ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Ecole Doctorale ABIES
Agriculture
Agriculture
Alimentation
Food
Biologie
Biology
Environnement
Environment
Santé
Health

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Microbiologie

présentée et soutenue publiquement par

Betty FLEUCHOT

le 6 décembre 2011

– Les régulateurs transcriptionnels Rgg –

Confirmation de leur implication dans des phénomènes de quorum-sensing et identification de leurs cibles

Jury :

M. Ivan MIJAKOVIC, Professeur, INRA, AgroParisTech	Président
M. Jean-Pierre CLAVERYS, Directeur de Recherche, CNRS	Rapporteur
M. Tarek MSADEK, Chef de laboratoire, Institut Pasteur	Rapporteur
M. Jean GUZZO, Professeur, Université de Bourgogne	Examinateur
M. Didier LERECLUS, Directeur de Recherche, INRA, AgroParisTech	Examinateur
Mme Véronique MONNET, Directeur de Recherche, INRA, AgroParisTech	Directrice de thèse
Mme Rozenn GARDAN, Chargé de Recherche, INRA, AgroParisTech	Directrice des travaux

AgroParisTech
INRA UMR1319 Micalis, Equipe Peptides et Communication Bactérienne
Domaine de Vilvert, Bat. 526 - 78352 Jouy-en-Josas

Remerciements

*Mes remerciements s'adressent en premier lieu aux membres de mon jury de thèse et en particulier à **Ivan Mijakovic** qui a bien voulu accepter de présider le jury de cette thèse, **Jean-Pierre Claverys** et **Tarek Msadek** qui ont si gentiment consenti à juger ce travail en qualité de rapporteurs, ainsi que **Jean Guzzo** et **Didier Lereclus** qui m'ont donné leur aval avec autant de plaisir pour leur rôle d'examinateur. Je les remercie grandement pour le temps qu'ils ont consacré à l'évaluation de mon travail de thèse.*

*Je remercie également les membres de mon comité de suivi de thèse, **Nathalie Leblond-Bourget** et **Didier Lereclus** pour m'avoir permis d'avoir des comités si enrichissants et stimulants ! Merci à vous pour la pertinence de vos remarques, pour vos conseils, pour votre gentillesse et vos encouragements qui auront pleinement contribué à l'avancée de mes travaux.*

Parce qu'une thèse, c'est être encadrée…

*Je tiens à remercier **Véronique Monnet** pour m'avoir accueillie au sein de l'équipe « Peptides et Communication Bactérienne » ; tout particulièrement pour son attention et sa disponibilité tout au long de cette thèse. Véronique, merci de m'avoir permis d'évoluer dans de bonnes conditions grâce à ton regard critique, la confiance que tu m'as témoignée et parce que tu m'as tant transmis de ton expérience ! Merci aussi pour le soutien et la présence que tu m'as apportés lors de la dernière ligne droite.*

*Je tiens également à remercier **Rozenn Gardan** qui m'a encadrée et accompagnée durant mes trois ans de thèse. Merci, Rozenn, pour m'avoir formée au métier de chercheur et appris tant de choses… Merci pour ta patience, ta disponibilité et tes conseils avisés qui m'auront fait grandir d'un point de vue scientifique et également guidée dans le monde professionnel. Merci de m'avoir accordé ta confiance comme tu l'as fait et pour la liberté que tu as su me donner dans mon travail de thèse. Je garderai en mémoire nos discussions scientifiques si enrichissantes et stimulantes… parfois même accompagnées d'une petite pause « chocolat » surprise (c'était toujours très bon !). Merci aussi pour cette oreille attentive – si appréciable – et pour toutes tes petites attentions…*

Parce qu'une thèse, c'est aussi des collaborations…

*Je remercie vivement **Pierre Nicolas** pour son implication dans mon travail de thèse et pour sa collaboration que j'ai tant appréciée. Merci Pierre pour ce regard « extérieur » si constructif que tu as su apporter à mon travail de thèse. Merci aussi pour ta disponibilité, ton écoute et ta patience concernant le « re-moulinage » des données et pour m'avoir permis d'analyser autant de scans !*

*Je remercie également **Jasmina Vidic** pour m'avoir initiée à la SPR. Merci à toi, Jasmina, pour ta bonne humeur et cet enthousiasme constant qui auront rendu mes passages à la VIM bien agréables ! Un grand merci aussi pour ton adaptabilité et ta disponibilité lors de l'attente des peptides… et les re-planifications de manips !*

*Je remercie également la Nancy's team et en particulier **Emmanuelle** qui m'a accueillie lors de mon passage à Nancy et transmis la technique du gel shift. Un clin d'œil aux autres PhD (Romain, Nico, Sarah et Aurore) pour leur accueil, leur gentillesse et pour m'avoir fait découvrir « the » restaurant de fondues !*

Parce qu'une thèse, c'est aussi faire partie d'une équipe…

J'adresse un immense merci aux membres de l'équipe ComBac et de la plateforme PAPPSO ! Les pauses café du matin resteront un bon souvenir de ce bâtiment…

Un merci tout particulier à ***Colette Besset*** *dont je suis reconnaissante pour tous les bons conseils qu'elle m'a apportés. Sans toi Colette les transformations, les étalements, les PCR,… n'auraient pas été aussi sympas à faire ! J'ai été très heureuse de partager mes 3 ans de thèse avec toi. Merci aussi à* ***Alain Guillot*** *avec qui j'ai pu découvrir la LC-MS/MS. Merci pour ton expérience Alain, pour tes réflexions pertinentes et enrichissantes qui auront fait avancer l'histoire des SHP ! Un grand merci aussi à* ***Christophe Gitton*** *pour son aide et sa contribution dans l'avancement de ma thèse. Merci à toi, Christophe, pour ta disponibilité et ta générosité que ce soit pour les manips, pour des problèmes informatiques ou techniques… ton aide aura été bien appréciable ! Un grand merci également à* ***Emilie Chambellon.*** *J'ai été très heureuse de partager le même labo que toi suite au déménagement. Merci beaucoup Emilie pour ton implication, ton aide et tes conseils… et aussi parce que la fin de cette thèse n'aurait pas été pareille sans toi !*

