., Quesne, G., Coulon, S., Berche, P., and Trieu-Cuot, P. (1998) Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**: 41-47.
- Pritchard, D.G., Dong, S., Baker, J.R., and Engler, J.A. (2004) The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology* **150**: 2079-2087.
- Priyadarshini, R., Popham, D.L., and Young, K.D. (2006) Daughter cell separation by penicillin-binding proteins and peptidoglycan amidases in *Escherichia coli*. *J Bacteriol* **188**: 5345-5355.
- Pucci, M.J., Thanassi, J.A., Discotto, L.F., Kessler, R.E., and Dougherty, T.J. (1997) Identification and characterization of cell wall-cell division gene clusters in pathogenic gram-positive cocci. *J Bacteriol* **179**: 5632-5635.
- Pugsley, A.P., Francetic, O., Driessen, A.J., and de Lorenzo, V. (2004) Getting out: protein traffic in prokaryotes. *Mol Microbiol* **52**: 3-11.
- Raffelsbauer, D., Bubert, A., Engelbrecht, F., Scheinplflug, J., Simm, A., Hess, J., Kaufmann, S.H., and Goebel, W. (1998) The gene cluster *inlC2DE* of *Listeria monocytogenes* contains additional new internalin genes and is important for virulence in mice. *Mol Gen Genet* **260**: 144-158.
- Rajagopal, L., Clancy, A., and Rubens, C.E. (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**: 14429-14441.
- Rallu, F., Taillez, P., Ehrlich, S.D., and Renault, P. (2002) Common scheme of evolution between *eps* clusters of the food bacteria *Streptococcus thermophilus* and *cps* clusters of pathogenic streptococci. Genetics, I.P.o.t.A.S.o.M.C.o.S. (ed.) Asheville, NC.
- Rashel, M., Uchiyama, J., Takemura, I., Hoshiba, H., Ujihara, T., Takatsuji, H., Honke, K., and Matsuzaki, S. (2008) Tail-associated structural protein gp61 of *Staphylococcus aureus* phage phi MR11 has bifunctional lytic activity. *FEMS Microbiol Lett* **284**: 9-16.
- Rasic, J.L., and Kurmann, J.A. (1978) Fermented fresh milk products. Vol. 1. Yogurt: Scientific Grounds, T., Manufacture and Preparations (ed): Technical Dairy Publishing House, Copenhagen, DK.
- Rathsam, C., and Jacques, N.A. (1998) Role of C-terminal domains in surface attachment of the fructosyltransferase of *Streptococcus salivarius* ATCC 25975. *J Bacteriol* **180**: 6400-6403.
- Redko, Y., Courtin, P., Mezange, C., Huard, C., and Chapot-Chartier, M.P. (2007) *Lactococcus lactis* gene *yjgB* encodes a gamma-D-glutaminy-L-lysyl-endopeptidase which hydrolyzes peptidoglycan. *Appl Environ Microbiol* **73**: 5825-5831.
- Reinscheid, D.J., Gottschalk, B., Schubert, A., Eikmanns, B.J., and Chhatwal, G.S. (2001) Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J Bacteriol* **183**: 1175-1183.
- Reinscheid, D.J., Ehlert, K., Chhatwal, G.S., and Eikmanns, B.J. (2003) Functional analysis of a PcsB-deficient mutant of group B streptococcus. *FEMS Microbiol Lett* **221**: 73-79.
- Reis, K.J., Ayoub, E.M., and Boyle, M.D. (1984) Streptococcal Fc receptors. I. Isolation and partial characterization of the receptor from a group C streptococcus. *J Immunol* **132**: 3091-3097.

- Renault, P., and Reitz-Ausseau, J. (2007) Qualified Presumption of Safety (QPS), un concept européen en construction. 15ème colloques du Club des Bactéries Lactiques, Rennes.
- Rigden, D.J., Jedrzejewski, M.J., and Galperin, M.Y. (2003) Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem Sci* **28**: 230-234.
- Riley, M., and Anilionis, A. (1978) Evolution of the bacterial genome. *Annu Rev Microbiol* **32**: 519-560.
- Rioux, S., Martin, D., Ackermann, H.W., Dumont, J., Hamel, J., and Brodeur, B.R. (2001) Localization of surface immunogenic protein on group B streptococcus. *Infect Immun* **69**: 5162-5165.
- Rocha, E.P., Pradillon, O., Bui, H., Sayada, C., and Denamur, E. (2002) A new family of highly variable proteins in the *Chlamydomonas reinhardtii* genome. *Nucleic Acids Res* **30**: 4351-4360.
- Rocha, E.P. (2003a) An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: from duplications to genome reduction. *Genome Res* **13**: 1123-1132.
- Rocha, E.P. (2003b) DNA repeats lead to the accelerated loss of gene order in bacteria. *Trends Genet* **19**: 600-603.
- Rogers, H.J., Perkins, H.R., and Ward, J.B. (1980) *Microbial cell walls and membranes*. London.
- Rothfield, L., Justice, S., and Garcia-Lara, J. (1999) Bacterial cell division. *Annu Rev Genet* **33**: 423-448.
- Sanchez-Puelles, J.M., Ronda, C., Garcia, J.L., Garcia, P., Lopez, R., and Garcia, E. (1986) Searching for autolysin functions. Characterization of a pneumococcal mutant deleted in the *lytA* gene. *Eur J Biochem* **158**: 289-293.
- Sanchez-Puelles, J.M., Sanz, J.M., Garcia, J.L., and Garcia, E. (1990) Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. *Gene* **89**: 69-75.
- Sanz, J.M., Garcia, P., and Garcia, J.L. (1996) Construction of a multifunctional pneumococcal murein hydrolase by module assembly. *Eur J Biochem* **235**: 601-605.
- Sass, P., and Bierbaum, G. (2007) Lytic activity of recombinant bacteriophage phi11 and phi12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. *Appl Environ Microbiol* **73**: 347-352.
- Schleifer, K.H., and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36**: 407-477.
- Schuster, C., Dobrinski, B., and Hakenbeck, R. (1990) Unusual septum formation in Streptococcus pneumoniae mutants with an alteration in the D,D-carboxypeptidase penicillin-binding protein 3. *J Bacteriol* **172**: 6499-6505.
- Sebulsky, M.T., Hohnstein, D., Hunter, M.D., and Heinrichs, D.E. (2000) Identification and characterization of a membrane permease involved in iron-hydroxamate transport in *Staphylococcus aureus*. *J Bacteriol* **182**: 4394-4400.
- Sebulsky, M.T., Speziali, C.D., Shilton, B.H., Edgell, D.R., and Heinrichs, D.E. (2004) FhuD1, a ferric hydroxamate-binding lipoprotein in *Staphylococcus aureus*: a case of gene duplication and lateral transfer. *J Biol Chem* **279**: 53152-53159.
- Sela, S., Yogev, D., Razin, S., and Bercovier, H. (1989) Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. *J Bacteriol* **171**: 581-584.
- Sheehan, M.M., Garcia, J.L., Lopez, R., and Garcia, P. (1996) Analysis of the catalytic domain of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *FEMS Microbiol Lett* **140**: 23-28.
- Sheehan, M.M., Garcia, J.L., Lopez, R., and Garcia, P. (1997) The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Mol Microbiol* **25**: 717-725.

- Sherman, J.M. (1937) The Streptococci. *Bacteriol Rev* **1**: 3-97.
- Shibata, Y., Kawada, M., Nakano, Y., Toyoshima, K., and Yamashita, Y. (2005) Identification and characterization of an autolysin-encoding gene of *Streptococcus mutans*. *Infect Immun* **73**: 3512-3520.
- Shockman, G.D., and Barrett, J.F. (1983) Structure, function, and assembly of cell walls of gram-positive bacteria. *Annu. Rev. Microbiol.* **37**: 501-527.
- Shockman, G.D., and Holtje, J.V. (1994) Microbial peptidoglycan (murein) hydrolases. In *Bacterial cell wall*. Ghuyssen, J.M. and Hakenbeck, R. (eds). Amsterdam, The Netherlands: Elsevier Science B.V.
- Siezen, R.J., Kuipers, O.P., and de Vos, W.M. (1996) Comparison of lantibiotic gene clusters and encoded proteins. *Antonie Van Leeuwenhoek* **69**: 171-184.
- Smith, T.J., Blackman, S.A., and Foster, S.J. (1996) Peptidoglycan hydrolases of *Bacillus subtilis* 168. *Microb Drug Resist* **2**: 113-118.
- Solow, B.T., and Somkuti, G.A. (2000) Comparison of low-molecular-weight heat stress proteins encoded on plasmids in different strains of *Streptococcus thermophilus*. *Curr Microbiol* **41**: 177-181.
- Steen, A., Buist, G., Leenhouts, K.J., El Khattabi, M., Grijpstra, F., Zomer, A.L., Venema, G., Kuipers, O.P., and Kok, J. (2003) Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J Biol Chem* **278**: 23874-23881.
- Steen, A., Buist, G., Horsburgh, G.J., Venema, G., Kuipers, O.P., Foster, S.J., and Kok, J. (2005) AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. *Febs J* **272**: 2854-2868.
- Suzuki, T., and Tahara, Y. (2003) Characterization of the *Bacillus subtilis* *ywtD* gene, whose product is involved in gamma-polyglutamic acid degradation. *J Bacteriol* **185**: 2379-2382.
- Swartley, J.S., McAllister, C.F., Hajjeh, R.A., Heinrich, D.W., and Stephens, D.S. (1993) Deletions of Tn916-like transposons are implicated in tetM-mediated resistance in pathogenic *Neisseria*. *Mol Microbiol* **10**: 299-310.
- Tamime, A.Y., and Deeth, H.C. (1980) Yogurt: technology and biochemistry. *J. Food Protect* **43**: 939-977.
- Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouiri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**: 498-506.
- Thibessard, A., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2002a) Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can J Microbiol* **48**: 473-478.
- Thibessard, A., Fernandez, A., Gintz, B., Leblond-Bourget, N., and Decaris, B. (2002b) Effects of rodA and pbp2b disruption on cell morphology and oxidative stress response of *Streptococcus thermophilus* CNRZ368. *J Bacteriol* **184**: 2821-2826.
- Thibessard, A., Borges, F., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2004) Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. *Appl Environ Microbiol* **70**: 2220-2229.
- Thibodeau, E.A., and Ford, C.M. (1991) Chain formation and de-chaining in *Streptococcus sobrinus* SL-1. *Oral Microbiol Immunol* **6**: 313-315.

- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.
- Thulin, S., Olcen, P., Fredlund, H., and Unemo, M. (2006) Total variation in the *penA* gene of *Neisseria meningitidis*: correlation between susceptibility to beta-lactam antibiotics and *penA* gene heterogeneity. *Antimicrob Agents Chemother* **50**: 3317-3324.
- Tomasz, A., and Fischer, W. (2000) The cell of *Streptococcus pneumoniae*. Fischetti, V.A., Novick, R.P., Ferretti, J.J., Portnoy, D.A. and Rood, J.I. (eds). Washington, D.C: ASM press, pp. 191-2000.
- Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M., and Venter, J.C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539-547.
- Tsai, Y.H., Orsi, R.H., Nightingale, K.K., and Wiedmann, M. (2006) *Listeria monocytogenes* internalins are highly diverse and evolved by recombination and positive selection. *Infect Genet Evol* **6**: 378-389.
- van den Bogaard, P.T., Hols, P., Kuipers, O.P., Kleerebezem, M., and de Vos, W.M. (2004) Sugar utilisation and conservation of the gal-lac gene cluster in *Streptococcus thermophilus*. *Syst Appl Microbiol* **27**: 10-17.
- Vaughan, E.E., van den Bogaard, P.T., Catzeddu, P., Kuipers, O.P., and de Vos, W.M. (2001) Activation of silent gal genes in the lac-gal regulon of *Streptococcus thermophilus*. *J Bacteriol* **183**: 1184-1194.
- Vazquez-Boland, J.A., Dominguez-Bernal, G., Gonzalez-Zorn, B., Kreft, J., and Goebel, W. (2001) Pathogenicity islands and virulence evolution in *Listeria*. *Microbes Infect* **3**: 571-584.
- Veiga, P., Bulbarela-Sampieri, C., Furlan, S., Maisons, A., Chapot-Chartier, M.P., Erkelenz, M., Mervelet, P., Noirot, P., Frees, D., Kuipers, O.P., Kok, J., Gruss, A., Buist, G., and Kulakauskas, S. (2007) SpxB regulates O-acetylation-dependent resistance of *Lactococcus lactis* peptidoglycan to hydrolysis. *J Biol Chem* **282**: 19342-19354.
- Welch, R.A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., and Blattner, F.R. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 17020-17024.
- Winkler, M.E. (1978) Expression of the *histidine* operon in rho mutants of *Escherichia coli*. *J Bacteriol* **135**: 721-725.
- Wootton, J.C., and Drummond, M.H. (1989) The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng* **2**: 535-543.
- Wootton, J.C., and Federhen, S. (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* **266**: 554-571.
- Wuenschel, M.D., Kohler, S., Bubert, A., Gerike, U., and Goebel, W. (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.
- Yakota, T., Sugisaki, H., Takamami, M., and Kaziro, Y. (1980) The nucleotide sequence of the cloned *tuf* gene of *E. coli*. *Gene* **12**: 25-31.
- Yamada, S., Sugai, M., Komatsuzawa, H., Nakashima, S., Oshida, T., Matsumoto, A., and Suginaka, H. (1996) An autolysin ring associated with cell separation of *Staphylococcus aureus*. *J Bacteriol* **178**: 1565-1571.

- Yamaguchi, H., Furuhashi, K., Fukushima, T., Yamamoto, H., and Sekiguchi, J. (2004) Characterization of a new *Bacillus subtilis* peptidoglycan hydrolase gene, *yvcE* (named *cwlO*), and the enzymatic properties of its encoded protein. *J Biosci Bioeng* **98**: 174-181.
- Yamamoto, H., Kurosawa, S., and Sekiguchi, J. (2003) Localization of the vegetative cell wall hydrolases LytC, LytE, and LytF on the *Bacillus subtilis* cell surface and stability of these enzymes to cell wall-bound or extracellular proteases. *J Bacteriol* **185**: 6666-6677.
- Yokoi, K.J., Kawahigashi, N., Uchida, M., Sugahara, K., Shinohara, M., Kawasaki, K., Nakamura, S., Taketo, A., and Kodaira, K. (2005) The two-component cell lysis genes *holWMY* and *lysWMY* of the *Staphylococcus warneri* M phage varphiWMY: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. *Gene* **351**: 97-108.
- Yoshimura, G., Komatsuzawa, H., Hayashi, I., Fujiwara, T., Yamada, S., Nakano, Y., Tomita, Y., Kozai, K., and Sugai, M. (2006) Identification and molecular characterization of an N-Acetylmuraminidase, Aml, involved in *Streptococcus mutans* cell separation. *Microbiol Immunol* **50**: 729-742.



*Annexe*





## **Annexe**

---

### **Articles**

Publication n°2 :

---

**High genetic variability of the *Streptococcus thermophilus* cse central part, a repeat rich region required for full cell segregation activity.**

Frédéric Borges, Séverine Layec, Annabelle Fernandez, Bernard Decaris et Nathalie Leblond-Bourget.

**Antonie Van Leeuwenhoek.** Octobre 2006, p. 245-255, Vol. 90, No. 3

---

Publication n°3 :

---

**Characterization of proteins belonging to the CHAP-related superfamily within the Firmicutes.**

Séverine Layec, Bernard Decaris et Nathalie Leblond-Bourget.

**Journal of Molecular Microbiology and Biotechnology.** 2008, p. 31-40, Vol. 14, No. 1-3

---

Publication n°4 :

---

**Diversity of Firmicutes peptidoglycan hydrolases and specificities of those involved in daughter cell separation.**

Séverine Layec, Bernard Decaris et Nathalie Leblond-Bourget.

**Research in Microbiology.** Juillet 2008, sous presse

---



# High genetic variability of the *Streptococcus thermophilus* *cse* central part, a repeat rich region required for full cell segregation activity

Frédéric Borges · Séverine Layec ·  
Annabelle Fernandez · Bernard Decaris ·  
Nathalie Leblond-Bourget

Received: 30 January 2006 / Accepted: 11 April 2006  
© Springer Science+Business Media B.V. 2006

**Abstract** The *cse* gene of *Streptococcus thermophilus* encodes an extracytoplasmic protein involved in cell segregation. The Cse protein consists of two putative domains: a cell wall attachment LysM domain and a catalytic CHAP domain. These two domains are spaced by an interdomain linker, known as Var-Cse, previously reported to be highly divergent between two *S. thermophilus* strains. The aim of this study was to assess the extent of this intraspecific variability and the functional involvement of the *var-cse* region in cell segregation. Analysis of the *var-cse* sequence of 19 different strains allowed detection of 11 different alleles, varying from 390 bp to 543 bp, all containing interspersed and tandem nucleotides repeats. Overall, 11 different repeat units were identified and some series of these small repeats, named supermotifs, form large repeats. Results suggested that *var-cse* evolved by

deletion of all or part of the repeats and by duplication of repeats or supermotifs. Moreover, sequence analysis of the whole *cse* locus revealed that the *cse* ORF is mosaic suggesting that *var-cse* polymorphism resulted from horizontal transfer. The partial deletion of the *var-cse* region of the *S. thermophilus* strain CNRZ368 led to the lengthening of the number of cells per streptococcal chain, indicating that this region is required for full cell segregation in *S. thermophilus* strain CNRZ368.

**Keywords** Cell separation · Horizontal transfer · Mosaic · Mutation · Repeat

## Abbreviations

CR Coding Repeat  
Indel Insertion/deletion

---

F. Borges · S. Layec · A. Fernandez · B. Decaris ·  
N. Leblond-Bourget (✉)  
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  
e-mail: bourget@nancy.inra.fr

F. Borges  
Department of Microbiology and Biotechnology,  
University of Ulm, Albert-Einstein-Allee 11, D-89069  
Ulm, Germany

## Introduction

The most relevant property of repeat sequences, from an evolutionary point of view, is their genetic plasticity. For intragenic repeats, known as coding repeats (CRs), the consequences of this instability, regarding the Open Reading Frame (ORF) integrity and the encoded amino acid sequence, are different according to their size and

their respective reading phase positioning. When the repeat motif size is not a multiple of three nucleotides, genetic instability of the repeat rich region can induce, at a high frequency, protein shortening or lengthening, inactivating or activating the ORF. Usually, genetic instabilities of these kinds of repeats induce switching between activation and inactivation of the gene and is a mechanism of gene expression regulation by phase variation (Hallet 2001). When the repeat size is a multiple of three and when the repeats are positioned in the same phase, the size and the amino acid sequences are modified but ORFs are always finished by the same stop codon whatever the allele and therefore are not inactivated by frameshifts. Intragenic repeats from the last category are widely studied in pathogenic bacteria but less in non-pathogenic bacteria. Furthermore, these genetically unstable regions are mainly used as molecular tags for typing pathogenic strains, but their biological function and evolution mechanism are not often studied (van Belkum et al. 1998; van Belkum 1999).

The size of CRs can vary from one to one hundred nucleotides. Genes, and especially those encoding surface proteins, can contain numerous sets of repeats. In most cases, CRs are organized in direct tandem repeat and are mainly built with one type of motif (Fischetti et al. 1991; Dramsi et al. 1993; van Belkum et al. 1998). Exceptions, however, can be found such as the gene encoding the G protein from the GX7805 strain of group G *Streptococcus* that contains 4 different types of motifs that are organized as alternated interspersed repeats (Filpula et al. 1987). One of the most complex reported examples of CRs is found in the *cse* gene from *Streptococcus thermophilus* involved in Cell Segregation (Borges et al. 2005). It encodes an extracytoplasmic protein with one putative cell wall attachment (LysM) domain and a putative Cysteine Histidine-dependent Amidohydrolase/Peptidase (CHAP) domain. These two domain-encoding sequences are spaced by a region named *var-cse*. Analysis and comparison of the *var-cse* allele from two *S. thermophilus* strains (CNRZ368 and LMG18311) revealed that the *var-cse* sequence: (i) is almost completely built with numerous (up to 8) different types of small interspersed and tandem direct repeats, (ii)

encodes an amino acid sequence of low complexity, and (iii) is highly divergent. The sequence of the two analyzed *S. thermophilus* strains diverge from approximately 40%. However, the comparison of two alleles only provided little information concerning the origin of this variability. Moreover, the replacement of the *var-cse* region of strain LMG18311 by that of strain CNRZ368 had little or no effect on cell segregation activity of Cse (Borges et al. 2005). This suggested that this region may be dispensable for cell segregation activity of Cse.

The aim of this study was to assess the extent of *var-cse* variability and to determine if this region is required for Cse activity. For this purpose, the sequence of *var-cse* alleles from numerous strains of *S. thermophilus* were analyzed and compared. In addition, the impact of a partial deletion of *var-cse* on cell segregation was investigated.

## 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, M17 (Terzaghi and Sandine 1975) or HJL (Stingele and Mollet 1996) media, without shaking. Milk medium was used for strain storage, M17 was used for mutant generation and HJL was employed for phenotypic analyses. Phenotypic analyses were performed at 42°C, the optimal *S. thermophilus* growth temperature. Erythromycin was added at 2 µg ml<sup>-1</sup> when required. *S. thermophilus* derivative strains containing pGh9 (Maguin et al. 1992) plasmid derivatives were cultivated at 30°C when plasmid self-maintenance was required and at 42°C for selection of clones containing chromosomally integrated plasmid.

Recombinant plasmids derived from pGh9 were transformed into *Escherichia coli* EC101, a TG1 strain containing a chromosomal copy of the pWV01 *repA* gene (Buist et al. 1995), and were selected at 37°C on Luria-Bertani (LB) medium

**Table 1** Strains and plasmids used in this study

| Strain or plasmid                 | Relevant characteristic(s)   | Source or reference  |
|-----------------------------------|--|--|
| <b>Strains</b>                    |  |  |
| <i>Streptococcus thermophilus</i> |  |  |
| CNRZ368                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ385                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ445                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ702                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ308                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ71                            | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ1066                          | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ464                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ388                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ1402                          | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ302                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| CNRZ307                           | Wild-type strain   | INRA-CNRZ, strain collection   |
| NST1                              | Wild-type strain   | INRA-CNRZ, strain collection   |
| A054                              | Wild-type strain   | Laboratoire de Génétique et Microbiologie, Nancy (Mercenier et al. 1987) |
| ATCC19258                         | Wild-type strain   | American Type Culture Collection   |
| IP6757                            | Wild-type strain   | Institut Pasteur strain collection                                       |
| Sf6                               | Wild-type strain   | Nestlé strain collection   |
| LMG18311                          | Wild-type strain   | BCCM <sup>TM</sup> /LMG, strain collection                               |
| CNRZ368- $\Delta_{67-411}var-cse$ | <i>S. thermophilus</i> CNRZ368 with deletion of 345bp of <i>var-cse</i> .  | This work  |
| CNRZ368- $\Delta_{cse}$           | <i>S. thermophilus</i> CNRZ368 with $\Delta_{cse}$ mutation  | (Borges et al. 2005)   |
| <i>Escherichia coli</i>           |  |  |
| EC101                             | <i>supE hsd-5 thi</i> $\Delta$ (lac-proAB) F <sup>+</sup> ( <i>traD6 proAB<sup>+</sup> lacZ<math>\Delta</math>M15</i> ) <i>repA<sup>+</sup></i> , derivative of TG1 strain | (Leenhouts 1995)   |
| DH5 $\alpha$                      | <i>supE44 <math>\Delta</math>lacU169 (<math>\phi</math>80 lacZ <math>\Delta</math>M15) hsdR17 endA1 gyrA96 thi-1 relA1</i>   | (Sambrook et al.1989)  |
| <b>Plasmids</b>                   |  |  |
| pGh9                              | Em <sup>R</sup> , thermosensitive replication origin from pVE6002  | (Maguin et al. 1992)   |
| pGh9:: $\Delta_{67-411}var-cse$   | pGh9 with $\Delta_{1,29-411}var-cse$ and CNRZ368 surrounding regions   | This work  |

(Sambrook et al. 1989) containing 150  $\mu\text{g ml}^{-1}$  of erythromycin.

### DNA manipulations

Preparation of chromosomal and plasmid DNA, and Southern analysis were performed according to standard protocols (Sambrook et al. 1989). Sequences were aligned with CLUSTALW (Thompson et al. 1994). DOT PLOTS were constructed by using the Nucleic Acid Dot Plot from the Molecular Toolkit (<http://www.arbl-cvmb.colostate.edu/molkit/dnadot/index.html>). The *cse* region encompassing *var-cse* was amplified by PCR using the primers 5'-CTGTAGTAGCAGAATCTAAC-3' and 5'-GCACTAGCAATCCAGTCTT-3'. The reaction mixture was incubated at 95°C for 5 min. This was followed by 30 cycles of 45 s at 95°C, 45 s at 50°C, and 45 s at 72°C. After the last cycle, extension of the products was made at 72°C for 5 min. The *cse* locus was amplified by PCR using the following primers 5'-TTGTTGGCGCGCTTTAGCCGATTTGGC-TTTTGAAG-3' and 5'-TTGTTGGCGCGCCC-CATTATTTTCTCAGGATGAAT-3'. Amplifications were performed as follow: one cycle of 5 min at 95°C followed by 10 cycles of 45 s at 95°C, 45 s at 50°C and 90 s at 72°C. This was followed by 20 cycles of 45 s at 95°C, 45 s at 60°C and 90 s at 72°C. At last, the reaction mixture was incubated 5 min at 72°C. Taq polymerase (Sigma) was used for amplification. To avoid the sequencing of PCR induced mutations, four independent PCR reactions were performed in parallel for each *S. thermophilus* strain, and the products were pooled before sequencing.

### Analysis of *cse* mosaicism

A sequence identity matrix was elaborated, on the basis of sequence alignments of the *cse* locus, by using the BioEdit program (Hall 1999). One matrix was constructed for each of the A, B, D and E regions of the *cse* locus. The boundary separating the A and B regions was positioned on the 5' side of the first variable nucleotide position that separates the sequences as groups I and II. The boundary between the D and E regions was positioned similarly. The boundaries flanking the

C (or *var-cse*) region were positioned so that the highly variable repeat region is completely comprised in the C region. A mean and standard deviation were then calculated with the values resulting from the comparison of sequences that belong to the same group as well as from different groups. However, since group I was only represented by two sequences (from CNRZ368 and CNRZ445), the values presented for the group I does not correspond to a mean.

Construction of a  $\Delta_{67-411}$ *var-cse* mutant strain of *S. thermophilus*

The *var-cse* region from nucleotide 67 to 411 (with nucleotide 1 as the first 5' nucleotide of the *var-cse* region) was deleted by allelic replacement as previously described (Thibessard et al. 2004). Briefly, the two regions flanking the *cse* locus to be deleted were independently amplified by PCR using the following primers pairs: 5'-CCCCCAAGCTTTGAATGTTTTGGCTA-ATATC-3' and 5'-CCGGAATTCTTCAGTTGT TTCAGTTGC-3', 5'-CCGGAATTCCTAGCA GCTACATACGA-3' and 5'-CCCCCTGCAGTT ATGGATAAATATAATATAC-3'. The PCR products were digested by appropriate restriction enzymes, joined together and ligated with the plasmid pGh9. The ligation products were used to transform *E. coli* EC101. After introduction of the recombinant plasmid into *S. thermophilus*, deleted mutants displaying an EcoRI restriction site instead of the 345 bp of the *var-cse* region were selected as previously described (Thibessard et al. 2004). The *var-cse* partial deletion was checked by PCR, Southern hybridization and sequencing (data not shown).

### Microscopy

Cells were observed with a Nikon OPTIPHOT microscope mounted with phase contrast equipment "Ph" at  $\times 100$  magnification, using the condenser turret at position Ph4, or at magnification  $\times 1000$  by phase contrast.

### Nucleotide sequence accession numbers

DNA sequences reported in this paper have been deposited in GenBank under accession numbers

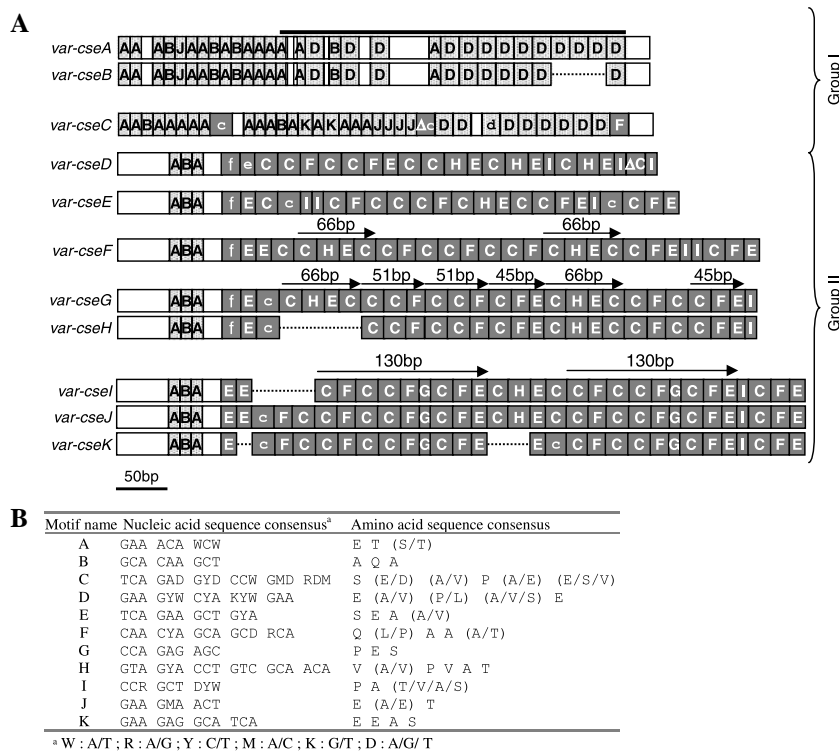
AY730642, AY695844, CP000024 and from DQ102334 to DQ102345.

**Results**

Structure of the *var-cse* alleles

In order to characterize the genetic variability of *var-cse*, the corresponding sequence of 15 *S. thermophilus* strains was determined. For this purpose,

a region of the *cse* ORF comprising *var-cse* was amplified by PCR from genomic DNA. A single PCR product was obtained for each strain and was sequenced. Since the sequence of the *cse* locus of 4 strains (CNRZ368, LMG18311, CNRZ1066 and LMD-9) was already available (Bolotin et al. 2004; Borges et al. 2005; [http://www.genome.jgi-psf.org/draft\\_microbes/strth/strth.home.html](http://www.genome.jgi-psf.org/draft_microbes/strth/strth.home.html)), a total of 19 sequences were analyzed. Among these sequences, 11 alleles (named *var-cseA* to *var-cseK*) differing by their sequence and their size were



**Fig. 1** Repetitive structure of *var-cse* alleles. **(A)** Schematic representation of the *var-cse* region of the following *S. thermophilus* strains: CNRZ368 NST1 and CNRZ385 (*var-cseA* allele); A054 (*var-cseB* allele); CNRZ445 (*var-cseC* allele); ATCC19258 (*var-cseD* allele); CNRZ702 (*var-cseE* allele); CNRZ308 (*var-cseF* allele); IP6757, CNRZ7, CNRZ1066 and Sfi6 (*var-cseG* allele); CNRZ464 (*var-cseH* allele); LMG18311 and CNRZ388 (*var-cseI* allele); LMD9 and CNRZ1402 (*var-cseJ* allele); CNRZ302 and CNRZ307 (*var-cseK* allele). Each repeat unit is represented by a letter-containing box (referring to part B). Repeat units in accordance with the consensus sequences presented in panel B are represented in uppercase letters and those that are slightly divergent from the consensus are represented by lowercase letters. The Δ symbol refers to partially deleted repeats. Repeats

that are characteristic of group I and group II are represented by light colored grey and dark grey rectangles respectively. Aligned sequences are represented by bracketing one above the other the schematic representations of the sequences aligned by CLUSTALW. Dotted lines indicate gaps in the alignments. Notice that because of the repeat richness of these regions, the positioning of gaps can not be favored between two possibilities and so representation of one of them was arbitrarily chosen. The repeated supermotifs corresponding to nucleotide duplications greater than 45 bp with less than 1% degeneracy are indicated by arrows. The region of *var-cseA* deleted in the CNRZ368-Δ67-411*var-cse* mutant strain is indicated by a thick line. **(B)** Table listing the consensus sequences of the repeated motifs identified in the *var-cse* alleles

detected. The size of the *var-cse* sequences varies from 390 to 543 bp, i.e. a maximum size variation of 153 bp was found. The repeat content was assessed for each allele. This analysis revealed that they are almost all built-up of direct repeats, with tandem and dispersed arrangement (Fig. 1A). A total of 11 different repeat motifs (named A to K), with sizes from 9 to 18 bp, were detected (Fig.1B). All of the repeated units are multiples of three and are thus in frame. Consequently, all the nucleic acid repeat sequences correspond to an amino acid repeat in Cse (Fig.1B). The sequence of 8 of these 11 motifs (A, C, D, E, F, H, I, and J) are degenerated. The most degenerated is the most frequent C motif, with 9 degenerated positions among 18.

Consensus sequence comparison of the motifs showed that some of them are similar to parts of others. For example, the nucleotide sequence encoding the PES part of the SEAPES amino acid sequence (one possible C motif) is identical to the G motif, whose corresponding amino acid sequence is also PES (Fig. 2).

Nucleotide sequence analysis by dot plot indicated that the *var-cse* alleles can contain repeats larger than 18 bp, corresponding to the size of the largest repeats described Fig. 1B. Indeed, the *var-cseF* and *var-cseG* alleles contain a perfectly duplicated 66 bp region (Fig. 1A). This region corresponds to the CHEC series of motifs, or supermotif. This supermotif is also present in the other alleles but has degenerated. Other examples are the CCF (51 bp) and CFE (45 bp) supermotifs that are perfectly duplicated in *var-cseG* and *var-cseH* (Fig. 1A). The largest duplication identified is 130 bp long, comprising the

CFCCFGCFE motif block. This region is duplicated almost perfectly (the two copies exhibits one divergent nucleotide) in *var-cseI* to *K* (Fig. 1A). In conclusion, the small repeated motifs can be organized as repeated supermotifs.

Features of *var-cse* intraspecific variability

Alignment of *var-cse* alleles was obtained using CLUSTALW. Taking into account their high divergence (nucleotide or amino acid), only the most related sequences could be aligned and are represented in Fig. 1A. Related sequences differed only by their repeat copy number. For instance, the D motif is found 13 times in *var-cseA* but only 10 times in *var-cseB*. Additionally, the number of supermotifs can also differ from one allele to the other. Indeed, *var-cseG* and *var-cseH* differ only by one copy of the CHEC supermotif.

In addition, the *var-cse* alleles can be classified in two groups on the basis of their repeat type content. The group I, bringing together *var-cseA* to *C*, comprises all of the D, J and K motifs, and the majority (74%) of the A motifs (Fig. 1A). Group II, bringing together *var-cseD* to *K*, comprises all of the E, G, H, and I motifs, and the majority (98%) of the C and F motifs.

*cse* is mosaic

Two alternative hypotheses could explain the origin of the *var-cse* alleles in these two groups. First, the common ancestor of the two corresponding groups of *S. thermophilus* strains is ancient each has since evolved by severe mutation accumulation. Alternatively, a *var-cse* allele was acquired in *S. thermophilus* by horizontal sequence transfer resulting in a mosaic *cse* locus.

To test these hypotheses, we compared the entire *cse* ORF and flanking regions from 10 *S. thermophilus* strains belonging to group I (CNRZ368, CNRZ445) and group II (LMG18311, LMD9, ATCC19258, CNRZ1066, CNRZ302, CNRZ308, CNRZ702, CNRZ1402). Sequence alignments revealed that the regions encompassing the 436 bp B region and the 228 bp D region upstream and downstream *var-cse* respectively, could be classified into group I and II, on the criterion of sequence identity. Indeed,

| Motif name     | Amino acid sequence   | Nucleotide sequence  |
|----------------|-----------------------|--|
| C:<br>I:       | SEAPAA<br>PAA         | TCAGAAGCA <b>CCAGCTGCA</b><br><b>CCAGCTGCT</b>                             |
| C:<br>G:       | SEAPES<br>PES         | TCAGAGGCA <b>CCAGAGAGC</b><br><b>CCAGAGAGC</b>                             |
| C:<br>E:<br>E: | SEAPA<br>SEAA<br>SEAV | <b>TCAGAAGCT</b> CCTGCTAGC<br><b>TCAGAAGCT</b> GCA<br><b>TCAGAAGCT</b> GTA |
| C:<br>D:       | SEAPAE<br>EAPAE       | TCAGAAGC <b>CCAGCAGAA</b><br>GAAGC <b>ACCAGCAGAA</b>                       |

**Fig. 2** Similarity between repeat motifs. Similar repeat motifs were aligned. Identical amino acids or nucleotides are indicated by bold face



sequence identity matrixes showed that along the B and D regions, percentage sequence identities are much higher between sequences belonging to the same group than between sequences that belong to different groups (Fig. 3). On the other hand, along the regions flanking BCD, *i.e.* the A and E regions, all the sequences are almost identical. Therefore, it is within the B, C (or *var-cse*) and D regions that the sequences can be classified into two groups. We conclude that these results show that the *cse* locus is mosaic.

**Involvement of *var-cse* in cell segregation**

The high intraspecific variability of the *var-cse* region could suggest that it is a dispensable region of the gene. As Cse is involved in the separation of *S. thermophilus* cells (Borges et al. 2005), the involvement of the *var-cse* region in cell segregation was investigated.

For this purpose, 80% of the *var-cse* region was deleted in-frame in the CNRZ368 strain (Fig. 1A) using an allelic replacement strategy. The resulting CNRZ368- $\Delta_{67-411}var-cse$  strain was grown, in parallel with CNRZ368 and CNRZ368- $\Delta cse$ , in liquid medium without shaking. The CNRZ368- $\Delta cse$  strain is a *cse* null mutant that was previously constructed by in-frame deletion of the *cse* ORF (Borges et al. 2005).

Phase-contrast photonic microscopy observation revealed that the CNRZ368- $\Delta_{67-411}var-cse$  cell chains were shorter than those of the  $\Delta cse$  mutant (Fig. 4A). This was reinforced by counting the number of cells per chain for each strain.

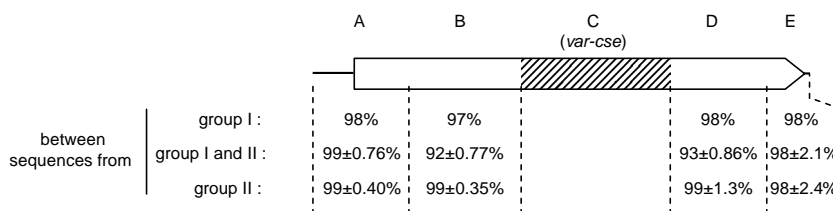
All the cells of the CNRZ368- $\Delta_{67-411}var-cse$  mutant formed chains of less than 351 cells, while more than 99% of the CNRZ368- $\Delta cse$  mutant cells were found in chains of more than 1000 cells (Fig. 4B). This result shows that a residual cell segregation activity is encoded by the *cse* mutant gene devoid of the  $_{67-411}var-cse$  region. On the other hand, 80% and 23% of the WT and the CNRZ368- $\Delta_{67-411}var-cse$  mutant cells, respectively, compose chains of 1 to 50 cells (Fig. 4B). Moreover, none of the WT chains contained more than 151 cells, yet 25% of mutant chains exhibited this length or longer. Thus, the CNRZ368- $\Delta_{67-411}var-cse$  chains are typically longer than those of the WT, indicating that the mutant with a partial deletion of the *var-cse* region is impaired in cell segregation. These data show that the *var-cse* region is required for full cell segregation in *S. thermophilus* strain CNRZ368.