Je tiens également à remercier ***Françoise Rul*** *pour son accueil, sa bonne humeur, ses conseils et pour tous ces bons moments de discussions scientifiques ou autres partagés ensemble. Merci aussi, Françoise, pour avoir contribué à l'obtention de ma bourse de thèse en ayant assisté à mes répètes lors de ma préparation au labo. Je remercie également* ***Vincent Juillard*** *pour son écoute, sa bienveillance et ses relectures critiques de ce manuscrit. Je garde un très bon souvenir de nos conversations, source de conseils et de recommandations. Et puis un clin d'œil particulier à vous deux, Françoise et Vincent pour vos encouragements… (si précieux en Italie… dans le hall pendant une certaine session poster…).*

Une pensée à ***Mireille Yvon****, partie à la retraite pendant ma thèse, que j'ai eu plaisir à connaître et dont je garderai en mémoire la bonne humeur communicative. Je la remercie pour l'intérêt qu'elle a su porter à mon travail de thèse et pour ses conseils d'un point de vue personnel et professionnel. Merci aussi à* ***Didier Chevret*** *pour sa bonne humeur et ses « dérapages » constants ainsi qu'à* ***Céline Henry*** *pour les échanges que nous avons eus ensemble.*

Un clin d'œil aux autres thésards qui, comme moi, ont vadrouillé dans les couloirs du bâtiment 526 : ***Benoit****, pour avoir partagé nos quotidiens de thèse pendant trois ans. Je te remercie pour tous les bons moments scientifiques et autres passés ensemble !* ***Leila****, pour avoir partagé notre quotidien de fin de thèse (et ce n'est pas rien…).* ***Sybille****, pour tes conseils et les discussions que nous pouvions avoir lors de nos petites pauses l'après-midi…* ***Luz****, pour ton amitié et tous les bons moments de franche rigolade passés ensemble. Je vous souhaite à tous une très belle continuation… déjà bien entamée pour certains !*

Enfin, merci aux post-docs (plus ou moins récents au labo), ***Sandeep*** *et* ***David****, pour nos échanges en anglais… pas toujours faciles !*

Et puis, merci aux personnes qui, de près ou de loin, ont fait partie de mon entourage de thèse – en particulier ***Lucy****,* ***Isabelle****,* ***Niriaina****,* ***Mylène****,* ***Sandrine****,* ***Landa****,* ***Virginie****,* ***l'équipe Paroi*** *et* ***l'équipe Flec*** *– pour les bons moments passés ensemble.*

Merci enfin à mon stagiaire, ***Vincent****. J'ai été très heureuse de découvrir l'expérience de l'encadrement à ses côtés !*

Parce qu'une thèse à l'INRA, c'est aussi Doc'J…

*Merci à **Bérénice**, **Aude**, **Poupoune**, et aux **Romain** x2 pour ces bons moments partagés ensemble à Doc'J. J'ai été très heureuse d'écrire une page de l'association à vos côtés !*

Parce qu'une thèse à l'INRA, c'est aussi le théâtre à l'ADAS…

*Une pensée bien particulière aux théâtreux de l'ADAS avec qui j'ai pu partager 3 belles années et m'aérer l'esprit… Un merci tout particulier à **Annie**, **Aline**, **Emmanuelle**, **Olivier**, **Elodie**, **Martine** et bien d'autres pour ces moments de délire et de stress partagés avant les représentations ! Merci aussi à **Romaric** et **Jean-Marc** de nous avoir menés !*

Et enfin, parce qu'une thèse, c'est aussi…

*… le souvenir de 2 ans ½ de colocation, merci à toi **Marie** (sans oublier ta famille et Nico) pour ton accueil si chaleureux à mon arrivée à Versailles, pour ton écoute et toutes les opportunités… gastronomiques et autres partagées ! Je te souhaite plein de bonnes choses pour la suite.*

*… le plaisir de se retrouver entre amis… et d'en profiter ! Je me dois de citer **Ludi**, **Jerem**, **Adrien**, **Milie**, **Mika**, **Vivi**, **Cat**, **Max** et **Cyril**.*

*… être soutenue et encouragée par sa famille : merci à mes **parents** (les 4 !), à mes **frère** et **sœurs** et aux **mamies** !*

… remercier sa Maman pour avoir lu et corrigé autant de fois que je lui ai demandé mes écrits (malgré son incompréhension de mon vocabulaire scientifique). On aura bien ri !

… pour l'anecdote, se faire charrier par ses frère et sœurs quand on mange du fromage ou du yaourt à la maison… « Hummm ! Il est bon Thermophilus ?!!!!!! »…

… être reconnaissante envers ses parents et les remercier d'avoir cru en mes choix ; de m'avoir « accompagnée » tout au long de mon parcours de Clermont-Ferrand à Jouy-en-Josas en passant par Aurillac, Marseille et Dijon ; et de m'avoir permis de grandir dans le monde de la microbiologie. Toutes ces expériences si enrichissantes, je les ai construites grâce à vous…

*Et enfin, parce qu'une thèse c'est aussi le bonheur de t'avoir rencontré… merci à toi, **Aurélien**, pour ta patience, ton soutien, ta compréhension… d'être là.*

Sommaire

Introduction générale **1**

Introduction bibliographique **4**

Chapitre 1 : Les streptocoques **5**
1. Une appartenance au groupe des bactéries lactiques **5**
2. *Streptococcus thermophilus* et les streptocoques **6**
2.1 Classification et relations taxonomiques 6
2.2 Les streptocoques pathogènes stricts et opportunistes 6
2.3 *Streptococcus thermophilus* 9
2.3.1 *S. thermophilus*, un streptocoque d'intérêt industriel 9
2.3.2 Taxonomie et identification de *S. thermophilus* 10
2.3.3 Histoire évolutive de *S. thermophilus* 11
2.3.4 Le transfert de l'information génétique chez *S. thermophilus* 13