**Discussion**

The *var-cse* variability: a possible consequence of repeat richness

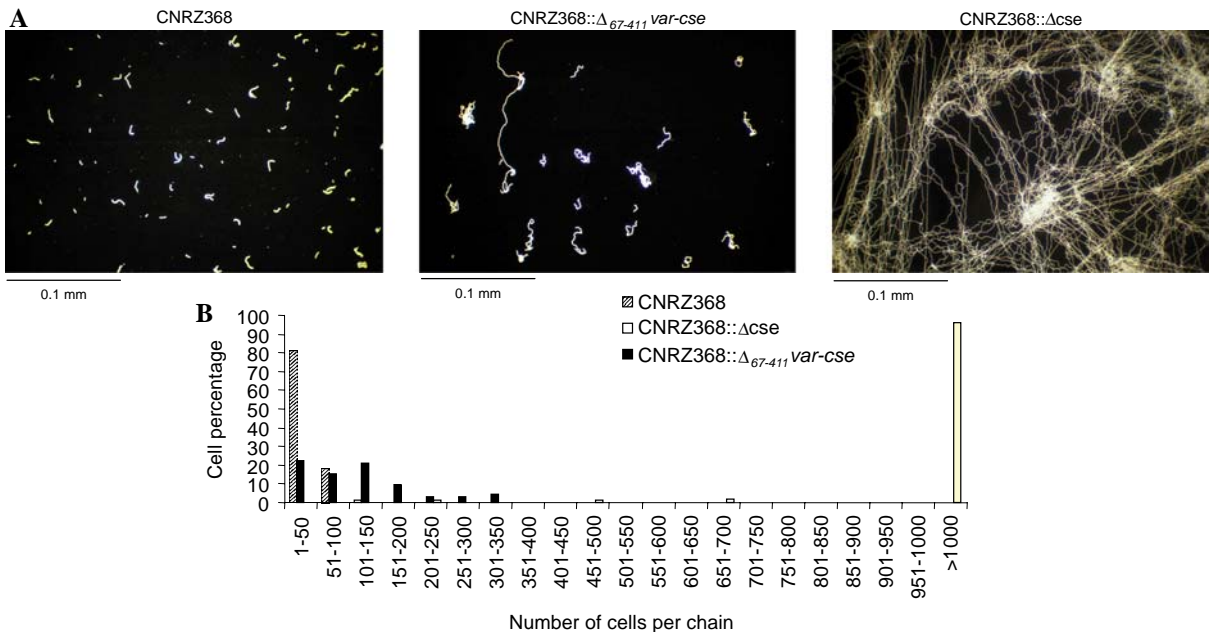
Analysis of the *var-cse* sequence from 19 *S. thermophilus* strains has revealed the extent of the intraspecific variability of this *cse* region. This variability is emphasized when considering that *S. thermophilus* strains are closely related (Bolotin et al. 2001; P. Renault, personal communication).

The *var-cse* polymorphism can result from repeat number variability. Such variability was



**Fig. 3** The *cse* locus is mosaic. The sequences of the *cse* locus were aligned with CLUSTALW. Based on the alignment, a sequence identity matrix was built and the mean value and standard deviation corresponding to comparison between sequences from group I, sequences from group I and group II, and sequences from group II

was reported for regions A, B, D and E of the *cse* locus. Values without standard deviation resulted from comparison of two sequences. Such analysis is not reported for the region 3 (corresponding to *var-cse*) since alignment is not valuable for this region. The open arrow represent the *cse* ORF and indicate its reading direction



**Fig. 4** Influence of *var-cse* partial deletion on cell segregation activity. **(A)** Cell chain morphology of *S. thermophilus* strains. **(B)** Results from counting the number of

cells per chain. At least 1000 cells were counted from three independent cultures. Cells were photographed and counted in stationary phase after 20 h of growth

already widely described for repeat regions (Yang and Gabriel 1995; van Belkum et al. 1998; Wilton et al. 1998; Lukowski et al. 2000; Schupp et al. 2000; Rasmussen and Bjorck 2001; Areschoug et al. 2002; Koroleva et al. 2002; Karatzas et al. 2003; de Benito et al. 2004). It is likely that insertions/deletions (indels) of repeat units result from polymerase slippage, during DNA replication, or by single strand annealing (SSA) (Michel 1999). These two mechanisms of illegitimate recombination imply pairing of ectopic complementary sequences. Sequence alignment of *var-cseI*, *J*, *K* suggests that an indel of an E motif occurred (Fig. 1A). This indel would result from ectopic pairing of two sequences, each one constituting a whole E repeat unit. Interestingly, considering the similarities between repeat motifs, it is likely that illegitimate recombination events are implicated in pairing of sequences from different repeat motifs. This hypothesis is supported by our *var-cseI*, *J*, *K* alignment. This alignment revealed an indel of a CH unit (Fig. 1A), potentially resulting from pairing of the 5' end of the inserted/deleted C motif, with the 5' end of the E motif downstream the CH indel

(Fig. 1A). Thus, the *var-cse* region could have evolved by recombination between identical repeats but also between different repeats. The ability of different repeats to undergo recombination would have increased the variety of recombination events in *var-cse* and could therefore be a factor responsible for the high polymorphism of this region.

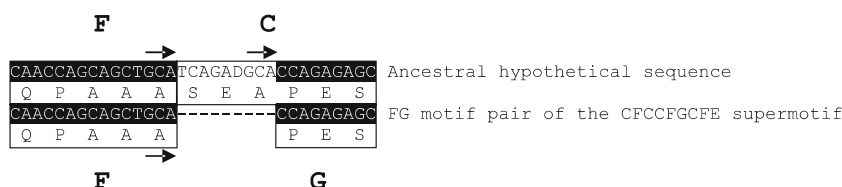
#### Origin of the diversity of the repeats

Partial pairing of repeated motifs could be responsible for indels of blocks of whole repeats units, but also they could induce partial deletion of repeat units. Indeed, since the G motif is entirely comprised within the C repeat (Fig. 2), and taking into account the abundance of the latter, a simple hypothesis is that the G motif has been generated by partial deletion of an ancestral C motif (Fig. 5). A probable scenario would implicate an ectopic pairing of the GCA trinucleotide repeats: one constituting the 3' end of the F repeat unit upstream of the hypothetical C repeat unit and the other encoding the alanine amino acid of the C hypothetical repeat unit (Fig. 5).

This ectopic pairing would have generated the deletion of the SEA encoding region of the hypothetical C repeat unit, generating the G motif. It is notable that at the end of this scenario, the G motif is not yet a repeat.

The alleles can contain large nucleotide duplications. Three alleles, grouping together the *var-cse* sequence of 11 *S. thermophilus* strains, exhibit an interspersed duplication of the CFCCFGCFE block of motifs (Fig.1). Two hypotheses could explain this duplication: that the same succession of repeat motifs appeared independently twice following multiple indels or, alternatively, that a single mutation resulted in the duplication of one of the two copies. Taking into account the high conservation of the CFCCFGCFE supermotif (1 nucleotide divergence between the 2 copies along 130 bp), it is more probable that the CFCCFGCFE duplication resulted from a unique mutation. Thus, the *var-cse* region would have evolved by duplication of motif blocks.

Interestingly, the CFCCFGCFE duplication would have duplicated the enclosed G motif that would have been created by partial deletion of a C hypothetical motif. This hypothesis would explain how the G motif, initially generated as a single copy, was duplicated and thus became a repeat. This hypothesis is supported by the observation that all the G motifs identified here are included in the CFCCFGCFE supermotifs (Fig. 1A). A similar mechanism of repeat generation by recombination between existing interspersed repeats has already been suggested for the G protein from two strains of Group G streptococci (Fahnestock et al. 1986; Filpula et al. 1987; Kehoe 1994).



**Fig. 5** Sequence alignment illustrating the hypothetical mechanism resulting in the origin of the G motif. Identical nucleotides between aligned sequences are represented by highlighted white letters. The trinucleotide sequence not duplicated in the hypothetical ancestral sequence and not duplicated in the actual sequence is indicated by an arrow.

In conclusion, a mutation series consisting of partial deletion of repeats followed by supermotif duplication could be responsible in part for the high repeat variety of the *var-cse* region.

#### Implication of horizontal transfer

Comparison of the *var-cse* region from 19 *S. thermophilus* strains and the flanking regions of 10 of them revealed that *cse* is mosaic. The sequences from group I exhibit 7% nucleotide divergence with the sequences from group II, along the B and C regions of *cse*, regions that are not supposed to evolve rapidly since they are not composed of repeats. It was previously reported that the divergence average between *S. thermophilus* strains sequences is 0.2% (Bolotin et al. 2001; P. Renault, personal communication). Therefore, the intraspecific divergence noticed in the *cse* locus is unusual for *S. thermophilus* sequences. Numerous mosaic genes have been previously described, especially in streptococci (Whatmore and Kehoe 1994; Kapur et al. 1995; Whatmore and Dowson 1999; Filipe et al. 2000; Hakenbeck 2000). These genes are typically formed by allelic recombination between a genomic locus and another allele acquired from horizontal transfer. It is likely that the *cse* mosaic nature has resulted from this mechanism.

#### Extent of the tolerance for *var-cse* variability regarding the *cse* cell segregation activity

Our partial *var-cse* deletion led to a decrease of cell segregation activity. Although the *var-cse* region is required for full cell segregation activity,

The hypothesis implies the generation of the G motif by partial deletion of the hypothetical and ancestral C motif. Partial deletion would have occurred by illegitimate recombination implicating an ectopic pairing of two copies of the duplicated GCA sequence

it exhibits high variability. This suggests that the function fulfilled by *var-cse* does not lead to counter-selection against this variability and has allowed the genomic fixation of a large number of alleles. This hypothesis is supported by results from *var-cse* allelic replacement experiments (Borges et al. 2005). Indeed, the replacement of the *var-cse* sequence of strain LMG18311 by that from CNRZ368, which results in shortening by 60 bp and the alteration of approximately 60% of the remaining amino acid sequence, did not have significant consequence on chain length. However, since the  $\Delta_{67-411}$ *var-cse* mutant is affected in cell segregation, this variability tolerance could be restricted to a minimal size of the *var-cse* region. Thus, the role of the *var-cse* region could be to maintain an appropriately orientated CHAP domain within the cell envelope i.e. with the Cse protein anchored by the LysM domain and the *var-cse* region acting as a spacer. This interpretation is consistent with the results from our previous *var-cse* allelic replacement experiments (Borges et al. 2005) since these would be predicted to maintain an appropriately displayed CHAP domain in the Cse constructs.

Analysis of the sequenced streptococcal genomes revealed two other LysM-CHAP domain proteins: SMU\_367 from *Streptococcus mutans* and BAB61101 from *Streptococcus intermedius*, whose functions are unknown. In SMU\_367 and BAB61101 proteins, the LysM and the CHAP domains are spaced by a short sequence of 31 amino acids and 36 amino acids, respectively. For Cse, the shortest linker sequence identified was that of *S. thermophilus* strain A054 which is 236 amino acids long, about seven times longer than the spacer in the other two LysM-CHAP proteins. Thus, the sequences spacing LysM and CHAP domains are not always long, suggesting that the functionality of this combination of domains does not necessarily require a long interdomain linker.

**Acknowledgements** We thank P. Renault (Institut National de la Recherche Agronomique, Jouy-en-Josas, France) for personal communications concerning the relatedness of *S. thermophilus* strains. 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.

## References

- Areschoug T, Linse S, Stalhammar-Carlemalm M, Heden LO, Lindahl G (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* 184:6376–6383
- Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, Fonstein M, Overbeek R, Kyprides N, Purnelle B, Prozzi D, Ngui K, Masuy D, Hancy F, Burteau S, Boutry M, Delcour J, Goffeau A, Hols P (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22:1554–1558
- Bolotin A, Wincker P, Mauger S, Jaillon O, Malmale K, Weissenbach J, Ehrlich SD, Sorokin A (2001) The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* 11:731–753
- Borges F, Layec S, Thibessard A, Fernandez A, Gintz B, Hols P, Decaris B, Leblond-Bourget N (2005) *cse*, a Chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J Bacteriol* 187:2737–2746
- Buist G, Kok J, Leenhouts KJ, Dabrowska M, Venema G, Haandrikman AJ (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
- de Benito I, Cano ME, Aguero J, Garcia Lobo JM (2004) A polymorphic tandem repeat potentially useful for typing in the chromosome of *Yersinia enterocolitica*. *Microbiology* 150:199–204
- Drams S, Dehoux P, Cossart P (1993) Common features of gram-positive bacterial proteins involved in cell recognition. *Mol Microbiol* 9:1119–1121
- Fahnestock SR, Alexander P, Nagle J, Filpula D (1986) Gene for an immunoglobulin-binding protein from a group G streptococcus. *J Bacteriol* 167:870–880
- Filipe SR, Severina E, Tomasz A (2000) Distribution of the mosaic structured *murM* genes among natural populations of *Streptococcus pneumoniae*. *J Bacteriol* 182:6798–6805
- Filpula D, Alexander P, Fahnestock SR (1987) Nucleotide sequence of the protein G gene from *Streptococcus* GX7805, and comparison to previously reported sequences. *Nucleic Acids Res* 15:7210
- Fischetti VA, Pancholi V, Schneewind O (1991) Common characteristics of the surface proteins from gram-positive cocci. In: Dunny GM, Cleary PP, Mc Kay LL (eds) *Genetics and Molecular Biology of Streptococci, Lactococci, and Enterococci*. ASM press, pp 290–294
- Hakenbeck R (2000) Transformation in *Streptococcus pneumoniae*: mosaic genes and the regulation of competence. *Res Microbiol* 151:453–456
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95–98

- Hallet B (2001) Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiol* 4:570–581
- Kapur V, Kanjilal S, Hamrick MR, Li LL, Whittam TS, Sawyer SA, Musser JM (1995) Molecular population genetic analysis of the streptokinase gene of *Streptococcus pyogenes*: mosaic alleles generated by recombination. *Mol Microbiol* 16:509–519
- Karatzas KA, Wouters JA, Gahan CG, Hill C, Abee T, Bennik MH (2003) The CtsR regulator of *Listeria monocytogenes* contains a variant glycine repeat region that affects piezotolerance, stress resistance, motility and virulence. *Mol Microbiol* 49:1227–1238
- Kehoe MA (1994) Cell-wall-associated proteins in Gram-positive bacteria. In: Hakenbeck J-MGaR (ed) *Bacterial Cell Wall*. Elsevier Science B.V, pp 217–261
- Koroleva IV, Efstratiou A, Suvorov AN (2002) Structural heterogeneity of the streptococcal C5a peptidase gene in *Streptococcus pyogenes*. *J Bacteriol* 184:6384–6386
- Leenhouts K (1995) Integration strategies and vectors. *Dev Biol Stand* 85:523–530
- Lukowski S, Nakashima K, Abdi I, Cipriano VJ, Ireland RM, Reid SD, Adams GG, Musser JM (2000) Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. *Infect Immun* 68:6542–6553
- Maguin E, Duwat P, Hege T, Ehrlich D, Gruss A (1992) New thermosensitive plasmid for gram-positive bacteria. *J Bacteriol* 174:5633–5638
- Mercenier A, Robert C, Romero DA, Slos PYL (1987) Transfection of *Streptococcus thermophilus* spheroblasts. In: Ferreti JJ, Curtiss R (eds) *Streptococcal genetics*. ASM press, pp 234–237
- Michel B (1999) Illegitimate recombination in bacteria. In: Charlebois RL (ed) *Organization of the prokaryotic genome*. American Society for Microbiology, pp 129–150
- Rasmussen M, Bjorck L (2001) Unique regulation of *SclB*—a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol Microbiol* 40:1427–1438
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press
- Schupp JM, Klevytska AM, Zinser G, Price LB, Keim P (2000) *vrpB*, a hypervariable open reading frame in *Bacillus anthracis*. *J Bacteriol* 182:3989–3997
- Stingele F, Mollet B (1996) Disruption of the gene encoding penicillin-binding protein 2b (*pbp2b*) causes altered cell morphology and cease in exopolysaccharide production in *Streptococcus thermophilus* Sf6. *Mol Microbiol* 22:357–366
- Terzaghi B, Sandine W (1975) Improved medium for lactic streptococci and their bacteriophages. *Appl Environ Microbiol* 29:807–813
- 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
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 22:4673–4680
- van Belkum A (1999) Short sequence repeats in microbial pathogenesis and evolution. *Cell Mol Life Sci* 56:729–734
- van Belkum A, Scherer S, van Alphen L, Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* 62:275–293
- Whatmore AM, Dowson CG (1999) The autolysin-encoding gene (*lytA*) of *Streptococcus pneumoniae* displays restricted allelic variation despite localized recombination events with genes of pneumococcal bacteriophage encoding cell wall lytic enzymes. *Infect Immun* 67:4551–4556
- Whatmore AM, Kehoe MA (1994) Horizontal gene transfer in the evolution of group A streptococcal *emm-like* genes: gene mosaics and variation in *Vir regulons*. *Mol Microbiol* 11:363–374
- Wilton JL, Scarman AL, Walker MJ, Djordjevic SP (1998) Reiterated repeat region variability in the ciliary adhesin gene of *Mycoplasma hyopneumoniae*. *Microbiology* 144(Pt 7):1931–1943
- Yang Y, Gabriel DW (1995) Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *J Bacteriol* 177:4963–4968



# Characterization of Proteins Belonging to the CHAP-Related Superfamily within the Firmicutes

S  verine Layec Bernard Decaris Nathalie Leblond-Bourget

Laboratoire de G  n  tique et Microbiologie, UMR INRA 1128, IFR 110, Facult   des Sciences et Techniques, Universit   Henri Poincar   Nancy 1, Vand  uvre-l  s-Nancy, France

## Key Words

Firmicutes · CHAP protein · Peptidoglycan hydrolase

## Abstract

Cell division is a dynamic process ending by separation of the daughter cells. This final step requires the cleavage of the murein septum synthesized during cell division. In *Streptococcus thermophilus*, *cse* plays an important role in cell separation. *Cse* protein contains, at its N-terminal end, a signal peptide and a putative LysM motif suggesting that it is secreted and able to bind to the cell wall. Furthermore, the C-terminus of *Cse* carries a putative cysteine, histidine-dependent amidohydrolases/peptidases (CHAP) domain conferring to the protein a potential catalytic activity. To gain insight into the role of *Cse* in the cell division process, in silico analysis of the Firmicutes proteins displaying CHAP-related domain was undertaken. This work allowed us to distinguish and characterize within the Firmicutes the 2 families of proteins (CHAP and NlpC/p60) belonging to the CHAP superfamily. These 2 families regroup mainly peptidoglycan hydrolases. Data from the literature indicate that NlpC/p60 and CHAP proteins cleave distinct peptidoglycan bonds. Among the enzymes characterized within the Firmicutes, NlpC/p60 proteins are  $\gamma$ -D-glutamate-meso-diaminopimel-

ate muropeptidase. Instead, CHAP enzymes involved in cell separation are N-acetylmuramoyl-L-alanine amidase and CHAP lysins have endopeptidase activity.

Copyright    2008 S. Karger AG, Basel

## Introduction

*Streptococcus thermophilus* belongs to the Firmicutes, and because of its ability to produce lactic acid, is used as a starter of fermentation for the conversion of milk into yogurt and many cheeses (such as Emmental, Gruy  re, Mozzarella and Cheddar) [Dellaglio et al., 1994].

Strains of *S. thermophilus* always grow as cell chains, indicating that the separation of daughter cells is not fully completed. Up to now, little is known about the *S. thermophilus* daughter cell splitting process. However, a role for *cse* in cell separation has been proposed based on the finding that *cse* deletion resulted in extremely long chains of cells (composed of more than 200 cells) compared to wild-type phenotype (containing an average of 40 cells) [Borges et al., 2005]. To prove that the formation of long chains was indeed the result of the lack of *cse*, a complementation experiment was performed with a plasmid expressing *cse*. The deleted mutant transformed with the

plasmid containing *cse* grew with a short-chain phenotype, while the mutant carrying only the empty plasmid grew as long chain [Borges et al., 2005].