Chapitre 2 : La communication cellulaire de type quorum-sensing **17**
1. Concepts généraux **17**
2. Le QS chez les bactéries à Gram positif **18**
2.1 Détection extracellulaire des peptides autoinducteurs 18
2.1.1 Les systèmes à deux composants 18
2.1.2 La virulence chez *Staphylococcus aureus* : le système *agrBDCA* 19
2.1.3 La compétence chez *Streptococcus pneumoniae* : le système *comCDE* 21
2.2 Détection intracellulaire des peptides autoinducteurs 25
2.2.1 Les transporteurs d'oligopeptides 25
2.2.2 Les effecteurs de la famille RNPP 26
2.2.3 Le système PapR/PlcR chez *Bacillus cereus* : virulence 27
2.2.4 Le système peptides phéromones / PrgX chez *Enterococcus faecalis*: conjugaison 31
3. Le QS chez les bactéries à Gram négatif **37**
3.1 Biosynthèse des *N*-AHL : les enzymes de la famille LuxI et LuxM 37
3.2 Caractéristiques des protéines de la famille LuxI 38
3.3 Sécrétion des *N*-AHL 38
3.4 Détection des *N*-AHL 39
3.4.1 Les systèmes *luxIR* : 39
3.4.2 Les systèmes *luxMN* : 40
4. La communication inter-espèces **40**
5. Autres molécules de signalisation connues chez les bactéries **42**
6. Conclusion **43**
6.1 La complexité des réseaux de communication cellulaire 43
6.2 Le QS et les notions de diffusion et efficiency-sensing 44

Chapitre 3 : Les régulateurs transcriptionnels de la famille des Rgg **45**
1. Présentation des protéines régulatrices Rgg **45**
2. Les protéines Rgg « stand-alone » **45**
2.1 Les protéines Rgg : production de glucosyltransférases 45
2.2 Les protéines RopB et RovS : régulation de la virulence 47
2.2.1 Le facteur de virulence SpeB chez *Streptococcus pyogenes* 47
2.2.2 L'adhérence aux cellules épithéliales chez *Streptococcus agalactiae* 49

2.3 Les protéines RggC et GadR : résistance au stress 50
2.3.1 La résistance au stress oxydatif chez *Streptococcus thermophilus* 50
2.3.2 Le système antiport glutamate/GABA chez *Lactococcus lactis* 51
2.4 Les protéines MutR et LasX : production de bactériocines 52
2.4.1 La mutacine I, II et III chez *Streptococcus mutans* 52
2.4.2 La lactocine S chez *Lactobacillus sakei* 54
2.5 Conclusion 54
3. Les protéines Rgg associées à de petits peptides hydrophobes 55
3.1 Détection de petits gènes dans l'environnement proche de gènes *rgg* 56
3.2 Un transporteur d'oligopeptides impliqué dans la réimportation de peptides phéromones chez les streptocoques 57
3.2.1 Hypothèse d'un nouveau mécanisme de QS : les systèmes SHP/Rgg 57
3.2.2 Le contrôle de la compétence au sein du groupe salivarius 58
3.2.3 Le contrôle de la compétence chez les streptocoques du groupe pyogenes et bovis : les systèmes XIP/ComR 60
3.3 Conclusion 61

Objectifs du travail de thèse 62

Résultats 65

Chapitre 1 : Implication des protéines Rgg dans des phénomènes de QS 66
1. Introduction et résultats préliminaires 66
2. Article 1 68
3. Conclusion 69
4. Résultats non publiés 70
4.1 Introduction 70
4.2 Matériel et méthode 70
4.3 Résultats principaux 71
4.4 Discussion 73

Chapitre 2 : Etude de l'interaction SHP/Rgg 75
1. Introduction 75
2. Matériel et méthode 76
3. Résultats 81
4. Discussion 86

Chapitre 3 : Prédiction des régulons SHP/Rgg 90
1. Contexte 90
2. Définitions 90
3. Matériel et méthode 91
4. Résultats 92
5. Discussion 97

Discussion générale et perspectives 102

Références bibliographiques 116

Annexes 129

Table des figures

Figure 1 Relations phylogénétiques liant les genres bactériens majeurs appartenant aux bactéries lactiques (adapté de Stiles and Holzapfel, 1997)..... p 5

Figure 2 Relations phylogénétiques, basées sur l'ARN 16S, liant 33 espèces appartenant au genre *Streptococcus* (adapté de Kawamura *et al.*, 1995a). ..... p 6

Figure 3 Ramifications au sein du groupe salivarius établies par inférence phylogénétique sur la séquence concaténée de l'ARNr 16S et des gènes *recA*, *secA* et *secY* (Pombert *et al.*, 2009). ..... p 10

Figure 4 Dendrogramme représentant les évènements évolutifs au sein de l'ordre des *Lactobacillales* (Makarova *et al.*, 2006). ..... p 11

Figure 5 Diagrammes de Venn représentant le pourcentage de gènes partagés au niveau de différentes espèces ou au sein d'une même espèce de streptocoques (Lefébure and Stanhope, 2007). ..... p 12

Figure 6 Composition du pan-génome de *Streptococcus thermophilus* (Rasmussen *et al.*, 2008). ..... p 12

Figure 7 Les mécanismes de transfert horizontal chez les bactéries. ..... p 13

Figure 8 Représentation schématique du déclenchement d'un système de quorum-sensing (d'après Keller and Surette, 2006). ..... p 17

Figure 9 Représentation schématique d'un système à deux composants. ..... p 18

Figure 10 Représentation schématique du système *agrBCDA* chez *Staphylococcus aureus* (inspiré de Novick and Geisinger, 2008 ; Thoendel and Horswill, 2009). ..... p 20

Figure 11 Représentation schématique du système *comCDE* chez *Streptococcus pneumoniae.* p 22

Figure 12 Organisation génétique des systèmes Com et Blp chez *Streptococcus pneumoniae* et *Streptococcus mutans* (d'après Martin *et al*., 2006). ..... p 24

Figure 13 Voie de biosynthèse des lipoprotéines chez les bactéries à Gram positif (d'après Hutchings *et al*., 2008). ..... p 26

Figure 14 Arbre phylogénétique des protéines OppA chez différents Firmicutes (d'après Nepomuceno *et al.*, 2007)..... p 26

Figure 15 Relations évolutives des protéines de la famille RNPP (Thèse L. Bouillaut, 2007). p 27

Figure 16 Représentation schématique du système PapR/PlcR chez *Bacillus cereus*. ..... p 29