Although the involvement of *cse* in the separation of daughter cells during cellular division is clear, the mechanism by which Cse acts is still unknown. Indeed, the long-chain phenotype of the *cse* mutant could either be caused by changes in the cell wall (directly or indirectly involving *cse*) or be related to a potential murein hydrolase activity of Cse. Involvement of multiple peptidoglycan hydrolases in daughter cell separation has been reported in a number of bacteria. The 3 amidases AmiA, AmiB and AmiC of *Escherichia coli*, the LytA and LytB cell wall hydrolases of *Streptococcus pneumoniae*, and the major autolysins AcmA of *Lactococcus lactis* and Atl1 of *Staphylococcus aureus* are all involved in cell separation [Buist et al., 1995; De Las Rivas et al., 2002; Garcia et al., 1999; Heidrich et al., 2001; Yamada et al., 1996]. The peptidoglycan hydrolases LytE (CwlF), LytF (CwlE) and YojL (CwlS) have also been demonstrated to be important for cell division in *Bacillus subtilis* [Fukushima et al., 2006; Ishikawa et al., 1998; Margot et al., 1998; Ohnishi et al., 1999]. Simultaneous inactivation of *lytE*, *lytF* and *yoyL* in *B. subtilis* is concomitant with the formation of longer chains of cells [Fukushima et al., 2006]. In *S. pneumoniae*, depletion of *pcsB* elicits the generation of long chains [Ng et al., 2004; Reinscheid et al., 2001]. In *Listeria monocytogenes*, the depletion of the autolysin p60 results in the generation of long-chain filaments and is associated with invasion into a mammalian cell line [Kuhn and Goebel, 1989].

The Cse cell separation protein consists of 2 putative domains that are separated by a central region [Borges et al., 2006]. The N-terminal region of Cse contains a signal peptide involved in protein export [Borges et al., 2005] and a putative LysM domain that binds the surface protein to the cell wall [Steen et al., 2003, 2005]. The C-terminal region of Cse contains homology to the cysteine, histidine-dependent amidohydrolases/peptidases (CHAP) superfamily of proteins characterized by the presence of conserved cysteine and histidine involved in the enzymatic activity [Anantharaman and Aravind, 2003; Bateman and Rawlings, 2003; Rigden et al., 2003]. Two distinct domains, NLPC/P60 (PF00877) and CHAP (PF05257), characterize this superfamily. The CHAP superfamily of proteins has been proposed to function mainly, but not only, in peptidoglycan hydrolysis [Anantharaman and Aravind, 2003; Bateman and Rawlings, 2003; Rigden et al., 2003].

Since the CHAP domain was assumed to be the Cse catalytic domain and to get insight into Cse role during

cell separation, in silico analysis of CHAP and NlpC/p60 families of proteins within the Firmicutes was done, allowing us to predict that Cse acts as a putative N-acetylmuramoyl-L-alanine amidase.

## Results and Discussion

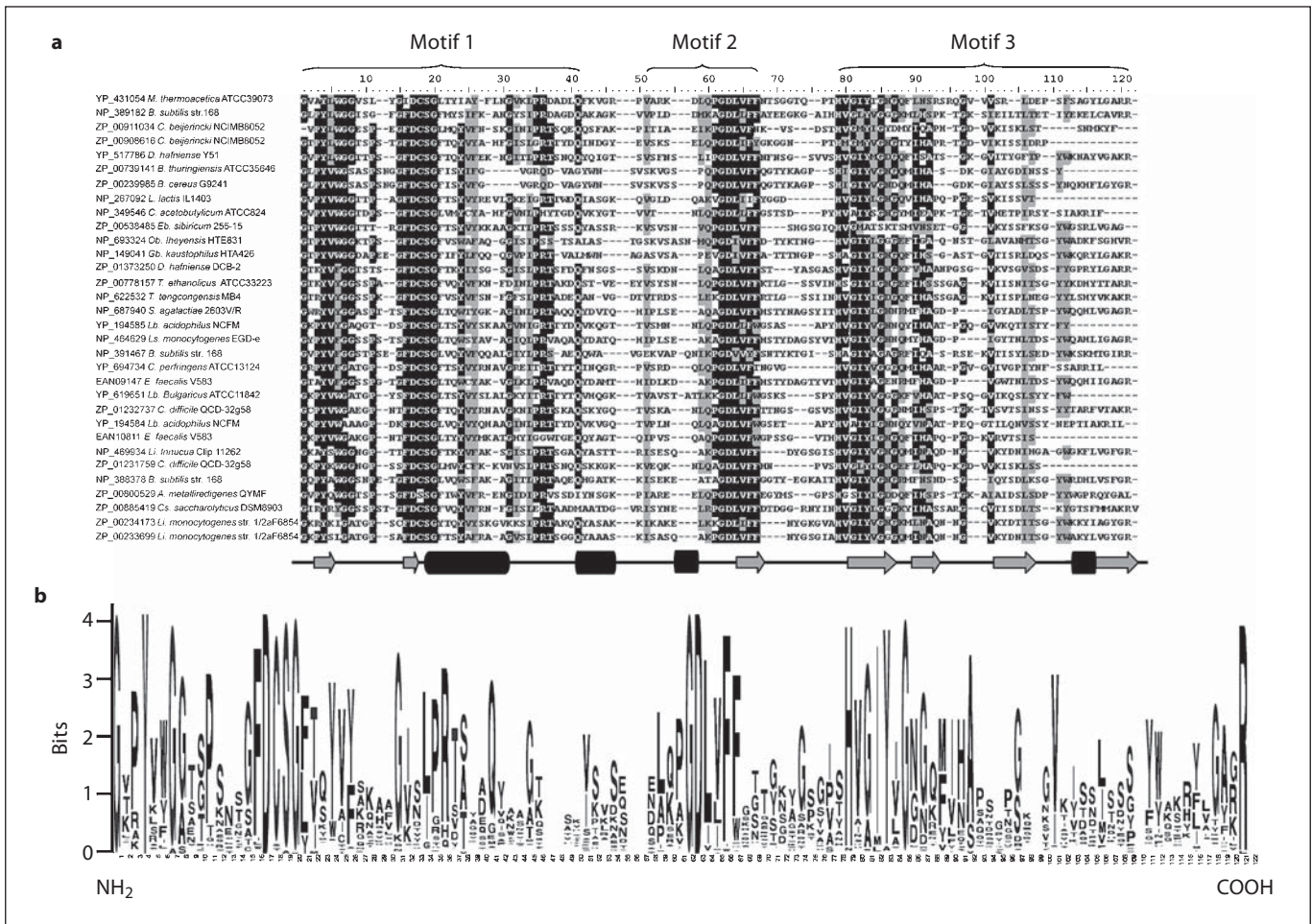
### Characterization of the CHAP and NLPC/P60 Domains within Firmicutes Proteins

#### In silico Analysis of the Firmicutes NlpC/p60 Family of Proteins

A BlastP [Altschul et al., 1997] search initiated with the NLPC/P60 consensus sequence (PF00877) given by the Pfam database [Bateman et al., 2000; Finn et al., 2006] and against the protein database of 150 entire Firmicutes genomes ([http://www.ncbi.nlm.nih.gov/sutils/genom\\_table.cgi](http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)) allowed the retrieval of 100 NlpC/p60 proteins. Analysis of these data indicates that these proteins are not equally distributed within the Firmicutes genomes. Indeed, the NlpC/p60 proteins are encoded by the Bacillales (37%, most of them by the *Bacillus* and *Listeria* genus), Clostridia (34%) and Lactobacilliales (20%). In contrast, no NlpC/p60 encoding gene was found in the genome of either Staphylococcaceae or Mollicutes.

Figure 1a shows an alignment of the NLPC/P60 domains of 32/100 Firmicutes proteins achieved by ClustalW Multiple Alignment software [Thompson et al., 1994]. Within the Firmicutes, NLPC/P60 domains comprise 84–112 amino acids and are characterized by 3 motifs containing conserved amino acids (see fig. 1a). Within them, the most conserved residues are Cys<sub>18</sub>, Gly<sub>62</sub>, Asp<sub>63</sub> and His<sub>79</sub> (present in 99, 100, 100 and 97% of the NlpC/p60 proteins, respectively) and the less conserved His<sub>91</sub> that is shared by 74% of these proteins. These conserved residues play a role in the catalytic activity of the NLPC/P60 domains [Anantharaman and Aravind, 2003; Rigden et al., 2003]. The consensus sequence of the Firmicutes NLPC/P60 domains is given in figure 1b and was determined using the Weblogo software (<http://weblogo.berkeley.edu/logo.cgi>) [Crooks et al., 2004]. Secondary structure predictions by the PSI-PRED Protein Structure Prediction Server (<http://bioinf.cs.ucl.ac.uk/psipred/>) [Jones, 1999] show that the NLPC/P60 consensus domain should belong to the  $\alpha$  and  $\beta$  structural class, with the N-terminal half containing 2 predicted  $\beta$  strands, followed by 3 predicted  $\alpha$  helices. The C-terminal half is mainly composed of 5 predicted  $\beta$  strands and 1 predicted  $\alpha$  helix (fig. 1a).





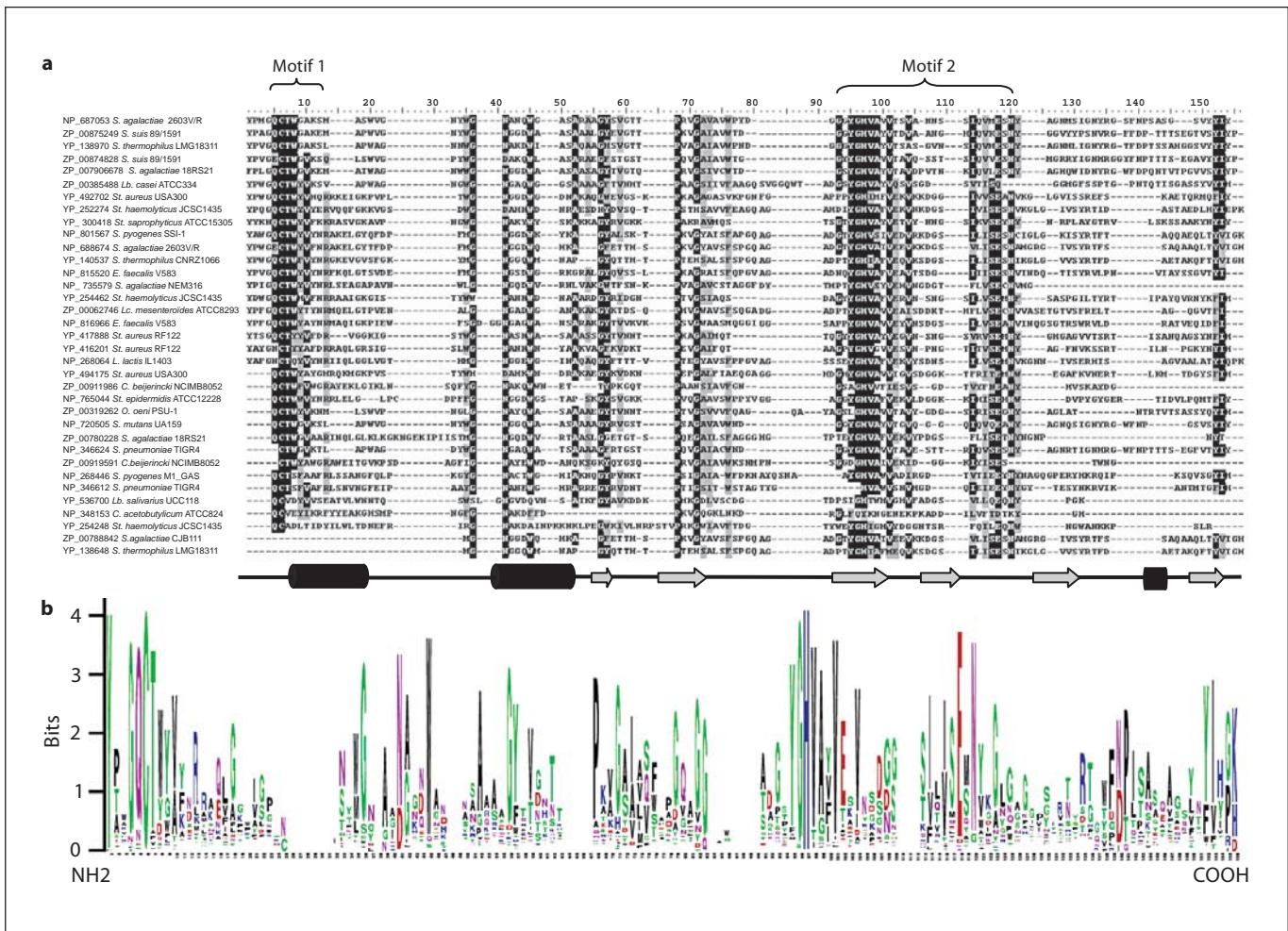
**Fig. 1.** NLPC/P60 domains of Firmicutes proteins. Only proteins displaying less than 70% identity between them are shown in this figure. **a** Multiple alignments of the NLPC/P60 domains of 32 proteins. The proteins identifiers are given as EMBL/GenBank accession number. Identical and similar amino acids are shown in darkly and lightly shaded boxes, respectively. Motifs with conserved residues are indicated within brackets. Secondary structure prediction of the NLPC/P60 domains was carried out with the PSI-PRED Protein Structure Prediction Server. Predicted  $\alpha$  helices and  $\beta$  strands are shown below the alignments as cylinders and arrows, respectively. *M.* = *Moorella*; *B.* = *Bacillus*; *C.* = *Clo-*

*tridium*; *D.* = *Desulfitobacterium*; *L.* = *Lactococcus*; *Eb.* = *Exiguobacterium*; *Ob.* = *Oceanobacillus*; *Gb.* = *Geobacillus*; *T.* = *Thermoanaerobacter*; *S.* = *Streptococcus*; *Lb.* = *Lactobacillus*; *Ls.* = *Listeria*; *E.* = *Enterococcus*; *A.* = *Alkaliphilus*; *Cs.* = *Caldicellulosiruptor*. **b** Consensus sequence of the Firmicutes NLPC/P60 domain based on an alignment of 100 NlpC/p60 proteins identified by BlastP and drawn in the SeqLogo format using the Weblogo tool. The size of each letter is proportional to its degree of conservation. The size of each column represents the statistical importance of the given position.

**In silico Analysis of the Firmicutes CHAP Family of Proteins**

In order to retrieve the CHAP proteins encoded by the Firmicutes, a BlastP search was initiated with the CHAP (PF05257) consensus sequence and the COG3942 sequence from the NCBI Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) [Marchler-Bauer et al., 2005]. The COG3942 sequence is a conserved protein sequence that is composed of a CHAP

domain and of another sequence that is not defined as a domain. On the whole, 230 proteins were extracted. Within the Firmicutes, the distribution of the CHAP proteins differs from that of NlpC/p60 proteins. Indeed, the majority of the CHAP proteins is encoded by the genomes of Lactobacillales (130/230), from which 50% were encoded by *Streptococcus*. Within the Bacillales, *Bacillus* and *Listeria* are deprived of CHAP proteins in opposition to *Staphylococcus* that encode nearly 42% of them. Few



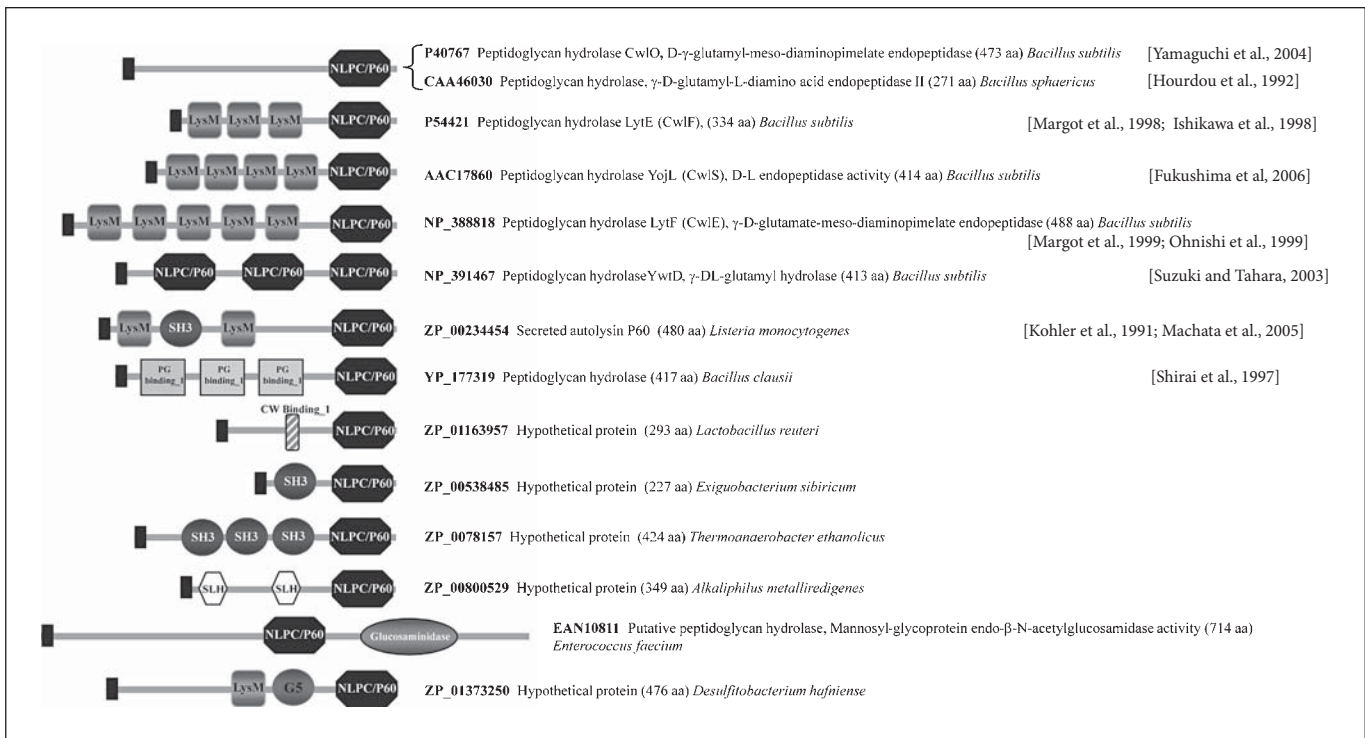
**Fig. 2.** CHAP domains of Firmicutes proteins. **a** Multiple alignments of the CHAP domains of 35 proteins. The proteins identifiers are given as EMBL/GenBank accession number. Identical and similar amino acids are shown in darkly and lightly shaded boxes, respectively. Motifs with conserved residues are indicated within brackets. Secondary structure prediction of the CHAP domain was carried out with the PSI-PRED Protein Structure Prediction Server. Predicted  $\alpha$  helices and  $\beta$  strands are shown below the alignments as cylinders and arrows, respectively. *S.* = *Strepto-*

*coccus*; *Lb.* = *Lactobacillus*; *St.* = *Staphylococcus*; *E.* = *Enterococcus*; *Lc.* = *Leuconostoc*; *L.* = *Lactococcus*; *C.* = *Clostridium*; *O.* = *Oenococcus*. **b** Consensus sequence of the Firmicutes CHAP domain based on the alignment of 230 CHAP proteins identified by BlastP and drawn in the the SeqLogo format using the Weblogo tool. The size of each letter is proportional to its degree of conservation. The size of each column represents the statistical importance of the given position.

are also encoded by *Clostridium*, 2 by *Clostridium beijerinckii* and 1 by *Clostridium acetobutylicum*. Thus, most of the Firmicutes genome encodes either NlpC/p60 or CHAP proteins.

Figure 2a shows an alignment of 35/230 of the CHAP proteins. The CHAP domain contains 53–117 amino acids and displays 2 highly conserved motifs (fig. 2a). These are characterized by the presence of the highly conserved Cys<sub>6</sub> and His<sub>97</sub> (present in at least 95% of the CHAP proteins). Between these 2 motifs, the well-conserved resi-

dues Gly<sub>36</sub>, Asn<sub>41</sub>, Trp<sub>44</sub>, Ala<sub>52</sub>, Gly<sub>56</sub>, Pro<sub>68</sub>, Gly<sub>71</sub> (shared by more than 74% of the proteins) and Tyr<sub>57</sub> (shared by 67% of the proteins) are present (fig. 2a). Point mutation of the conserved Cys and His residues or total deletion of this domain reveals that it retains peptidoglycan hydrolyase activity [Donovan et al., 2006; Huard et al., 2003; Navarre et al., 1999; Nelson et al., 2006; Pritchard et al., 2004; Yokoi et al., 2005]. Additional experiments demonstrated that the Asn residue (within motif 2; fig. 2a) is also crucial for the enzymatic function. The consensus se-



**Fig. 3.** Domain architecture of NlpC/p60 proteins within the Firmicutes. The NLPC/P60 domain can be found in the Pfam database, accession number PF00877. For more information about the other domains, see Pfam [Bateman et al., 2000], using the following accession numbers: LysM (PF01476), G5 (PF07501), SH3 (PF08460), CW-binding\_1 (PF01473), PG-binding\_1 (PF01471),

SLH (PF00395) and Glucosaminidase (PF01832). Small black boxes on the N-termini indicate signal peptides, predicted by the SignalPHMM server. For each kind of architecture, only 1 protein identified by its EMBL/GenBank accession number is given and characterized by the description available in the data banks.

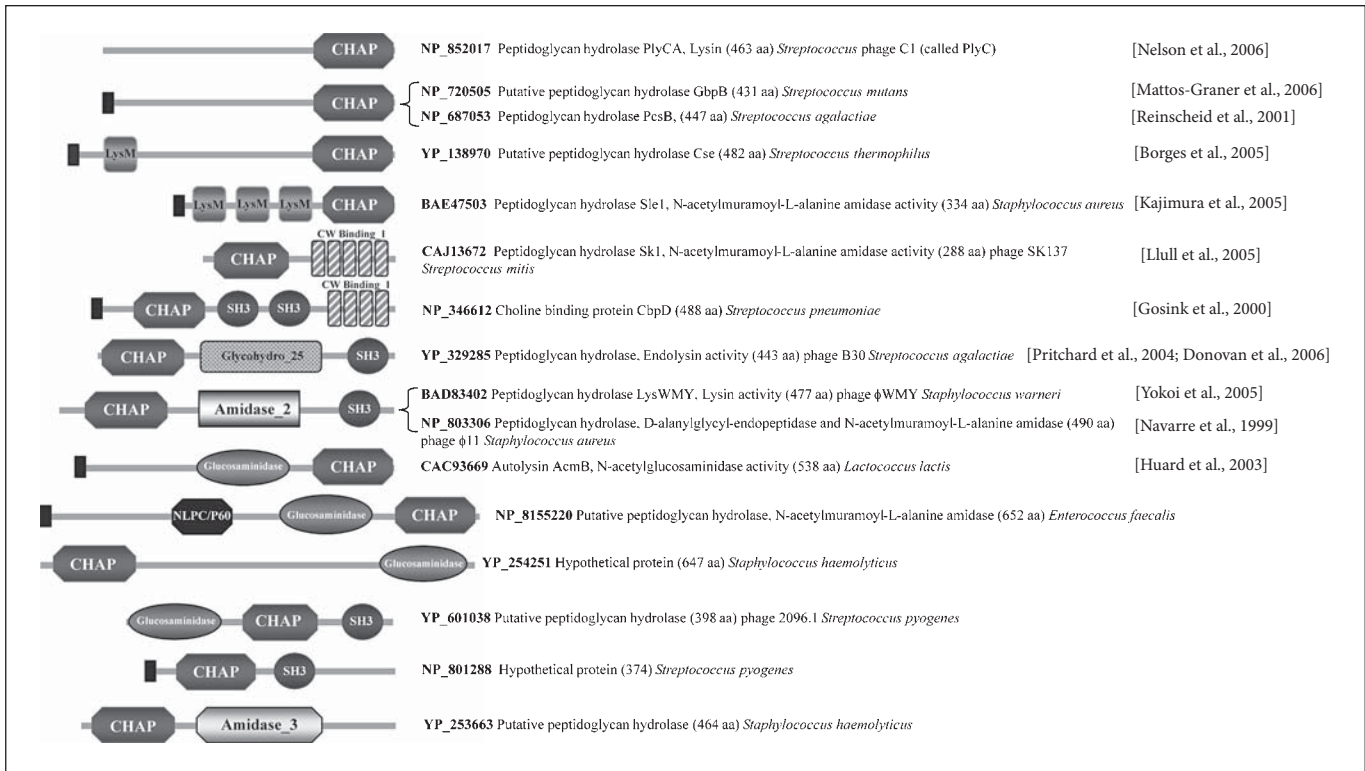
quence of the Firmicutes CHAP domain is given in figure 2b. Secondary structure predictions by the PSI-PRED Protein Structure Prediction Server (<http://bioinf.cs.ucl.ac.uk/psipred/>) show that the CHAP domain, like the NLPC/P60 domain, belongs to the  $\alpha$  and  $\beta$  structural class with the N-terminal half containing 2 predicted  $\alpha$  helices followed by 2 predicted  $\beta$  strands. The C-terminal half is mainly composed of 4 predicted  $\beta$  strands and 1 predicted  $\alpha$  helix in its COOH end (fig. 2a).

Although the primary sequence of the CHAP and NLPC/P60 domains of Firmicutes proteins are distinct, they adopt a common  $\alpha$  and  $\beta$  topology and are both characterized by the well-conserved Cys and His residues. Other conserved residues were demonstrated to be important as well [Anantharaman and Aravind, 2003; Bateman and Rawlings, 2003; Rigden et al., 2003; Yokoi et al., 2005]. In the model by Anantharaman and Aravind [2003], hydrolysis of the peptidoglycan bonds takes place as a general acid/base catalysis that requires Cys acting as the nucleophile that attacks the peptide bonds,

while His acts as a proton donor. The polar (Asn/His/Asp) residues serve as the orienting residue for the catalytic His residue [Anantharaman and Aravind, 2003; Rigden et al., 2003].

#### Domain Organization of the CHAP and NlpC/p60 Proteins within the Firmicutes

The 100 NlpC/p60 and 230 CHAP proteins found within the Firmicutes were submitted to a search for additional domains using the Pfam database (<http://www.sanger.ac.uk/Software/Pfam/>) and for a signal peptide using the SignalPHMM Server (<http://www.ebi.ac.uk/InterProScan/>) [Bendtsen et al., 2004]. Interestingly, the domains' organization of the NlpC/p60 proteins is distinct from that of the CHAP proteins. All NlpC/p60 proteins possess a signal peptide and in most cases multiple copies of given cell wall attachment domains or different cell wall binding domains (fig. 3). This suggests that all NlpC/p60 proteins are potentially secreted and located on the bacterial cell surface. In contrast, CHAP proteins



**Fig. 4.** Domain architecture of the CHAP proteins within the Firmicutes. The CHAP domain can be found in the Pfam database, accession number PF05257. For more information about the other domains, see Pfam using the following accession numbers: LysM (PF01476), SH3 (PF08460), CW-binding\_1 (PF01473),

Glycohydro\_25 (PF01183), Amidase\_2 (PF01510), Amidase\_3 (PF01520), Glucosaminidase (PF01832) and NLPC/P60 (PF00877). Small black boxes on the N-termini indicate signal peptides, predicted by the SignalPHMM server.

do not always contain a signal peptide (fig. 4) and several seem to be deprived of a cell wall attachment domain. This suggests that some of them are potentially non-secreted, whereas others are secreted and may be exposed to the bacterial cell surface.

Moreover, NlpC/p60 proteins are usually characterized by a single catalytic NLPC/P60 domain. One exception is YwtD from *B. subtilis* that carries 3 copies of the NLPC/P60 domain of which only 1 is functional [Bateman and Bycroft, 2000; Suzuki and Tahara, 2003]. In contrast, CHAP proteins often display amidase or glucosaminidase domains in addition to their CHAP domain. This multiplicity of catalytic domains confers to the proteins' multifunctional activities [Eckert et al., 2006; Navarre et al., 1999; Pritchard et al., 2004]. Within the Firmicutes, complex protein architectures are found only in *Enterococcus* genomes. In silico analyses reveal that *Enterococcus* NLPC/P60 domain is always associated with other catalytic domains (fig. 3, 4).

#### Cell Wall Binding Domains Encountered in NlpC/p60 and CHAP Proteins

The CHAP superfamily of proteins includes surface proteins containing various attachment domains that interact with the components in cell wall. Some of these binding domains are described in the articles by Desvaux et al. [2006] and Scott and Barnett [2006]. Bacterial cell surface binding domains frequently encountered in the CHAP superfamily of proteins are LysM (PF01476), CW-binding\_1 for choline binding (PF01473) and SH3 (PF08460; fig. 3, 4). The LysM domain is one of the most common attachment modules in bacterial cell surface proteins [Bateman and Bycroft, 2000]. The autolysins Atla of *Enterococcus faecalis*, AcmA of *L. lactis* and p60 of *L. monocytogenes* contain tandem repeats of the LysM motif [Eckert et al., 2006; Kohler et al., 1991; Steen et al., 2003, 2005]. Truncation of the LysM repeats of Atla and AcmA led to an important reduction in their activity, indicating that the LysM domain is important for peptido-

glycan binding and biological functioning [Eckert et al., 2006; Steen et al., 2005]. The function of LysM is thought to direct the enzyme to its site of action, that is cell separation sites and cell poles in the case of AcmA, YojL (CwlS) of *B. subtilis* and Sle1 of *S. aureus* [Fukushima et al., 2006; Kajimura et al., 2005; Steen et al., 2003].

The choline-binding domain (ChBD) is responsible for the noncovalent anchoring of the choline binding proteins to the choline moieties of both teichoic and lipoteichoic acids at the cell surface [Garcia et al., 1994; Hermoso et al., 2003; Janecek et al., 2000]. The *S. pneumoniae* cell wall hydrolases LytA, LytB and LytC possess several ChBDs positioned either in the C- or N-terminal end of the proteins [De Las Rivas et al., 2002; Garcia et al., 1999; Lopez et al., 1997]. The multiplicity of the ChBDs provides advantage for maintaining a close interaction with the cell envelope without perturbing the shape and function of the active catalytic modules of choline-binding proteins [Hermoso et al., 2003]. The ChBD also contributes to the virulence of *S. pneumoniae* [Fernandez-Tornero et al., 2001].

The SH3 (Src homology 3) domain was first discovered in eukaryotes and demonstrated to play an important role in cell-cell communication and signal transduction [Whisstock and Lesk, 1999]. The SH3 domain has recently been identified in bacteria and named SH3b. The basic role of bacterial SH3b domains from pathogen is considered to be the binding to receptors on eukaryotic cells to promote invasion [Whisstock and Lesk, 1999]. The SH3b domain of the *S. aureus* ALE-1 peptidoglycan hydrolase binds specifically to staphylococcal peptidoglycan bonds [Lu et al., 2006].

Additional cell wall binding domains such as PG-binding\_1 (PF01471), G5 (PF07501) and SLH (PF00395) are less frequent and specifically found in NlpC/p60 proteins (fig. 3). The G5 and PG-binding\_1 domains bind to the peptidoglycan [Bateman et al., 2005; Foster, 1991] and the SLH domain interacts with secondary cell wall polymers [May et al., 2006; Sara and Sleytr, 2000].

#### Function of the NlpC/p60 and CHAP Proteins

**NlpC/p60 Proteins.** One of the first characterized NlpC/p60 proteins was the autolysin p60 in *L. monocytogenes* [Kohler et al., 1991] that possesses an hydrolase activity required for a late step in cell division [Wuenscher et al., 1993]. Truncation of most of the NLP/C/P60 domain of p60 provokes its inactivation [Machata et al., 2005]. By analogy with p60, proteins displaying the same catalytic domain were hypothesized to function as peptidoglycan hydrolases involved in cell lysis. Within the Firmicutes

only a few NlpC/p60 proteins are characterized; all are found in *Bacillus* and are cell-separating rather than cell-lysing enzymes. The *Bacillus* cell wall hydrolases LytE (CwlF), LytF (CwlE) and YojL (CwlS) play a role in cell separation during vegetative growth [Fukushima et al., 2006; Ishikawa et al., 1998; Ohnishi et al., 1999]. CwlO is a *B. subtilis* enzyme with unknown function [Yamaguchi et al., 2004]. This protein is not essential for growth and seems not to be associated with daughter cell separation during cellular division [Yamaguchi et al., 2004]. The substrate bond specificity of LytF, CwlO, CwlS and the *Bacillus sphaericus* endopeptidase II indicates that these proteins are D,L-endopeptidase capable of hydrolyzing the linkage of a D- $\gamma$ -glutamyl-*meso*-diaminopimelic acid in the cell wall of *Bacillus* [Fukushima et al., 2006; Hourdou et al., 1992; Margot et al., 1998, 1999; Yamaguchi et al., 2004]. Another NlpC/P60 protein, YwtD, was identified in *B. subtilis* that degrades the poly- $\gamma$ -glutamic acid, a polypeptide secreted into the growth medium [Suzuki and Tahara, 2003]. YwtD cleaves the  $\gamma$ -glutamyl bond between D- and L-glutamic acid of the poly- $\gamma$ -glutamic acid. Thus, the NlpC/p60 proteins from *Bacillus* characterized so far hydrolyze diverse substrates that contain a  $\gamma$ -glutamyl moiety. This substrate specificity is consistent with that of other NlpC/p60 proteins found in other bacterial genus [Anantharaman and Aravind, 2003; Rigden et al., 2003] and allows us to hypothesize that all the Firmicutes NlpC/p60 enzymes hydrolyze  $\gamma$ -glutamyl-containing substrates.

If this hypothesis is true, NlpC/p60 proteins of *Enterococcus* would not be able to hydrolyze their own peptidoglycan, since they lack  $\gamma$ -glutamyl-*meso*-diaminopimelic acid [Schleifer and Kandler, 1972]. This would justify the presence of another catalytic domain to ensure the hydrolytic activity of these proteins and explain the complex architecture of the NlpC/p60 proteins of *Enterococcus*.

**CHAP Proteins.** Among the Firmicutes, only a few bacterial and phage-encoded CHAP proteins have been characterized and function mainly as murein hydrolases. Most of these CHAP murein hydrolases participate either in cell division, including daughter cell separation, or in cell lysis. Downexpression of *pcsB* from *S. pneumoniae* and *S. agalactiae* resulted in the formation of long chains of cells and led to abnormal, uncontrolled cell wall synthesis at misplaced septa and around large cells. These results show that PcsB protein has a role in cell division [Ng et al., 2004; Reinscheid et al., 2001]. Deletion of *cse* from *S. thermophilus* results in an extremely long-chain phenotype [Borges et al., 2005, 2006]. GbpB,

a secreted CHAP protein from *Streptococcus mutans*, is hypothesized to be part of a complex involved in peptidoglycan synthesis and cell division [Mattos-Graner et al., 2006]. Sle1 from *S. aureus* is involved in cell separation and was demonstrated to cleave N-acetylmuramoyl-L-alanine bonds in a cross bridge that interlinks the 2 murein strands in peptidoglycan [Kajimura et al., 2005]. The choline-binding protein Sk1 from *Streptococcus mitis*, whose function is unknown, displays the same N-acetylmuramoyl-L-alanine amidase activity [Llull et al., 2006].

Instead of being involved in cell separation, 4 other CHAP proteins have been implicated in cell lysis. These are AcmB autolysin from *L. lactis* [Huard et al., 2003], *S. agalactiae* phage B30 endolysin [Donovan et al., 2006; Pritchard et al., 2004], staphylococcal phage  $\Phi$ 11 [Navarre et al., 1999] and streptococcal PlyCA bacteriophage lysin [Nelson et al., 2006]. The autolysin AcmB displays a central glucosaminidase domain and a CHAP domain [Huard et al., 2003]. AcmB hydrolyzes the peptidoglycan bonds between N-acetylglucosamine and N-acetylmuramic acid and thus is an N-acetylglucosaminidase [Huard et al., 2003]. In this case, the specificity of the CHAP domain alone has not been determined. The *S. agalactiae* phage B30 endolysin contains an SH3-binding domain and 2 catalytic domains, CHAP and Acm [Pritchard et al., 2004]. Truncations and point mutations indicated that the Acm domain requires the SH3 domain for activity, while the CHAP domain is responsible for nearly all cell lysis activity [Donovan et al., 2006]. The Acm domain of the phage B30 lysin is associated with glycosidase activity, while its CHAP domain is associated with endopeptidase activity [Pritchard et al., 2004]. The CHAP domain of the staphylococcal phage  $\Phi$ 11 lysin is also responsible for D-alanyl-glycine endopeptidase activity [Navarre et al., 1999]. This enzyme also disposed of an amidase\_2 domain responsible for an N-acetylmuramoyl-L-alanine amidase activity [Navarre et al., 1999].

Although the cleavage bonds within peptidoglycan have been determined only in a few Firmicutes CHAP enzymes, it seems that CHAP proteins involved in cell separation have N-acetylmuramoyl-L-alanine amidase activity, whereas those implicated in cell lysis have an endopeptidase specificity. It should be noticed that CHAP enzymes involved in cellular lysis are in most cases bifunctional enzymes displaying more than 1 catalytic domain [Donovan et al., 2006; Navarre et al., 1999; Pritchard et al., 2004]. These domains may act in a cooperative manner to provide tightly regulated cleavage of peptidoglycan substrates [Bateman and Rawlings, 2003].

## Conclusion

Peptidoglycan is the major component of Gram-positive cell walls and ensures cell wall integrity and rigidity [Rogers et al., 1980]. The structure of peptidoglycan is an alternating copolymer of N-acetylmuramic acid-( $\beta$ 1-4)-N-acetylglucosamine with peptide side chains branching from the N-acetylmuramic acid residues. Wall peptides are cross-linked with other peptides that are attached to a neighboring glycan strand, thereby generating a three-dimensional molecular network that surrounds the cell.

Bacteria synthesize peptidoglycan hydrolases capable of hydrolyzing their own peptidoglycan to ensure important cellular process during cell growth and division [Smith et al., 1996]. One class of peptidoglycan hydrolase is the CHAP superfamily of proteins. Rigden et al. [2003] have recently pointed out that only 11 members of the CHAP superfamily are biochemically characterized, but some of them were completely unrelated to cell wall degradation function. Within the Firmicutes, this does not seem to be the case, since all the proteins belonging to the CHAP superfamily identified so far function as peptidoglycan hydrolases. However, various peptidoglycan bonds can be hydrolyzed depending on the nature of the catalytic (NlPC/P60 or CHAP) domain. Within the Firmicutes, CHAP proteins that play a role in the separation of daughter cells during cellular division have an N-acetylmuramoyl-L-alanine amidase specificity, whereas others (part of the CHAP and the NlPC/p60 family) have an endopeptidase activity. Among the Firmicutes CHAP superfamily of endopeptidases, NlPC/p60 proteins hydrolyze the peptide stem between D-glutamic acid and the meso-diaminopimelic acid bond that is not present in all Firmicutes. In *Staphylococcus* or *Streptococcus*, where the meso-diaminopimelic acid is replaced by an L-lysine, NlPC/p60 proteins are absent. Interestingly, the cleavage bond specificity of the CHAP endopeptidases varies according to the bacterial species in which the protein is found. Hence, the lysin encoded by phage B30 cleaves between D-Ala of the stem peptides and L-Ala of the interpeptide bridges of the streptococcal peptidoglycan [Pritchard et al., 2004]. Instead, the phage  $\Phi$ 11 endopeptidase cleaves between the D-Ala of the stem peptide and a glycine residue of the (Gly)<sub>5</sub> bridge in the staphylococcal peptidoglycan bridge that is absent in group B *Streptococcus*. Thus, the CHAP hydrolytic bond specificity is likely to depend on the composition of the peptidoglycan stem peptides and the way they are cross-linked. This could explain why the Mollicutes, characterized by a complete

lack of a cell wall [Trachtenberg, 1998], are deprived of NlpC/p60 and CHAP proteins.

The biological process in which the CHAP superfamily of enzymes is involved is also determined by the spatial location of the protein. This depends on the cell wall binding domains associated with the NLPC/P60-CHAP catalytic domain, but also on cell wall constituents that may hinder cell wall attachment of the peptidoglycan binding domains [Steen et al., 2003, 2005]. Nevertheless, in some of the NlpC/p60-CHAP proteins no binding domain is detected, suggesting either that they possess a nonidentified binding domain or are deprived of it. Interestingly, the example of PlyC demonstrated that the catalytic and binding domains can be encoded by 2 distinct genes [Nelson et al., 2006].