Figure 17 Représentation schématique de l'organisation du domaine HTH du régulateur PlcR en présence ou absence du peptide PapR (Declerck *et al.*, 2007). ..... p 30

Figure 18 Arbre phylogénétique des protéines PlcR et des peptides PapR apparentés (Slamti and Lereclus, 2005). ..... p 31

Figure 19 Représentation schématique du transfert conjugatif pCF10 chez *Enterococcus faecalis.* ..... p 33

Figure 20 Influence de l'oligomérisation du régulateur transcriptionnel PrgX sur l'expression des gènes de conjugaison (adapté de Dunny, 2007). ..... p 34

Figure 21 Contrôle post-transcriptionnel de l'expression des gènes de conjugaison (adapté de Dunny and Johnson, 2011 et Bae *et al.*, 2002). ..... p 34

Figure 22 Biosynthèse de la sex-phéromone cCF10 à partir du produit du gène *ccfA* chez *Enterococcus faecalis* (d'après Antiporta and Dunny, 2002 ; Chandler *et al.*, 2008).p 35

Figure 23 Exemple d'une voie de biosynthèse des *N*-acyl homoserines lactones par les enzymes LuxI (Bassler, 2002). ..... p 37

Figure 24 Représentation schématique des *N*-acyl synthétases de la famille LuxI (adapté de Fuqua and Greenberg, 2002). ..... p 38

Figure 25 Représentation schématique du système *luxIR* contrôlant la bioluminescence chez *Vibrio fischeri* (inspiré de Miller and Bassler, 2001). ..... p 39

Figure 26 Représentation schématique des systèmes de quorum-sensing contrôlant la bioluminescence chez *Vibrio harveyi* (adapté de Ng and Bassler, 2009). ..... p 40

Figure 27 Voie de biosynthèse de l'autoinducteur 2 (Rezzonico and Duffy, 2008). ..... p 41

Figure 28 Paramètres influençant le déclenchement d'une réponse physiologique coordonnée, de type quorum-sensing, chez les bactéries (d'après Hense *et al.*, 2007). ..... p 44

Figure 29 Organisation génétique du locus *gtfG* chez *Streptococcus gordonii* (adapté Vickerman *et al.*, 2001). ..... p 46

Figure 30 Présentation des régions promotrices des gènes *rgg* et *gtfG* de *Streptococcus gordonii*. ..... p 46

Figure 31 Représentation schématique de la région intergénique *ropB*/*speB* chez *Streptococcus pyogenes* (adapté de Neely *et al.*, 2003). ..... p 47

Figure 32 Contrôle du régulon RopB au cours de la croissance de *Streptococcus pyogenes* (inspiré de McIver, 2009). ..... p 48

Figure 33 Organisation génétique des loci codant les voies de biosynthèse de la mutacine II, I et III chez *Streptococcus mutans* (d'après Qi *et al.*, 1999ab ; 2000). ..... p 52

Figure 34 Représentation schématique des régions promotrices des gènes *mutA* et *mutR* impliqués dans la voie de biosynthèse de la mutacine II chez *Streptococcus mutans* (d'après Qi *et al.*, 1999b). ..... p 53

Figure 35 Modèle proposé pour le contrôle de l'expression de la mutacine I par LuxS chez *Streptococcus mutans* (d'après Merritt *et al.*, 2005). ..... p 53

Figure 36 Le locus de la lactocine S chez *Lactobacilles sakei* (d'après Skaugen *et al.*, 2002 et Rawlinson *et al.*, 2002). ..... p 54

Figure 37 Alignement des régions promotrices de plusieurs gènes cibles de régulateurs Rgg (adapté de Rawlinson *et al.*, 2005). ..... p 55

Figure 38 Arbre phylogénétique des protéines Rgg chez les streptocoques (Ibrahim *et al.*, 2007b). ..... p 56

Figure 39 Relations phylogénétiques basées sur l'ARN 16S et distribution du système ComCDE (et de ses paralogues) chez les streptocoques (adapté de Martin *et al.*, 2006). ..... p 58

Figure 40 Représentation schématique des systèmes SHP/Rgg et XIP/ComR chez *Streptococcus thermophilus* LMD-9. ..... p 61

Figure 41 Représentation schématique des différentes étapes intervenant dans un mécanisme de quorum-sensing. ..... p 66

Figure 42 Comparaison du niveau d'expression des promoteurs des gènes *shp1358* et *ster_1357*. ..... p 67

Figure 43 Nombre de protéines Rgg et Rgg-like, associées ou non à des spCDS, chez les bactéries à Gram positif à bas GC% appartenant au taxon Rgg. ..... p 70

Figure 44 Nombre de protéines Rgg et Rgg-like, associées ou non à des spCDS, au sein des différents genres bactériens appartenant au taxon Rgg. ..... p 70

Figure 45 Nombre de protéines Rgg et Rgg-like, associées à des spCDS, au sein du genre *Streptococcus*. ..... p 70

Figure 46 Représentation schématique des plasmides pGICB004 et pGICB004::*aphA3*... p 78

Figure 47 Fusions transcriptionnelles construites pour l'étude de la fonctionnalité des systèmes SHP/Rgg. ..... p 78

Figure 48 Croissance et activité luciférase de souches présentant une fusion P*shp::luxAB* dans leur génome et dans différentes conditions de cultures. ..... p 83

Figure 49 Croissance et activité luciférase de différentes souches de *Streptococcus thermophilus* LMG18311 présentant une fusion $P_{shp0182}$*::luxAB* dans leur génome. p 84

Figure 50 Spectres de fragmentation des ions m/z (799,5 et 872,5) permettant la validation des séquences des phéromones SHP issues des loci *shp/gbs1555* de *Streptococcus agalactiae* NEM316 et *shp/smu.1509* de *Streptococcus mutans* UA159. ..... p 85

Figure 51 SDS-PAGE des fractions protéiques issues des cultures d'*Escherichia coli* permettant la surproduction de protéines Rgg recombinantes. ..... p 86

Figure 52 Arbre de classification reflétant la divergence des sites synonymes au sein des protéines Rgg. ..... p 91