The CHAP domain is present in proteins containing other types of catalytic domains. One of them is PECACE,

a potential peptidoglycan hydrolyzing domain in Gram-positive bacteria [Pagliero et al., 2005]. Several NlpC/p60 and CHAP proteins harbor a multimodular architecture with multiple hydrolytic activities, probably needed for the regulation of peptidoglycan metabolism [Bateman and Rawlings, 2003; Pagliero et al., 2005].

Analysis of the CHAP superfamily of proteins allowed us to better understand their role within the Firmicutes and to gain insight into Cse activity. The Cse protein from *S. thermophilus* belongs to the CHAP family of enzymes and, like some other CHAP proteins, is involved in cell separation during cell division [Borges et al., 2005, 2006]. By analogy with other CHAP proteins, it is therefore conceivable that Cse acts as a peptidoglycan hydrolase with a potential N-acetylmuramoyl-L-alanine amidase specificity.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- Anantharaman V, Aravind L: Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* 2003;4:R11.
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: The Pfam protein families database. *Nucleic Acids Res* 2000;28:263–266.
- Bateman A, Bycroft M: The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* 2000;299:1113–1119.
- Bateman A, Holden MT, Yeats C: The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics* 2005;21:1301–1303.
- Bateman A, Rawlings ND: The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem Sci* 2003;28:234–237.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S: Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–795.
- Borges F, Layec S, Fernandez A, Decaris B, Leblond-Bourget N: High genetic variability of the *Streptococcus thermophilus* cse central part, a repeat rich region required for full cell segregation activity. *Antonie Van Leeuwenhoek* 2006;90:245–255.
- Borges F, Layec S, Thibessard A, Fernandez A, Gintz B, Hols P, Decaris B, Leblond-Bourget N: cse, a chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J Bacteriol* 2005;187:2737–2746.
- Buist G, Kok J, Leenhouts KJ, Dabrowska M, Venema G, Haandrikman AJ: Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation. *J Bacteriol* 1995;177:1554–1563.
- Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–1190.
- De Las Rivas B, Garcia JL, Lopez R, Garcia P: Purification and polar localization of pneumococcal LytB, a putative endo- $\beta$ -N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* 2002;184:4988–5000.
- Dellaglio F, De Roissard H, Torriani S, Curk MC, Janssens D: Caractéristiques générales des bactéries lactiques; in De Roissard, Luquet FM (eds): *Bactéries lactiques*. Uriage, Lorica, 1994, pp 25–116.
- Desvaux M, Dumas E, Chafsey I, Hebraud M: Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett* 2006;256:1–15.
- Donovan DM, Foster-Frey J, Dong S, Rousseau GM, Moineau S, Pritchard DG: The cell lysis activity of the *Streptococcus agalactiae* bacteriophage B30 endolysin relies on the cysteine, histidine-dependent amidohydrolase/peptidase domain. *Appl Environ Microbiol* 2006;72:5108–5112.
- Eckert C, Lecerf M, Dubost L, Arthur M, Mesnage S: Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. *J Bacteriol* 2006;188:8513–8519.
- Fernandez-Tornero C, Lopez R, Garcia E, Gimenez-Gallego G, Romero A: A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat Struct Biol* 2001;8:1020–1024.
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A: Pfam: clans, web tools and services. *Nucleic Acids Res* 2006;34:D247–D251.
- Foster SJ: Cloning, expression, sequence analysis and biochemical characterization of an autolytic amidase of *Bacillus subtilis* 168 trpC2. *J Gen Microbiol* 1991;137:1987–1998.
- Fukushima T, Afkham A, Kurosawa S, Tanabe T, Yamamoto H, Sekiguchi J: A new D,L-endo-peptidase gene product, YojL (renamed CwlS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis*. *J Bacteriol* 2006;188:5541–5550.
- Garcia JL, Diaz E, Romero A, Garcia P: Carboxy-terminal deletion analysis of the major pneumococcal autolysin. *J Bacteriol* 1994;176:4066–4072.
- Garcia P, Gonzalez MP, Garcia E, Lopez R, Garcia JL: LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol* 1999;31:1275–1277.
- Heidrich C, Templin MF, Ursinus A, Merdanovic M, Berger J, Schwarz H, de Pedro MA, Holtje JV: Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* 2001;41:167–178.

- Hermoso JA, Monterroso B, Albert A, Galan B, Ahrazem O, Garcia P, Martinez-Ripoll M, Garcia JL, Menendez M: Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* 2003;11:1239–1249.
- Hourdou ML, Duez C, Joris B, Vacheron MJ, Guinand M, Michel G, Ghuysen JM: Cloning and nucleotide sequence of the gene encoding the  $\gamma$ -D-glutamyl-L-diamino acid endopeptidase II of *Bacillus sphaericus*. *FEMS Microbiol Lett* 1992;70:165–170.
- Huard C, Miranda G, Wessner F, Bolotin A, Hansen J, Foster SJ, Chapot-Chartier MP: Characterization of AcmB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*. *Microbiology* 2003;149:695–705.
- Ishikawa S, Hara Y, Ohnishi R, Sekiguchi J: Regulation of a new cell wall hydrolase gene, *cwIF*, which affects cell separation in *Bacillus subtilis*. *J Bacteriol* 1998;180:2549–2555.
- Janecek S, Svensson B, Russell RR: Location of repeat elements in glucanucrases of *Leuconostoc* and *Streptococcus* species. *FEMS Microbiol Lett* 2000;192:53–57.
- Jones DT: Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195–202.
- Kajimura J, Fujiwara T, Yamada S, Suzawa Y, Nishida T, Oyamada Y, Hayashi I, Yamagishi J, Komatsuzawa H, Sugai M: Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase *SleI* involved in cell separation of *Staphylococcus aureus*. *Mol Microbiol* 2005;58:1087–1101.
- Kohler S, Bubert A, Vogel M, Goebel W: Expression of the *iap* gene coding for protein p60 of *Listeria monocytogenes* is controlled on the posttranscriptional level. *J Bacteriol* 1991; 173:4668–4674.
- Kuhn M, Goebel W: Identification of an extracellular protein of *Listeria monocytogenes* possibly involved in intracellular uptake by mammalian cells. *Infect Immun* 1989;57: 55–61.
- Llull D, Lopez R, Garcia E: Skl, a novel choline-binding N-acetylmuramoyl-L-alanine amidase of *Streptococcus mitis* SK137 containing a CHAP domain. *FEBS Lett* 2006;580:1959–1964.
- Lopez R, Garcia E, Garcia P, Garcia JL: The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb Drug Resist* 1997;3:199–211.
- Lu JZ, Fujiwara T, Komatsuzawa H, Sugai M, Sakon J: Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *J Biol Chem* 2006;281:549–558.
- Machata S, Hain T, Rohde M, Chakraborty T: Simultaneous deficiency of both MurA and p60 proteins generates a rough phenotype in *Listeria monocytogenes*. *J Bacteriol* 2005;187: 8385–8394.
- Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DJ, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH: CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 2005;33:D192–D196.
- Margot P, Pagni M, Karamata D: *Bacillus subtilis* 168 gene *lytF* encodes a  $\gamma$ -D-glutamate-meso-diaminopimelate peptidase expressed by the alternative vegetative sigma factor,  $\sigma^D$ . *Microbiology* 1999;145:57–65.
- Margot P, Wahlen M, Gholamhosseinian A, Piggot P, Karamata D: The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. *J Bacteriol* 1998;180:749–752.
- Mattos-Graner RO, Porter KA, Smith DJ, Hosogi Y, Duncan MJ: Functional analysis of glucan binding protein B from *Streptococcus mutans*. *J Bacteriol* 2006;188:3813–3825.
- May A, Pusztahelyi T, Hoffmann N, Fischer RJ, Bahl H: Mutagenesis of conserved charged amino acids in SLH domains of *Thermoanaerobacterium thermosulfurigenes* EM1 affects attachment to cell wall sacculi. *Arch Microbiol* 2006;185:263–269.
- Navarre WW, Ton-That H, Faull KF, Schneewind O: Multiple enzymatic activities of the murein hydrolase from staphylococcal phage  $\phi$ phi11: identification of a D-alanyl-glycine endopeptidase activity. *J Biol Chem* 1999; 274:15847–15856.
- Nelson D, Schuch R, Chahales P, Zhu S, Fischetti VA: PlyC: a multimeric bacteriophage lysin. *Proc Natl Acad Sci USA* 2006;103:10765–10770.
- Ng WL, Kazmierczak KM, Winkler ME: Defective cell wall synthesis in *Streptococcus pneumoniae* R6 depleted for the essential PcsB putative murein hydrolase or the VicR (YycF) response regulator. *Mol Microbiol* 2004;53: 1161–1175.
- Ohnishi R, Ishikawa S, Sekiguchi J: Peptidoglycan hydrolase *LytF* plays a role in cell separation with *CwIF* during vegetative growth of *Bacillus subtilis*. *J Bacteriol* 1999;181:3178–3184.
- Pagliero E, Dideberg O, Vernet T, Di Guilmi AM: The PECACE domain: a new family of enzymes with potential peptidoglycan cleavage activity in Gram-positive bacteria. *BMC Genomics* 2005;6:19.
- Pritchard DG, Dong S, Baker JR, Engler JA: The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology* 2004;150:2079–2087.
- Reinscheid DJ, Gottschalk B, Schubert A, Eikmanns BJ, Chhatwal GS: Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J Bacteriol* 2001;183:1175–1183.
- Rigden DJ, Jedrzejewski MJ, Galperin MY: Amidase domains from bacterial and phage autolysins define a family of  $\gamma$ -D,L-glutamate-specific amidohydrolases. *Trends Biochem Sci* 2003; 28:230–234.
- Rogers HJ, Perkins HR, Ward JB: *Microbial Cell Walls and Membranes*. London, Chapman and Hall, 1980.
- Sara M, Sleytr UB: S-Layer proteins. *J Bacteriol* 2000;182:859–868.
- Schleifer KH, Kandler O: Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* 1972;36:407–477.
- Scott JR, Barnett TC: Surface proteins of Gram-positive bacteria and how they get there. *Annu Rev Microbiol* 2006;60:397–423.
- Smith TJ, Blackman SA, Foster SJ: Peptidoglycan hydrolases of *Bacillus subtilis* 168. *Microb Drug Resist* 1996;2:113–118.
- Steen A, Buist G, Horsburgh GJ, Venema G, Kuipers OP, Foster SJ, Kok J: AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. *FEBS J* 2005;272: 2854–2868.
- Steen A, Buist G, Leenhouts KJ, El Khattabi M, Grijpstra F, Zomer AL, Venema G, Kuipers OP, Kok J: Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *J Biol Chem* 2003;278:23874–23881.
- Suzuki T, Tahara Y: Characterization of the *Bacillus subtilis* *ywtD* gene, whose product is involved in  $\gamma$ -polyglutamic acid degradation. *J Bacteriol* 2003;185:2379–2382.
- Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
- Trachtenberg S: Mollicutes-wall-less bacteria with internal cytoskeletons. *J Struct Biol* 1998;124:244–256.
- Whisstock JC, Lesk AM: SH3 domains in prokaryotes. *Trends Biochem Sci* 1999;24:132–133.
- Wuenschel MD, Kohler S, Bubert A, Gerike U, Goebel W: The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity. *J Bacteriol* 1993;175:3491–3501.
- Yamada S, Sugai M, Komatsuzawa H, Nakashima S, Oshida T, Matsumoto A, Suginata H: An autolysin ring associated with cell separation of *Staphylococcus aureus*. *J Bacteriol* 1996;178:1565–1571.
- Yamaguchi H, Furuhashi K, Fukushima T, Yamamoto H, Sekiguchi J: Characterization of a new *Bacillus subtilis* peptidoglycan hydrolase gene, *yvcE* (named *cwLO*), and the enzymatic properties of its encoded protein. *J Biosci Bioeng* 2004;98:174–181.
- Yokoi KJ, Kawahigashi N, Uchida M, Sugahara K, Shinohara M, Kawasaki K, Nakamura S, Taketo A, Kodaira K: The two-component cell lysis genes *holWMy* and *lysWMy* of the *Staphylococcus warneri* M phage varphiWMy: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. *Gene* 2005;351:97–108.





INSTITUT PASTEUR

Research in Microbiology xx (2008) 1–9



www.elsevier.com/locate/resmic

# Diversity of Firmicutes peptidoglycan hydrolases and specificities of those involved in daughter cell separation

S  verine Layec, Bernard Decaris, Nathalie Leblond-Bourget\*

Laboratoire de G  n  tique et Microbiologie, UMR INRA/UHP 1128, IFR 110, Facult   des Sciences et Techniques, Nancy-Universit  , BP 239, 54506 Vandoeuvre-l  s-Nancy, France

Received 7 March 2008; accepted 10 June 2008

## Abstract

Within *Streptococcus thermophilus*, Cse was identified as the major cell disconnecting peptidoglycan hydrolase (PGH) and was demonstrated to be species-specific. To identify cell disconnecting PGHs encoded by other *Streptococcus* genomes, we explored the diversity of domains carried by Firmicutes PGHs, and especially that of enzymes involved in daughter cell separation. This work brings to light the diversity of PGHs and reveals that each species recruits its own cell-separating enzyme distinct from that of the others. This specificity is probably correlated with the diversity of substrates found in the bacterial cell wall.

   2008 Elsevier Masson SAS. All rights reserved.

**Keywords:** Peptidoglycan hydrolase (PGH); Diversity; Modular architecture of enzymes; Daughter cell separation; *Streptococcus*

## 1. Introduction

Bacterial cellular division ends with cell disconnection, a still poorly understood process that requires peptidoglycan hydrolases (PGHs). Indeed, in many bacteria, a correlation between the absence of PGH activity and failure of cell separation has been reported [76].

*Streptococcus thermophilus*, a lactic acid bacterium massively used for dairy product fermentation, grows in the form of cell chains [61], indicating that cell separation is never fully completed within this species. Indeed, in *S. thermophilus*, cell separation is ensured exclusively (or mainly) by Cse, a modular enzyme thought to possess peptidoglycan hydrolase activity. The *S. thermophilus*  $\Delta cse$  mutant forms extraordinarily long chains of cells due to incompleteness of cell daughter separation [6]. The *cse* gene encodes a modular protein containing a signal peptide and a putative LysM motif, suggesting that Cse is secreted and able to bind to the cell wall, respectively. Cse also carries a putative CHAP (cysteine,

histidine-dependent aminohydrolase/peptidase) domain potentially conferring catalytic activity upon the protein [37]. Surprisingly, Cse was found to be specific to *S. thermophilus*. Instead, *Streptococcus pneumoniae* disposes of LytB [26] and *Streptococcus mutans* of Aml [80], key daughter cell-separating enzymes. One common feature of cell-separating enzymes described thus far is that they are all PGHs.

Bacteria are known to possess a variety of specific PGHs [62]. In order to obtain an overview of Firmicutes PGHs, and especially those involved in cell separation, we searched for PGHs encoded by Firmicutes genomes. We studied their modular organization so as to elucidate the main characteristics of PGHs involved in daughter cell disconnection. Special attention was paid to *Streptococcus* cell-separating PGHs.

## 2. PGH substrate specificities

Within the last decade, an increasing number of PGHs have been characterized in Gram-positive bacteria using biochemical techniques leading to identification of their activities. In order to search for PGHs, enzymes annotated as assuming one of these hydrolytic activities were extracted from the

\* Corresponding author. Tel.: +33 3 83 68 42 10; fax: +33 3 83 68 44 99.  
E-mail address: bourget@nancy.inra.fr (N. Leblond-Bourget).

complete genome sequence of the 129 complete Firmicutes genomes. This enabled us to identify their catalytic domains. The Pfam database was then used to collect all Firmicutes proteins displaying these domains and their sequences were retrieved from EMBL/GenBank (<http://www.sanger.ac.uk/Software/Pfam/>) [21].

Peptidoglycan is the major component of cell walls of Gram-positive bacteria. Peptidoglycan is a polymer of alternating units of the disaccharide *N*-acetylmuramic acid (MurNAc)-( $\beta$ -1,4)-*N*-acetylglucosamine (GlcNAc), that are cross-linked by short peptides (Fig. 1). Wall peptides are cross-linked with other peptides that are attached to a neighboring glycan strand, thereby generating a three-dimensional molecular network that surrounds the cell [62].

Firmicutes are characterized by a huge diversity of PGHs. Specific enzymes exist for almost all peptidoglycan covalent bonds. According to the specific bond that is split down in the peptidoglycan, distinct classes of PGHs are defined [27,64,66]. *N*-acetylmuramoyl-L-alanine amidases cleave the amide bond between the lactic acid side chain of MurNAc and L-alanine of the short peptide. Peptidases are capable of cleaving short peptides (carboxypeptidases) or peptide cross-bridges (endopeptidases). *N*-acetylglucosaminidases and *N*-acetylmuramidases hydrolyze  $\beta$ -1,4-bonds between disaccharide units of peptidoglycan (Fig. 1).

Cleavage specificity depends on the catalytic domain carried by the PGHs (Fig. 1).

In most cases, catalytic domains are specialized in cleavage of one specific peptidoglycan bond and the corresponding

enzyme belongs to one of the previously described functional classes. Nevertheless, one catalytic domain, VanY, was reported to be able to cleave several different peptidoglycan linkages (for an extensive review see Ref. [22]). To our knowledge, the only catalytic domain that was reported to display two distinct functions is the CHAP domain: it confers *N*-acetylmuramoyl-L-alanine amidase activity upon proteins involved in cell separation [33,38] and shows endopeptidase specificity in those implicated in cell lysis [19,50,56].

### 3. Modular architecture of the PGHs

Most Firmicutes PGHs are characterized by a modular structure composed of two domains: one having catalytic activity and the other containing a recognition/cell wall binding domain. Within Firmicutes PGHs, a total of 14 different catalytic domains were listed, as well as 27 distinct surface association domains. Fig. 2 illustrates all possibilities of domain associations found within Firmicutes PGHs along with the diversity of their architectures. Some rare Firmicutes PGHs (often carboxypeptidases) are composed of a single catalytic domain. Alternatively, others display multiple distinct or identical catalytic domains associated with one or more substrate/binding domains (Fig. 2).

Commonly, the presence of multiple distinct catalytic domains confers upon PGHs multifunctional activities [20,50,56]. For instance, glucosaminidase and amidase activities of Atl from *Staphylococcus aureus* are conferred by glucosaminidase and amidase\_2 domains, respectively [4,52,76].

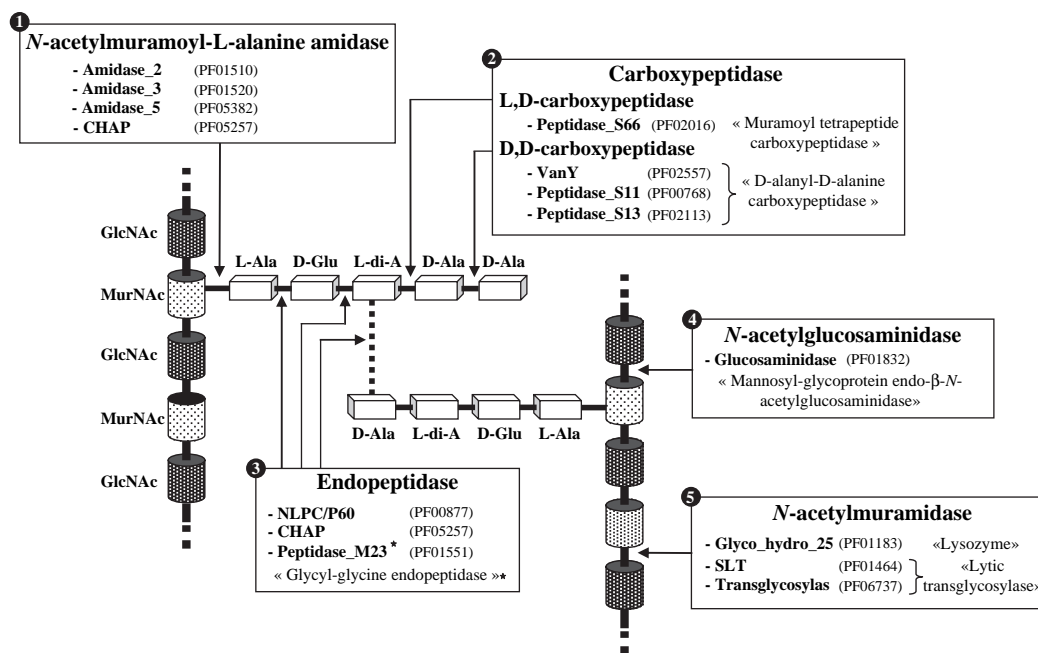


Fig. 1. Schematic structure of peptidoglycan from Gram-positive bacteria and PGH substrate specificity. Peptidoglycan is a polymer of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, linked by  $\beta$ -1,4-glycosidic bonds. The chemistry of glycan strands varies only slightly between different bacteria. However, the composition of short peptides which are linked to the lactyl group of MurNAc is variable [62]. For Gram-positive bacteria, these short peptides are most often L-alanyl-L-lysyl-D-alanyl-D-alanine, except for *Bacillus* and *Listeria* species, wherein a dibasic amino acid is a mesodiaminopimelyl residue [60]. PGHs are classified into five categories according to their specific site of cleavage. Arrows point to specific bonds cleaved by each category of PGH. Catalytic domains are indicated by their Pfam database accession number (<http://www.sanger.ac.uk/Software/Pfam/>).