Figure 53 Comparaison des motifs ADN identifiés dans les régions promotrices des gènes cibles du locus *shp/rgg1358* de *Streptococcus thermophilus* LMD-9. ..... p 92

Liste des tableaux

Tableau 1 Histoire de la classification des bactéries et des Archaea (Schleifer, 2009). ..... p 5

Tableau 2 Caractéristiques générales de l'ensemble des génomes séquencés appartenant à l'espèce *Streptococcus thermophilus*. ..... p 11

Tableau 3 Peptides autoinducteurs AgrD connus chez *Staphylococcus aureus*. ..... p 21

Tableau 4 Le peptide CSP de *Streptococcus pneumoniae.* ..... p 23

Tableau 5 Le peptide PapR7 de *Bacillus cereus*. ..... p 29

Tableau 6 Les deux principaux plasmides conjugatifs décrits chez *Enterococcus faecalis*. ..... p 32

Tableau 7 Précurseurs des sex-phéromones (adapté de Clewell *et al.*, 2000). ..... p 35

Tableau 8 Précurseurs des anti-phéromones (adapté de Clewell *et al.*, 2000) ..... p 35

Tableau 9 Caractéristiques des principales molécules de signalisation appartenant à la famille des *N*-AHL chez les bactéries à Gram négatif. ..... p 37

Tableau 10 Les principaux régulateurs transcriptionnels Rgg caractérisés et leur gène cible proximal associé chez les bactéries à Gram positif. ..... p 45

Tableau 11 Prévalence du gène *fbsA* et de ses régulateurs au sein des principaux complexes clonaux de *Streptococcus agalactiae* (adapté de Al Safadi *et al.*, 2011). ..... p 50

Tableau 12 Caractéristiques des peptides « phéromones » SHP et XIP. ..... p 71

Tableau 13 Familles de peptides$_{SHOW}$ identifiées dans le groupe mitis et le genre *Listeria*. ..... p 72

Tableau 14 Peptides$_{SHOW}$ « souche spécifique » présentant des caractéristiques intéressantes. . p 73

Tableau 15 Identification d'une famille de peptides « sans » critère au sein de l'espèce *Streptococcus pneumoniae*. ..... p 73

Tableau 16 Description des souches utilisées dans cette étude. ..... p 76

Tableau 17 Oligonucléotides utilisés dans cette étude. ..... p 77

Tableau 18 Plasmides utilisés dans cette étude. ..... p 77

Tableau 19 Loci *shp/rgg* sélectionnés dans cette étude. ..... p 82

Tableau 20 Pourcentage d'identité des protéines affines AmiA chez *Streptococcus thermophilus* LMD-9 et CNRZ1066. ..... p 88

Tableau 21 Récapitulatif des trois groupes de systèmes SHP/Rgg décrit dans l'article 1. ..... p 90

Tableau 22 Alignements MEME et MUSCLE des références des groupes I, II et III. ..... p 92-93

Tableau 23 Squelettes ADN des motifs putatifs de fixation des régulateurs Rgg à l'ADN identifiés dans les régions promotrices des gènes *shp*. ..... p 92-93

Tableau 24 Environnement génétique des loci *shp/rgg* associés au groupe I. ..... p 94

Tableau 25 Environnement génétique des loci *shp/rgg* associés au groupe II. ..... p 95

Tableau 26 Environnement génétique des loci *shp/rgg* associés au groupe III. ..... p 96

Tableau 27 Environnement génétique des gènes cibles distaux présentant une distance « motif / codon start » conservée. ..... p 97

Tableau 28 Cibles putatives sélectionnées pour validation *in vitro*. ..... p 101

Introduction générale

De par la diversité des niches écologiques qu'elles colonisent, les bactéries possèdent de nombreux mécanismes de réponse leur permettant de s'adapter aux fluctuations environnementales, en particulier en régulant l'expression de gènes spécifiques. A titre d'exemples, la capacité des bactéries à détecter et traduire divers signaux physico-chimiques de l'environnement leur permet de contrôler des fonctions en lien avec le chimiotastisme, le transport du phosphate ou encore de s'adapter à un changement d'osmolarité. Au cours des dernières décennies, il a été admis que les bactéries sont également capables de répondre à des signaux qu'elles synthétisent et sécrètent. Ces petites molécules de signalisation ont été appelées autoinducteurs ou phéromones. Cette découverte de l'existence d'interactions au sein du monde des procaryotes, qui a profondément bouleversé la notion que les bactéries fonctionnent de manière autonome, confère à une population bactérienne la capacité d'initier un comportement collectif en réponse à leur environnement (Atkinson and Williams, 2009).

Les molécules de signalisation sont utilisées dans un processus appelé « quorum-sensing » (QS) et l'une des préoccupations des microbiologistes est de mieux comprendre ce langage, permettant aux bactéries de contrôler différentes fonctions physiologiques telles que la virulence, la formation de biofilms, la production de métabolites secondaires, le transfert horizontal d'ADN, la bioluminescence ou encore la motilité (Di Cagno *et al.*, 2011; Winans, 2006). Aussi, le terme de molécule de signalisation désigne une molécule produite qui induit *in fine* une réponse permettant de coordonner les activités entre des bactéries émettrices du « signal » et des bactéries receveuses ; ces dernières pouvant être à la fois les bactéries émettrices ou provenant d'une autre population bactérienne (Platt and Fuqua, 2010). Dans ce dernier cas, l'effet de la molécule de signalisation peut être positif ou négatif sur la population en question. Aussi, une molécule de signalisation se définit par quatre critères : (i) une production spécifique à un moment donné au cours de la croissance des bactéries et sous certaines conditions physiologiques, (ii) une accumulation dans le milieu extracellulaire et une reconnaissance *via* un senseur spécifique, (iii) la mise en place d'une réponse concertée dès lors qu'une concentration seuil en molécule de signalisation a été atteinte et (iv) un changement physiologique propre à la molécule de signalisation s'étendant au delà des changements physiologiques requis pour métaboliser ou détoxifier la molécule de signalisation en question (Winzer *et al.*, 2002). Aujourd'hui, ce mécanisme de régulation semble effectif dans la plupart des bactéries.

Malgré leur importance et les enjeux économiques qu'elles représentent, les bactéries industrielles, telles que les bactéries lactiques, n'ont encore été que faiblement étudiées dans le domaine de la communication cellulaire, contrairement aux bactéries pathogènes.

Chez *Streptococcus thermophilus*, par exemple, une seule étude concernant la production de bactériocines est décrite dans la littérature. La généralisation de ces systèmes de QS suggérant l'existence de plusieurs mécanismes de QS au sein des espèces bactériennes, la recherche et la caractérisation de nouveaux systèmes de QS chez *S. thermophilus* a été entreprise au laboratoire. Mon travail de thèse s'est inscrit dans cette thématique et a consisté à définir un nouveau mécanisme de QS spécifique du groupe des streptocoques. Aussi, pour la première fois, l'implication de régulateurs transcriptionnels appartenant à la famille des Rgg a été démontrée dans de tels phénomènes.

*** *** ***

Une introduction bibliographique présentera dans un premier temps, les différentes notions scientifiques en lien avec mon sujet de thèse. Un premier chapitre sera consacré à la famille des streptocoques et plus particulièrement à *S. thermophilus*, bactérie modèle dans cette étude. Un second chapitre sera ensuite consacré à la communication cellulaire de type QS et détaillera les principaux systèmes décrits actuellement dans la littérature. Enfin, une troisième partie présentera les données connues sur les protéines régulatrices de la famille Rgg.

Les résultats seront ensuite présentés au sein de trois chapitres. Le premier fait l'objet d'un article publié et regroupe l'ensemble des résultats ayant permis la validation du nouveau système de QS *via* l'étude d'un locus génétique de *S. thermophilus*. Une analyse *in silico* présente, par ailleurs, la distribution de ce système au sein des bactéries à bas GC% et souligne leur spécificité au groupe des streptocoques. Les chapitres suivants présentent, quant à eux, des résultats qui seront prochainement soumis. Aussi, dans un second chapitre, la fonctionnalité de ce nouveau système de QS a été évaluée au sein d'autres espèces de streptocoques. Ce travail ayant permis d'élargir la fonctionnalité de ce mécanisme de QS à des streptocoques pouvant être pathogènes pour l'Homme et les animaux, l'étude de l'existence de phénomènes de cross-talk ou de communication croisée entre des streptocoques pathogènes et/ou non-pathogènes est actuellement en cours au laboratoire. Dans un troisième chapitre sont présentés les résultats d'une approche *in-silico*, développée dans le but de définir les cibles de ces systèmes de QS et de mieux appréhender les fonctions physiologiques qu'ils contrôlent.

Enfin, une discussion générale soulignera l'apport de ce travail pour la communauté scientifique et proposera les perspectives envisageables à ce travail de thèse.

Introduction bibliographique

Tableau 1 – Histoire de la classification des bactéries et des Archaea (Schleifer, 2009)**.**

	Classification basée principalement sur
Fin du 19ème siècle	Morphologie, Besoin spécifique de croissance, Potentiel pathogène
1900 à 1960	Morphologie, Physiologie, Biochimie
1960 à 1980	Chimiotaxonomie, Taxonomie numérique, Hybridation ADN-ADN
1980 à aujourd'hui	Analyses génotypiques, MLSA[1], ANI[2], Analyse du génome

[1] MLSA *Multilocus Sequence Analyses* ; [2] ANI *Average Nucleotide Identity*.

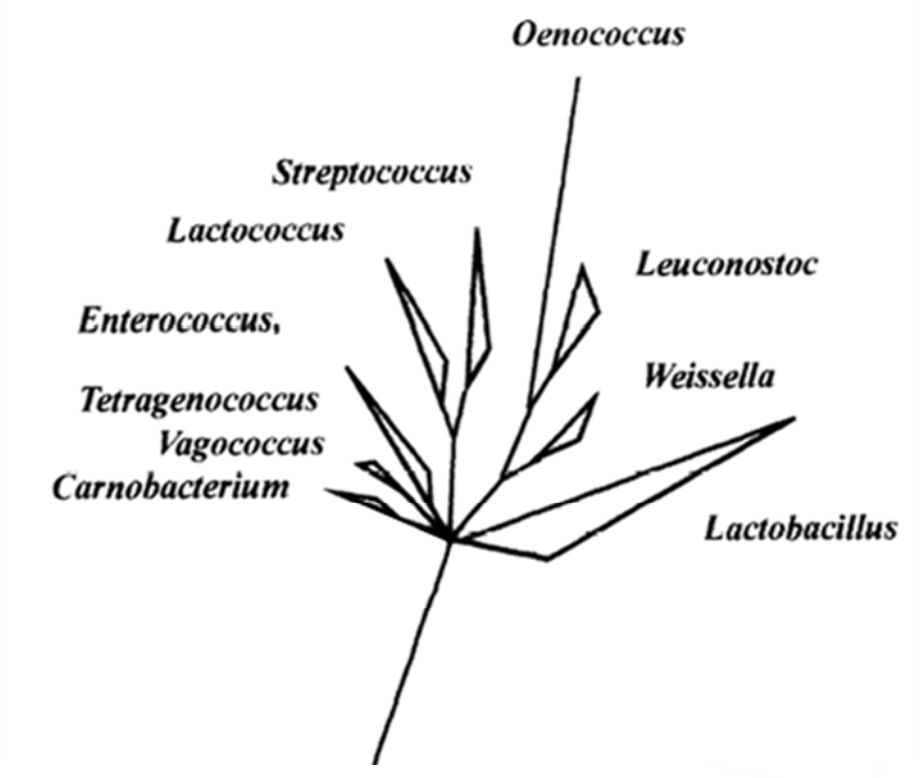


Figure 1 – Relations phylogénétiques liant les genres bactériens majeurs appartenant aux bactéries lactiques (adapté de Stiles and Holzapfel, 1997)**.**