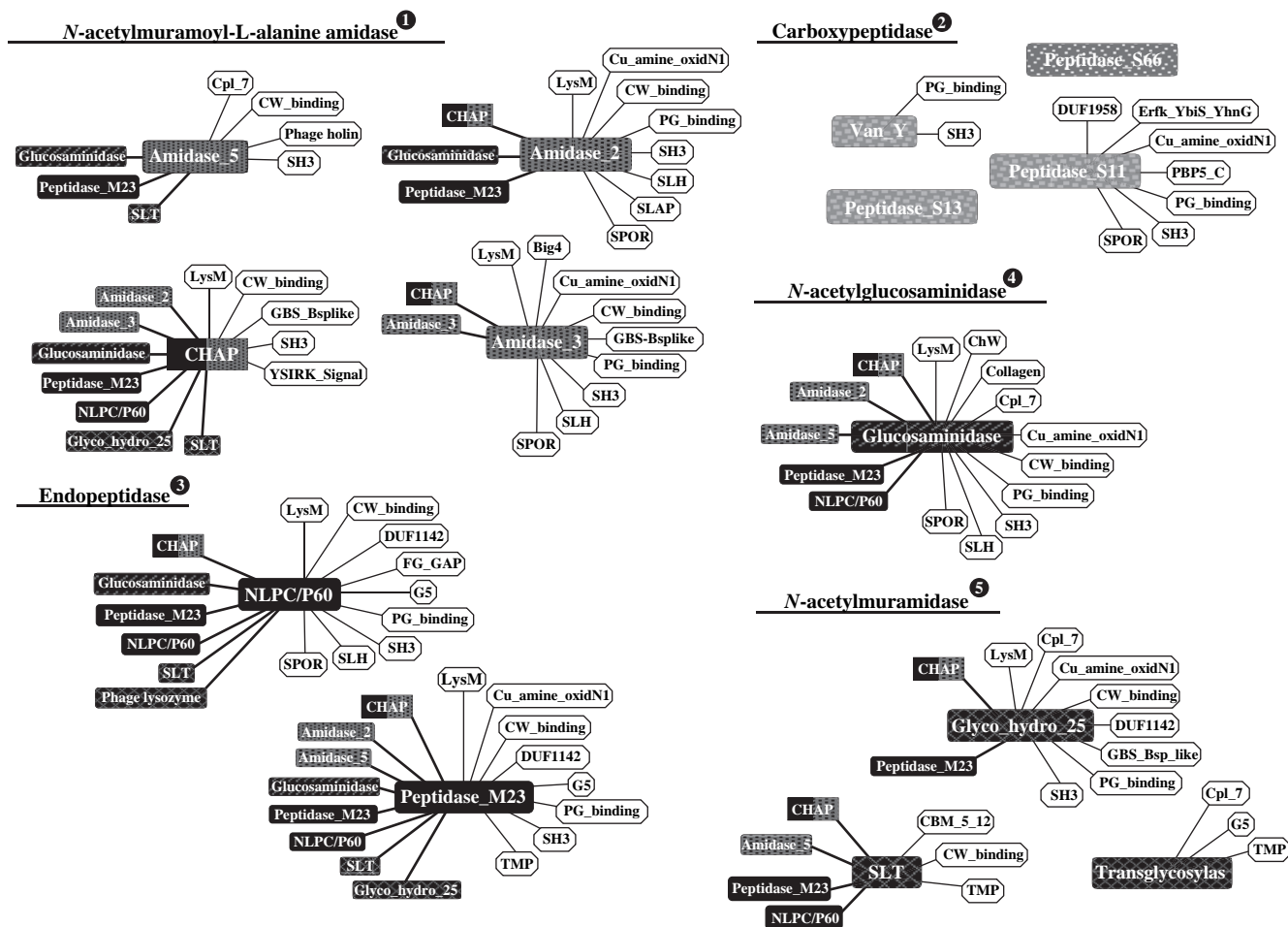


Fig. 2. Diversity of modular structure of PGHs within the Firmicutes. Schematic representation of domain combinations forming Firmicutes PGHs. PGH sequences were submitted to a search for additional domains using the Pfam database [21] (<http://www.sanger.ac.uk/Software/Pfam/>). For more information about non-catalytic domains, see Pfam using the following accession numbers: Big\_4 (PF07532), CBM\_5\_12 (PF02839), ChW (PF07538), Collagen (PF01391), Cpl\_7 (PF08230), Cu\_amine\_oxidN1 (PF07833), CW\_binding\_1 (PF01473), CW\_binding\_2 (PF04122), DUF1142 (PF06605), DUF1958 (PF09211), Erfk\_YbiS\_YhnG (PF03734), FG\_GAP (PF01839), GBS\_Bsp\_like (PF08481), G5 (PF07501), LysM (PF01476), PBPS\_C (PF07943), PG\_binding\_1 (PF01471), Phage holin (PF04688), SH3\_2 (PF07653), SH3\_3 (PF08239), SH3\_4 (PF06347), SH3\_5 (PF08460), SLAP (PF03217), SLH (PF00395), SPOR (PF05036), TMP (PF05017), YSIRK\_Signal (PF04650).

GBS bacteriophage B30 endolysin shows both glycosidase and endopeptidase activities resulting from independent action of Acm (Glyco\_hydro\_25) and CHAP catalytic domains [56]. Nevertheless, the combined action of different catalytic domains is also possible. The full LysWMY enzyme of the *Staphylococcus warneri* phage is more active than the isolated CHAP or amidase domains, suggesting that they cooperate to cleave cell walls of intact cells [79].

The level of activity of hydrolytic enzymes not only results from the efficiency of catalytic domains, but is also controlled by cell wall binding domains. Their number may be essential for efficient PGH binding to the cell wall [7,32]. For instance, the copy number of the choline binding domain (ChBD/CW\_binding\_1), LysM and SH3, has been demonstrated to be important for high catalytic efficiency of the PGHs: their inactivation leads to a sharp reduction [20,25,69] or even abolition of PGH activity [59]. Targeting domains are also responsible for the PGH binding specificity; for example, ChBD of *S. pneumoniae* LytA promotes its binding to choline within the pneumococcal

cell wall [58]. The SH3 domain of *S. aureus* ALE-1 binds specifically to staphylococcal peptidoglycan bonds [39].

The diversity of the PGHs is reflected not only in the huge variety of their catalytic and surface association domains, but also in the multiplicity of their associations. To explain PGH diversity, evolution by modular recombination was hypothesized and further confirmed by in vitro construction of functional chimeric PGHs [12,16,17]. The creation of Cse, the major *S. thermophilus* cell separation enzyme, results from natural domain shuffling events combining the LysM cell wall binding module of the *sip* gene and the CHAP catalytic module of the *pcsB* gene [6]. Atl, a natural chimera, also results from evolution of multimodular enzymes by gene fusion and gene duplication [52].

#### 4. Firmicutes genomes encode different sets of PGHs

Table 1 illustrates domains encountered within Firmicutes proteins and shows that each genus encodes at least one





oligosaccharides in its environment to provide carbon sources [11]. In addition, PGH-associated domains were also demonstrated to play roles other than binding to the cell surface. Thus, the YSIRK-G/S domain of *Staphylococcus* protein A is not essential for anchoring to the cell surface, but is required for efficient protein A secretion [1].

## 5. High interspecific variability of PGH number

In silico analysis revealed the presence of multiple PGH-encoding genes within each Firmicutes genomes except for those of *Mycoplasma* bacteria deprived of peptidoglycan. As

described in Table 2, every genome encodes at least one (and often several) PGHs of each class. One remarkable feature is the intergeneric disparity of the PGH number between strains. The lactic acid bacteria *Streptococcus* and *Lactobacillus* are characterized by a relatively small number of PGHs. In contrast, *Bacillus* and *Clostridium* are better supplied. In general, bacteria displaying a complex life cycle encode numerous PGHs.

Why this need for so many PGHs? The huge number of PGHs probably reflects the plurality of their functions, since they participate in numerous biological processes. These include cellular growth and division that require cell wall expansion,

Table 3  
*Streptococcus* PGHs potentially involved in daughter cell separation

| Protein (Gene) name                          | Accession number | Predicated activity  | Modular structure | Size (aa) | SP |
|--|------------------|--|-------------------|-----------|----|
| <i>Streptococcus agalactiae</i><br>SAM_0643  | ZP_00787586      | <i>N</i> -acetylglucosaminidase and/or<br><i>N</i> -acetylmuramoyl-L-alanine amidase |                   | 466       | ND |
| <i>Streptococcus downei</i><br><i>atlh</i>   | BAF76789         | <i>N</i> -acetylmuramidase   |                   | 879       | +  |
| <i>Streptococcus dysgalactiae</i><br>/       | ABV55414         | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>and/or Endopeptidase                   |                   | 489       | ND |
| <i>Streptococcus gordinii</i><br>orfO        | AAN64579         | <i>N</i> -acetylmuramidase   |                   | 1366      | +  |
| SGO_2094                                     | YP_001451342     | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 549       | +  |
| <i>Streptococcus intermedius</i><br>orf1     | BAB61101         | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 226       | +  |
| <i>Streptococcus mitis</i><br><i>lytB</i>    | CAI34856         | <i>N</i> -acetylglucosaminidase  |                   | 568       | +  |
| <i>Streptococcus rattii</i><br>/             | BAB88829         | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 125       | ND |
| <i>Streptococcus pyogenes</i><br>/           | AAU29461         | <i>N</i> -acetylglucosaminidase and/or<br>Endopeptidase                              |                   | 398       | ND |
| <i>Streptococcus sanguinis</i><br>SSA_0304   | YP_001034311     | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 221       | +  |
| SSA_0036                                     | YP_001034053     | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 649       | +  |
| SSA_0164                                     | YP_001034175     | <i>N</i> -acetylglucosaminidase and/or<br>Endopeptidase                              |                   | 479       | +  |
| <i>Streptococcus sobrinus</i><br><i>atlg</i> | BAF76791         | <i>N</i> -acetylmuramidase   |                   | 863       | +  |
| <i>Streptococcus suis</i><br>SSU98_1309      | YP_001200867     | <i>N</i> -acetylmuramoyl-L-alanine amidase   |                   | 1035      | +  |
| SsuiDRAFT_1874                               | ZP_00876248      | <i>N</i> -acetylmuramoyl-L-alanine amidase<br>or Endopeptidase                       |                   | 304       | +  |

PGHs displaying one or several of the following catalytic domains, “CHAP, NLPC/P60, amidase, glyco-hydro\_25 and glucosaminidase” were selected. These domains were encountered in the previously described cell-separating enzymes. Protein activity was predicted according to the potential enzymatic activity of its catalytic domain(s).

Proteins identifiers are given as EMBL/GenBank accession number. The proteins were submitted to a search for additional domains using the Pfam database (<http://www.sanger.ac.uk/Software/Pfam/>) [21] and for signal peptide (SP) using the SignalPHMM server (<http://www.ebi.ac.uk/InterProScan/>) [3]. For more information on cell wall binding domains, see Pfam using the following accession numbers: LysM (PF01476), GBS Bsp-like (PF08481), CW\_binding\_1 (PF01473) and SH3\_5 (PF08460). ND = not determined.

peptidoglycan turnover, recycling and maturation and daughter cell separation [66,67,72]. However, PGHs are also involved in cell lysis and peptidoglycan remodelling in response to environmental factors [34,65,67]. In addition, some PGHs are required for cellular differentiation. In *B. subtilis*, several PGHs (LytC, CwIC, CwID, CwIJ, SleB) have been implicated in spore formation and germination [5,31,46,47,55,63,68]. There is also evidence that some PGHs contribute to the pathogenicity of Gram-positive bacteria either by facilitating release of potent pro-inflammatory agents such as cell wall components or toxins [9,13,18] or by mediating bacterial adherence [36,44]. Finally, other roles have been attributed to PGHs, such as motility, competence at genetic transformation and protein secretion (for a review, see Ref. [67]).

Moreover, there is also intragenic disparity in the PGH number that possibly reflects the different biological properties (virulence, competence, motility, etc.) of the species.

## 6. Cell-separating PGHs are species-specific

Although the process of daughter cell disconnection remains to be investigated, PGHs have been demonstrated to be key actors in this process [8,14,24,26,31,41,42,51]. Fig. 3 schematizes all cell separation enzymes identified thus far within the Firmicutes: they are all PGHs. Most are composed of a unique CHAP, NLPC/P60 or a glucosaminidase (more rarely glyco-<sub>L</sub>hydro-<sub>25</sub>) catalytic domain (Fig. 3), and of several identical copies of a cell wall binding domain. The most frequent binding domain is LysM, sometimes GBS Bsp-like, CW\_binding or SH3 that probably drive the protein at a specific location. Cell disconnecting enzymes present, as a common characteristic, a location restricted to the site for cell separation [14,24,33,69,76,78]. The targeting of PGHs to their specific location is probably an essential factor for avoiding cell lyses while assuming daughter cell disconnection.

Another interesting aspect is that each cell-separating enzyme is encoded only by a small group of related bacteria and is, in most cases, species-specific. For instance, Cse is encoded by *S. thermophilus* and absent from pathogenic streptococci. Cell separation of *S. pneumoniae* is assumed by LytB [26] and that of *S. mutans* by Aml [80], with the presence of both enzymes restricted in one species. Why were distinct PGH genes recruited in distinct species?

One explanation for the need for species-specific PGHs is that distinct species would have different cell wall composition at the poles, thus necessitating the action of distinct cell-separating enzymes. Recruitment of species-specific cell-separating enzymes would therefore allow each bacterial species to specifically control the length of its chains without interfering with that of other bacteria. The strong substrate specificity of Aml that selectively lyses *S. mutans* and *S. sobrinus* but no other oral streptococci is in favor of this hypothesis [80].

## 7. Concluding remarks

To identify potential cell-separating enzymes encoded by *Streptococcus* genomes, all genes encoding catalytic and

binding domains characteristic of these proteins were retrieved and are presented in Table 3. Few candidates were recovered with, as expected, at least one per genome (Table 3). Of course, this list is not exhaustive and further experimental evidence will be useful to determine the exact biological function of the candidates. Nevertheless, these data confirm the diversity of cell-separating enzymes recruited by *Streptococci*. Moreover, most of the candidate genes were found in a unique species, corroborating species specificity.

## Acknowledgements

This study was supported by grants from the Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche.

## References

- [1] Bae, T., Schneewind, O. (2003) The YSIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing. *J. Bacteriol.* 185, 2910–2919.
- [2] Bateman, A., Holden, M.T., Yeats, C. (2005) The G5 domain: a potential *N*-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics* 21, 1301–1303.
- [3] Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- [4] Biswas, R., Voggu, L., Simon, U.K., Hentschel, P., Thumm, G., Gotz, F. (2006) Activity of the major staphylococcal autolysin Atl. *FEMS Microbiol. Lett.* 259, 260–268.
- [5] Boland, F.M., Atrih, A., Chirakkal, H., Foster, S.J., Moir, A. (2000) Complete spore-cortex hydrolysis during germination of *Bacillus subtilis* 168 requires SleB and YpeB. *Microbiology* 146, 57–64.
- [6] Borges, F., Layec, S., Thibessard, A., Fernandez, A., Gintz, B., Hols, P., et al. (2005) *cse*, a chimeric and variable gene, encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. *J. Bacteriol.* 187, 2737–2746.
- [7] Braun, L., Dramsi, S., Dehoux, P., Bierne, H., Lindahl, G., Cossart, P. (1997) InlB: an invasion protein of *Listeria monocytogenes* with a novel type of surface association. *Mol. Microbiol.* 25, 285–294.
- [8] Buist, G., Kok, J., Leenhouts, K.J., Dabrowska, M., Venema, G., Haandrikman, A.J. (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.
- [9] Canvin, J.R., Marvin, A.P., Sivakumaran, M., Paton, J.C., Boulnois, G.J., Andrew, P.W., et al. (1995) The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 *pneumococcus*. *J. Infect. Dis.* 172, 119–123.
- [10] Carroll, S.A., Hain, T., Technow, U., Darji, A., Pashalidis, P., Joseph, S.W., et al. (2003) Identification and characterization of a peptidoglycan hydrolase, MurA, of *Listeria monocytogenes*, a muramidase needed for cell separation. *J. Bacteriol.* 185, 6801–6808.
- [11] Clarke, V.A., Platt, N., Butters, T.D. (1995) Cloning and expression of the beta-*N*-acetylglucosaminidase gene from *Streptococcus pneumoniae*. Generation of truncated enzymes with modified aglycon specificity. *J. Biol. Chem.* 270, 8805–8814.
- [12] Croux, C., Ronda, C., Lopez, R., Garcia, J.L. (1993) Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Mol. Microbiol.* 9, 1019–1025.
- [13] De Kimpe, S.J., Kengatharan, M., Thiernemann, C., Vane, J.R. (1995) The cell wall components peptidoglycan and lipoteichoic acid from *Staphylococcus aureus* act in synergy to cause shock and multiple organ failure. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10359–10363.