Chapitre 1 : Les streptocoques

1. Une appartenance au groupe des bactéries lactiques

L'appartenance des streptocoques au groupe des bactéries lactiques a été définie par Orla-Jensen, en 1919 (Stiles and Holzapfel, 1997), sur la base de critères phénotypiques. Depuis cette période, de nombreuses techniques (Tableau 1) sont venues bouleverser la classification au sein du groupe des bactéries lactiques et l'appartenance de certains genres bactériens à ce groupe taxonomique a été fortement discutée. En 1996, Vandamme et ses collaborateurs définissent un groupe hétérogène de bactéries composé de 11 genres bactériens : *Carnobacterium*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Oenococcus*, *Pediococcus*, *Streptococcus*, *Tetragenococcus*, *Vagococcus* et *Weissella* (Figure 1). Cette large diversité – regroupant des bactéries d'intérêt industriel, pathogènes stricts ou encore opportunistes – contribue à la présence des bactéries lactiques dans de nombreuses niches écologiques allant de la matrice alimentaire (lait, viande, légume, vin) aux muqueuses de l'Homme (cavité orale, vagin et tractus gastro-intestinal) (Pfeiler and Klaenhammer, 2007). A l'heure actuelle, ce groupe taxonomique est composé d'environ 400 espèces et une controverse demeure concernant les espèces du genre *Lactobacillus*, *Pediococcus*, *Oenococcus* et *Leuconostoc* (Zhang *et al.*, 2011).

D'un point de vue phylogénétique, les bactéries lactiques sont des bactéries à Gram positif (coques ou bacilles), anaérobies aérotolérantes, dépourvues de catalase, immobiles, asporulées et possédant un bas pourcentage en bases G+C (inférieur à 55%) (Stiles and Holzapfel, 1997). En dehors de ces caractères communs, l'appartenance au groupe des bactéries lactiques porte en réalité sur leur capacité à fermenter les glucides en produisant de l'acide lactique. Aussi, la fermentation est dite « homolactique » quand l'acide lactique est le principal produit formé et « hétérolactique » lorsque d'autres composés tels que l'acide acétique, l'éthanol ou le dioxyde de carbone (CO_2) sont également produits (Liu, 2003).

De par l'importance économique de certaines souches en tant que ferments, l'utilisation du terme « bactéries lactiques » a tendance à être restreint aux bactéries d'intérêt industriel. Dans ce cas, la voie métabolique de l'utilisation des oses est à la base de leur intérêt en assurant la fermentation de nombreux produits alimentaires (laitiers, carnés, céréaliers) et favorisant *in fine* le bon développement des produits d'un point de vue conservation ou organoleptique (arômes, textures). *Sensu largo*, un bon nombre de souches restent, quant à

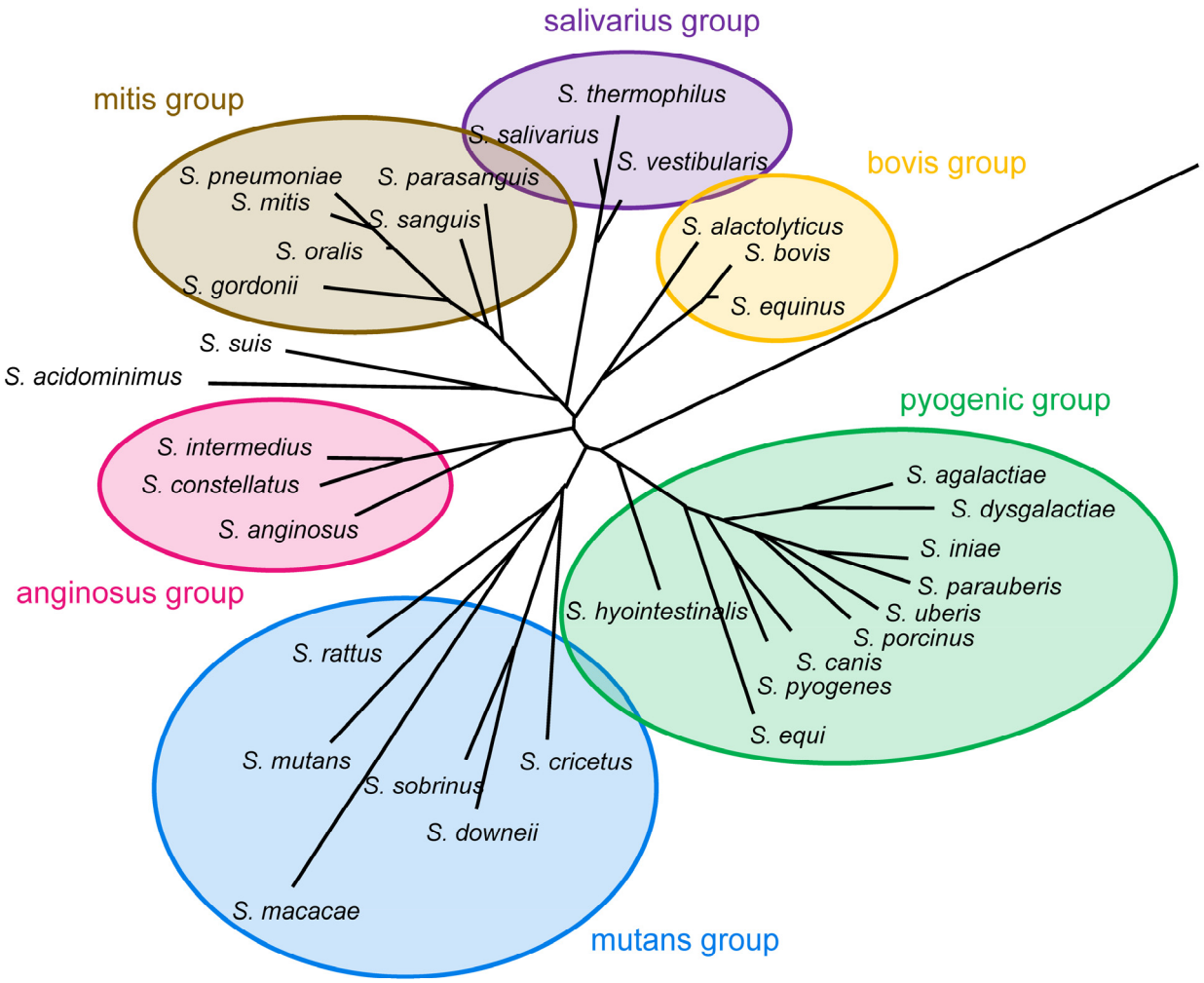


Figure 2 – Relations phylogénétiques, basées sur l'ARN 16S, liant 33 espèces appartenant au genre *Streptococcus* (adapté de Kawamura *et al.*, 1995a)**.**

elles, impliquées dans de nombreux cas de maladies humaines. Le genre *Streptococcus* – regroupant une bactérie d'intérêt industriel *Streptococcus thermophilus* et de nombreux pathogènes – étant un bon exemple, le chapitre suivant lui sera consacré.