- [14] De Las Rivas, B., Garcia, J.L., Lopez, R., Garcia, P. (2002) Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J. Bacteriol.* 184, 4988–5000.
- [15] Dhalluin, A., Bourgeois, I., Pestel-Caron, M., Camiade, E., Raux, G., Courtin, P., et al. (2005) Acd, a peptidoglycan hydrolase of *Clostridium difficile* with N-acetylglucosaminidase activity. *Microbiology* 151, 2343–2351.
- [16] Diaz, E., Lopez, R., Garcia, J.L. (1990) Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8125–8129.
- [17] Diaz, E., Lopez, R., Garcia, J.L. (1991) Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J. Biol. Chem.* 266, 5464–5471.
- [18] Diaz, E., Lopez, R., Garcia, J.L. (1992) Role of the major pneumococcal autolysin in the atypical response of a clinical isolate of *Streptococcus pneumoniae*. *J. Bacteriol.* 174, 5508–5515.
- [19] Donovan, D.M., Dong, S., Garrett, W., Rousseau, G.M., Moineau, S., Pritchard, D.G. (2006) Peptidoglycan hydrolase fusions maintain their parental specificities. *Appl. Environ. Microbiol.* 72, 2988–2996.
- [20] Eckert, C., Lecerf, M., Dubost, L., Arthur, M., Mesnage, S. (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. *J. Bacteriol.* 188, 8513–8519.
- [21] Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., et al. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.* 34, D247–251.
- [22] Firczuk, M., Bochtler, M. (2007) Folds and activities of peptidoglycan amidases. *FEMS Microbiol. Rev.* 31, 676–691.
- [23] Fujiwara, T., Aoki, S., Komatsuzawa, H., Nishida, T., Ohara, M., Suganaka, H., et al. (2005) Mutation analysis of the histidine residues in the glycyglycine endopeptidase ALE-1. *J. Bacteriol.* 187, 480–487.
- [24] Fukushima, T., Afkham, A., Kurosawa, S., Tanabe, T., Yamamoto, H., Sekiguchi, J. (2006) A new D,L-endopeptidase gene product, YojL (renamed CwIS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis*. *J. Bacteriol.* 188, 5541–5550.
- [25] Garcia, J.L., Diaz, E., Romero, A., Garcia, P. (1994) Carboxy-terminal deletion analysis of the major pneumococcal autolysin. *J. Bacteriol.* 176, 4066–4072.
- [26] Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R., Garcia, J.L. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol. Microbiol.* 31, 1275–1277.
- [27] Ghuyssen, J.M. (1968) Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriol. Rev.* 32, 425–464.
- [28] Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., et al. (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* 11, 1239–1249.
- [29] Holtje, J.V. (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 62, 181–203.
- [30] Hourdou, M.L., Duez, C., Joris, B., Vacheron, M.J., Guinand, M., Michel, G., et al. (1992) Cloning and nucleotide sequence of the gene encoding the gamma-D-glutamyl-L-diamino acid endopeptidase II of *Bacillus sphaericus*. *FEMS Microbiol. Lett.* 70, 165–170.
- [31] Ishikawa, S., Hara, Y., Ohnishi, R., Sekiguchi, J. (1998) Regulation of a new cell wall hydrolase gene, cwlF, which affects cell separation in *Bacillus subtilis*. *J. Bacteriol.* 180, 2549–2555.
- [32] Janecek, S., Svensson, B., Russell, R.R. (2000) Location of repeat elements in glucanucrases of *Leuconostoc* and *Streptococcus* species. *FEMS Microbiol. Lett.* 192, 53–57.
- [33] Kajimura, J., Fujiwara, T., Yamada, S., Suzawa, Y., Nishida, T., Oyamada, Y., et al. (2005) Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of *Staphylococcus aureus*. *Mol. Microbiol.* 58, 1087–1101.
- [34] Koch, A.L. (2000) The bacterium's way for safe enlargement and division. *Appl. Environ. Microbiol.* 66, 3657–3663.
- [35] Kohler, A., Hubert, A., Vogel, M., Goebel, W. (1991) Expression of the *iap* gene coding for protein p60 of *Listeria monocytogenes* is controlled on the posttranscriptional level. *J. Bacteriol.* 173, 4668–4674.
- [36] Kuhn, M., Goebel, W. (1989) Identification of an extracellular protein of *Listeria monocytogenes* possibly involved in intracellular uptake by mammalian cells. *Infect. Immun.* 57, 55–61.
- [37] Layec, S., Decaris, B., Leblond-Bourget, N. (2008) Characterization of proteins belonging to the CHAP-related superfamily within the Firmicutes. *J. Mol. Microbiol. Biotechnol.* 14, 31–40.
- [38] Llull, D., Lopez, R., Garcia, E. (2006) SkI, a novel choline-binding N-acetylmuramoyl-L-alanine amidase of *Streptococcus mitis* SK137 containing a CHAP domain. *FEBS Lett.* 580, 1959–1964.
- [39] Lu, J.Z., Fujiwara, T., Komatsuzawa, H., Sugai, M., Sakon, J. (2006) Cell wall-targeting domain of glycyglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *J. Biol. Chem.* 281, 549–558.
- [40] Machata, S., Hain, T., Rohde, M., Chakraborty, T. (2005) Simultaneous deficiency of both MurA and p60 proteins generates a rough phenotype in *Listeria monocytogenes*. *J. Bacteriol.* 187, 8385–8394.
- [41] Margot, P., Pagni, M., Karamata, D. (1999) *Bacillus subtilis* 168 gene *lytF* encodes a gamma-D-glutamate-meso-diaminopimelate mureopeptidase expressed by the alternative vegetative sigma factor, sigmaD. *Microbiology* 145, 57–65.
- [42] Margot, P., Wahlen, M., Gholamhoseinian, A., Piggot, P., Karamata, D. (1998) The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. *J. Bacteriol.* 180, 749–752.
- [43] McLaughlan, A.M., Foster, S.J. (1998) Molecular characterization of an autolytic amidase of *Listeria monocytogenes* EGD. *Microbiology* 144, 1359–1367.
- [44] Milohanic, E., Jonquieres, R., Cossart, P., Berche, P., Gaillard, J.L. (2001) The autolysin Ami contributes to the adhesion of *Listeria monocytogenes* to eukaryotic cells via its cell wall anchor. *Mol. Microbiol.* 39, 1212–1224.
- [45] Mishima, M., Shida, T., Yabuki, K., Kato, K., Sekiguchi, J., Kojima, C. (2005) Solution structure of the peptidoglycan binding domain of *Bacillus subtilis* cell wall lytic enzyme CwlC: characterization of the sporulation-related repeats by NMR. *Biochemistry* 44, 10153–10163.
- [46] Moriyama, R., Hattori, A., Miyata, S., Kudoh, S., Makino, S. (1996) A gene (*sleB*) encoding a spore cortex-lytic enzyme from *Bacillus subtilis* and response of the enzyme to L-alanine-mediated germination. *J. Bacteriol.* 178, 6059–6063.
- [47] Moriyama, R., Kudoh, S., Miyata, S., Nonobe, S., Hattori, A., Makino, S. (1996) A germination-specific spore cortex-lytic enzyme from *Bacillus cereus* spores: cloning and sequencing of the gene and molecular characterization of the enzyme. *J. Bacteriol.* 178, 5330–5332.
- [48] Morlot, C., Noirclerc-Savoie, B., Zapun, A., Dideberg, O., Vernet, T. (2004) The D,D-carboxypeptidase PBP3 organizes the division process of *Streptococcus pneumoniae*. *Mol. Microbiol.* 51, 1641–1648.
- [49] Morlot, C., Pernot, L., Le Gouellec, A., Di Guilmi, A.M., Vernet, T., Dideberg, O., et al. (2005) Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from *Streptococcus pneumoniae*. *J. Biol. Chem.* 280, 15984–15991.
- [50] Navarre, W.W., Ton-That, H., Faull, K.F., Schneewind, O. (1999) Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity. *J. Biol. Chem.* 274, 15847–15856.
- [51] Ohnishi, R., Ishikawa, S., Sekiguchi, J. (1999) Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J. Bacteriol.* 181, 3178–3184.
- [52] Oshida, T., Sugai, M., Komatsuzawa, H., Hong, Y.M., Suganaka, H., Tomasz, A. (1995) A *Staphylococcus aureus* autolysin that has an N-acetylmuramoyl-L-alanine amidase domain and an endo-beta-N-acetylglucosaminidase domain: cloning, sequence analysis, and characterization. *Proc. Natl. Acad. Sci. U.S.A.* 92, 285–289.
- [53] Perez-Dorado, I., Campillo, N.E., Monterroso, B., Heseck, D., Lee, M., Paez, J.A., et al. (2007) Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. *J. Biol. Chem.* 282, 24990–24999.
- [54] Popham, D.L., Gilmore, M.E., Setlow, P. (1999) Roles of low-molecular-weight penicillin-binding proteins in *Bacillus subtilis* spore peptidoglycan synthesis and spore properties. *J. Bacteriol.* 181, 126–132.



- [55] Popham, D.L., Helin, J., Costello, C.E., Setlow, P. (1996) Muramic lactam in peptidoglycan of *Bacillus subtilis* spores is required for spore outgrowth but not for spore dehydration or heat resistance. Proc. Natl. Acad. Sci. U.S.A. 93, 15405–15410.
- [56] Pritchard, D.G., Dong, S., Baker, J.R., Engler, J.A. (2004) The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. Microbiology 150, 2079–2087.
- [57] Reinscheid, D.J., Stosser, C., Ehlert, K., Jack, R.W., Moller, K., Eikmanns, B.J., et al. (2002) Influence of proteins Bsp and FemH on cell shape and peptidoglycan composition in group B *streptococcus*. Microbiology 148, 3245–3254.
- [58] Sanchez-Puelles, J.M., Sanz, J.M., Garcia, J.L., Garcia, E. (1990) Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. Gene 89, 69–75.
- [59] Sass, P., Bierbaum, G. (2007) Lytic activity of recombinant bacteriophage phi11 and phi12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. Appl. Environ. Microbiol. 73, 347–352.
- [60] Scheffers, D.J., Pinho, M.G. (2005) Bacterial cell wall synthesis: new insights from localization studies. Microbiol. Mol. Biol. Rev. 69, 585–607.
- [61] Schleifer, K.H. (1986) Gram-positive cocci. In P.H.A. Sneath, N.S. Mair, M.E. Sharp, & J.G. Holt (Eds.), *Bergey's Manual of Systematic Bacteriology* (pp. 999–1002). Baltimore: The Williams & Wilkins Co.
- [62] Schleifer, K.H., Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol. Rev. 36, 407–477.
- [63] Sekiguchi, J., Akeo, K., Yamamoto, H., Khasanov, F.K., Alonso, J.C., Kuroda, A. (1995) Nucleotide sequence and regulation of a new putative cell wall hydrolase gene, *cwlD*, which affects germination in *Bacillus subtilis*. J. Bacteriol. 177, 5582–5589.
- [64] Shockman, G.D., Barrett, J.F. (1983) Structure, function, and assembly of cell walls of gram-positive bacteria. Annu. Rev. Microbiol. 37, 501–527.
- [65] Shockman, G.D., Daneo-Moore, L., Kariyama, R., Massidda, O. (1996) Bacterial walls, peptidoglycan hydrolases, autolysins, and autolysis. Microb. Drug. Resist. 2, 95–98.
- [66] Shockman, G.D., Holtje, J.V. (1994) Microbial peptidoglycan (murein) hydrolases. In J.M. Ghuysen, & R. Hakenbeck (Eds.), *Bacterial Cell Wall* (pp. 131–168). London: Elsevier Science.
- [67] Smith, T.J., Blackman, S.A., Foster, S.J. (2000) Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. Microbiology 146, 249–262.
- [68] Smith, T.J., Foster, S.J. (1995) Characterization of the involvement of two compensatory autolysins in mother cell lysis during sporulation of *Bacillus subtilis* 168. J. Bacteriol. 177, 3855–3862.
- [69] Steen, A., Buist, G., Horsburgh, G.J., Venema, G., Kuipers, O.P., Foster, S.J., et al. (2005) AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. Febs J. 272, 2854–2868.
- [70] Steen, A., Buist, G., Leenhouts, K.J., El Khattabi, M., Grijpstra, F., Zomer, A.L., et al. (2003) Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. J. Biol. Chem. 278, 23874–23881.
- [71] Sugai, M., Fujiwara, T., Akiyama, T., Ohara, M., Komatsuzawa, H., Inoue, S., et al. (1997) Purification and molecular characterization of glycyglycine endopeptidase produced by *Staphylococcus capitis* EPK1. J. Bacteriol. 179, 1193–1202.
- [72] Ward, J.B., Williamson, R. (1984) Bacterial autolysins: specificity and function. In C. Nombela (Ed.) (pp. 156–166). Amsterdam: Elsevier.
- [73] Whisstock, J.C., Lesk, A.M. (1999) SH3 domains in prokaryotes. Trends Biochem. Sci. 24, 132–133.
- [74] Wilkin, J.M. (2004) Penicillin-binding protein 4. In A.J. Barrett, N.D. Rawlings, & J.F. Woessner (Eds.), *Handbook of Proteolytic Enzymes*, 2 (pp. 1967–1970). London: Elsevier.
- [75] Wuenschel, M.D., Kohler, S., Bubert, A., Gerike, U., Goebel, W. (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.
- [76] Yamada, S., Sugai, M., Komatsuzawa, H., Nakashima, S., Oshida, T., Matsumoto, A., et al. (1996) An autolysin ring associated with cell separation of *Staphylococcus aureus*. J. Bacteriol. 178, 1565–1571.
- [77] Yamaguchi, H., Furuhashi, K., Fukushima, T., Yamamoto, H., Sekiguchi, J. (2004) Characterization of a new *Bacillus subtilis* peptidoglycan hydrolase gene, *yvcE* (named *cwlO*), and the enzymatic properties of its encoded protein. J. Biosci. Bioeng. 98, 174–181.
- [78] Yamamoto, H., Kurosawa, S., Sekiguchi, J. (2003) Localization of the vegetative cell wall hydrolases LytC, LytE, and LytF on the *Bacillus subtilis* cell surface and stability of these enzymes to cell wall-bound or extracellular proteases. J. Bacteriol. 185, 6666–6677.
- [79] Yokoi, K.J., Kawahigashi, N., Uchida, M., Sugahara, K., Shinohara, M., Kawasaki, K., et al. (2005) The two-component cell lysis genes *holWMY* and *lysWMY* of the *Staphylococcus warneri* M phage varphiWMY: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. Gene 351, 97–108.
- [80] Yoshimura, G., Komatsuzawa, H., Hayashi, I., Fujiwara, T., Yamada, S., Nakano, Y., et al. (2006) Identification and molecular characterization of an N-Acetylmuraminidase, Aml, involved in *Streptococcus mutans* cell separation. Microbiol. Immunol. 50, 729–742.



## Actes de congrès

- 2006 : Congrès nationale, 14<sup>ème</sup> Colloque du Club des Bactéries Lactiques à Paris

---

### Implication des domaines LysM, CHAP et de la région centrale du gène *cse* dans l'activité de séparation cellulaire de *Streptococcus thermophilus*.

Séverine Layec, Frédéric Borges, Bernard Decaris et Nathalie Leblond-Bourget.

#### Communication orale.

---

*Streptococcus thermophilus* est une bactérie lactique qui se développe en chaîne de cellules. Un gène chimérique et polymorphe, *cse* (cell segregation) a été identifié chez *S. thermophilus*. Chez la souche CNRZ368, il code une protéine extracellulaire de 461 acides aminés qui est impliquée dans la ségrégation cellulaire.

Cette protéine possède :

- à son extrémité N-terminale, un domaine putatif de liaison au peptidoglycane, précédé d'une séquence signal de 28 acides aminés permettant son exportation ;
- une région centrale polymorphe riche en motifs répétés;
- à son extrémité C-terminale, un domaine catalytique CHAP (Cysteine Histidine dependent Aminohydrolase/Peptidase), présent au sein de nombreuses hydrolases du peptidoglycane.

La particularité de la protéine Cse, est sa structure chimérique. En effet, son extrémité N- terminale est homologue de celle de la protéine SIP (Surface Immunogenic Protein) de *S. agalactiae*, alors que son extrémité C-terminale est homologue à celle de la protéine PcsB de *S. thermophilus*. De plus, la région centrale de Cse ne présente aucune similarité avec les séquences contenues dans les banques de données. L'ensemble de ces résultats suggère que la création du gène *cse* résulterait d'un mécanisme de réarrangement de domaines qui aurait conduit à sa formation au sein de l'espèce *S. thermophilus*.

La nature chimérique de *cse*, pose la question du rôle de chacun de ces domaines. La délétion des domaines LysM et CHAP du gène *cse* chez la souche CNRZ368 a révélé un phénotype « longue chaîne de cellules » mettant ainsi en évidence leur importance dans l'activité de ségrégation cellulaire. De plus, la délétion de la région centrale conduit à une réduction, mais non à une inhibition totale, suggérant qu'elle n'est pas essentielle à l'activité de ségrégation cellulaire. Le rôle de cette région est méconnu, son polymorphisme suggère une évolution rapide de ce gène pouvant jouer un rôle dans l'adaptation aux conditions environnementales, et en particulier sur l'adhésion de *S. thermophilus* aux surfaces biotiques et abiotiques.

- 2007 : Congrès nationale, 15<sup>ème</sup> Colloque du Club des Bactéries Lactiques à Rennes

---

**Le gène *cse* code une hydrolase du peptidoglycane indispensable à la séparation des cellules de *Streptococcus thermophilus*.**

Séverine Layec, Joëlle Gérard, Valérie Legué, Christine Delorme, Bernard Decaris et Nathalie Leblond-Bourget.

Communication orale.

---

*Streptococcus thermophilus* est une bactérie lactique qui se développe en chaîne de cellules.

La séparation des cellules de *S. thermophilus* est assurée par le gène *cse* (cell separation) qui code une protéine modulaire résultant d'un réarrangement de domaines [1]. A son extrémité C-terminale, Cse possède un domaine catalytique putatif, CHAP (Cysteine Histidine dependent Aminohydrolase/Peptidase), présent au sein de nombreuses hydrolases du peptidoglycane [2].

Dans cette étude, l'activité catalytique de la protéine Cse a été démontrée par zymogramme et celle-ci est conférée par le domaine catalytique CHAP. A son extrémité N-terminale, Cse possède un domaine putatif de liaison au peptidoglycane LysM suggérant que Cse soit une protéine de surface. La localisation de Cse par immunofluorescence et par microscopie électronique à transmission a été étudiée. Cse est localisée à des sites spécifiques de séparation cellulaire : les septa matures.

Ainsi, le gène *cse* code une hydrolase du peptidoglycane qui interviendrait dans une des dernières étapes de la séparation cellulaire de *S. thermophilus*.

De plus, la présence du gène *cse* a été recherchée dans les génomes des streptocoques par analyse *in silico* et par PCR. Les résultats ont montré que *cse* est spécifique de *S. thermophilus* et que celui-ci serait apparu récemment dans le groupe *salivarius*. Par ailleurs, l'évolution rapide du gène *cse* suggère que celui-ci jouerait un rôle dans l'adaptation de *S. thermophilus* aux conditions environnementales, et en particulier sur l'adhésion de *S. thermophilus* aux surfaces biotiques et abiotiques.

**Mots clés:** Hydrolase du peptidoglycane, protéine de surface, septum mature, séparation cellulaire, *Streptococcus thermophilus*

**Références:**

- [1] Frédéric Borges, S., Layec, A., Thibessard, A., Fernandez, B., Gintz, P., Hols, B., Decaris, N., Leblond-bourget, 2005. *cse*, a chimeric and variable gene encodes an extracellular protein involved in cellular segregation in *Streptococcus thermophilus*. Journal of Bacteriology 187, 2737-2746.
- [2] Séverine Layec, B., Decaris, N., Leblond-Bourget, 2008. Characterization of proteins belonging to the CHAP-related superfamily within the *Firmicutes*. JMMB (sous presse)







## ***Résumé***

---

Le gène *cse* est le premier exemple de création récente d'un gène chimérique et fonctionnel apparu au cours de l'évolution du génome de *Streptococcus thermophilus*. Le gène *cse* résulte d'un réarrangement spontané de modules ayant permis d'associer le module LysM de *sip* avec le module CHAP de *pcsB*. La présence de régions de faible complexité, dans les gènes *pcsB* et *sip*, aurait favorisé la création du gène *cse* chez *S. thermophilus*. Ainsi, ces régions de faible complexité ont été définies comme étant le siège des bornes du réassortiment de modules.

Le gène *cse* code la protéine majeure de la séparation cellulaire de *S. thermophilus*. Son inactivation conduit à un phénotype drastique avec des chaînes de cellules 10 à 100 fois plus longues que celles de la souche sauvage. Ce travail a permis de montrer que *cse* code une endopeptidase agissant au niveau des septa matures des cellules de *S. thermophilus*. Ainsi, la séparation cellulaire est un processus actif nécessitant l'action ciblée d'hydrolases du peptidoglycane pour disjoindre les cellules filles. Les données bibliographiques ainsi que ce travail indiquent que les enzymes de séparation cellulaire sont très diverses. Et de façon plus étonnante, ces enzymes sont spécifiques d'espèces, notamment au sein du genre *Streptococcus*. Quelle peut-être la nature des pressions de sélection qui s'exerce sur la diversité et la spécificité des enzymes de séparation cellulaire ?

Le gène *cse* spécifique de *S. thermophilus* suggère que sa sélection confère un avantage adaptatif à *S. thermophilus*. Le fait que *cse* soit impliqué dans la séparation cellulaire et par conséquent, ait un impact sur la longueur des chaînes de cellules, amène à s'interroger sur l'impact de la longueur des chaînes de cellules sur la valeur adaptative de *S. thermophilus*, et notamment dans son éventuelle action probiotique.

### **Mots clés :**

*Streptococcus thermophilus*, *cse*, Réassortiment de modules, fonction nouvelle, hydrolase du peptidoglycane, septum mature, séparation cellulaire.