2. *Streptococcus thermophilus* et les streptocoques

2.1 Classification et relations taxonomiques

Depuis le début du 20ème siècle, la classification des streptocoques n'a cessé d'évoluer et les changements dans la nomenclature et la taxonomie ont été nombreux et variés. Le pouvoir hémolytique (1903) puis antigénique (Lancefield, 1933) des souches, ont d'abord permis de différencier les streptocoques β-hémolytiques (tels que *S. pyogenes*, *S. agalactiae*, *S. dysgalactiae*, *S. equi*, *S. canis*, *S. anginosus*) des streptocoques non β-hémolytiques (tels que *S. pneumoniae*, *S. bovis*, *S. suis*, *S. mutans*, *S. salivrius*, *S. mitis*) puis de sous diviser les streptocoques hémolytiques en fonction de la nature de leur antigène polysaccharidique de paroi. En 1937, sur la base de caractères phénotypiques complémentaires, les streptocoques ont ensuite été regroupés en 4 catégories (Sherman, 1937) incluant les groupes « entérocoque » et « lactique », considérés par la suite en tant que genre à part entière et reclassés respectivement sous le terme « *Enterococcus* » et « *Lactococcus* ». Avec l'arrivée du séquençage, l'analyse comparative de l'ARN ribosomal (ARNr) 16S a ensuite permis d'organiser les relations au sein du genre *Streptococcus* en définissant 6 taxons de taille variable. Ainsi, ont été définis les groupes anginosus, bovis, mitis, mutans, pyogénique et salivarius (Figure 2) (Kawamura *et al.*, 1995a). Depuis, bien que de nombreuses souches ou espèces (en particulier non β-hémolytiques) aient enrichi et été associées aux différents groupes existants (Facklam, 2002), aucune étude n'a remis en cause cette classification.

2.2 Les streptocoques pathogènes stricts et opportunistes

Groupe hétérogène de bactéries, les streptocoques constituent une famille ubiquitaire regroupant de nombreuses espèces, isolées des plantes ou de l'environnement, capables de coloniser la peau et les muqueuses de l'homme et de certains animaux. Malgré la présence de pathogènes stricts au sein de ce groupe, de nombreux streptocoques sont importants sur le plan écologique et participent au bon équilibre de la flore microbienne des Hommes et des animaux. Ce n'est qu'au travers d'un déséquilibre immunitaire ou de la flore indigène de l'hôte que ces streptocoques deviendront des pathogènes opportunistes déployant ainsi des mécanismes spécifiques pour coloniser leur environnement.

Parmi les principales espèces pathogènes de l'homme, on retrouve :
- *S. pyogenes*, à l'origine d'infections « aigues » provoquant pharyngites, impétigos, scarlatine mais également « systémiques » contribuant au choc toxique streptococcique associé à de forts taux de morbidité et mortalité (Cunningham, 2000, 2008),
- *S. pneumoniae*, responsable de pneumonies et de méningites (Mitchell and Mitchell, 2010),
- *S. agalactiae*, agent causal de septicémies néonatales (Mullaney, 2001).

De manière générale, les déterminants impliqués dans le pouvoir pathogène d'une souche confèrent aux bactéries la capacité de se fixer à l'hôte, de le coloniser et d'échapper aux mécanismes de défenses de ce dernier.
La capacité d'adhérence aux tissus de l'hôte constitue l'étape clé à l'infection du pathogène. De nombreuses protéines de surfaces ou sécrétées, nécessaires à l'accomplissement de cette étape, sont par conséquent produites par la bactérie. Parmi celles-ci, les adhésines jouent un rôle primordial et leur spécificité confère aux streptocoques la capacité de coloniser l'hôte à divers endroits. Aussi, les glycoprotéines permettront l'adhésion des streptocoques oraux (principalement) aux cellules épithéliales, plaquettes et glycoprotéines salivaires par interaction directe avec l'acide sialique présent ou associé à ces dernières ; et les protéines de liaison à la fibronectine produites chez tous les streptocoques pathogènes favoriseront l'association du pathogène à la matrice extracellulaire de l'hôte (Nobbs *et al.*, 2009). Par ailleurs, chez *S. pyogenes*, le rôle de la protéine de surface M dans l'adhésion de la bactérie aux cellules épithéliales a été largement caractérisé et résulterait d'une interaction de la protéine avec la fibronectine présente au niveau des cellules épithéliales. Néanmoins, d'autres fonctions telles que la protection contre la phagocytose et un rôle dans le maintien de l'homéostasie de la bactérie ont été attribuées à ce facteur de virulence décrit comme multifactoriel (Oehmcke *et al.*, 2010). Plus récemment, les pili ont également été identifiés comme une nouvelle classe d'adhésines bactériennes chez les bactéries pathogènes à Gram positif (Kreikemeyer *et al.*, 2011). Jusqu'à maintenant, 3 pili ont été mis en évidence chez *S. agalactiae* (Lauer *et al.*, 2005; Rosini *et al.*, 2006), 2 chez *S. pneumoniae* (Bagnoli *et al.*, 2008; Barocchi *et al.*, 2006) et 9 chez *S. pyogenes* (Falugi *et al.*, 2008; Mora *et al.*, 2005) en ce qui concerne les trois pathogènes majeurs de l'Homme.
La capacité des pathogènes à former des biofilms joue également un rôle non négligeable dans leur pathogénicité en leur permettant de se protéger des agressions extérieures. Les streptocoques de la cavité buccale sont principalement associés à ce phénomène, leur permettant de mieux résister aux attaques physico-chimiques et frictions mécaniques provoquées par le brossage des dents (Nicolas and Lavoie, 2011). Ce phénomène, largement étudié chez *S. mutans*, responsable des caries dentaires (Keyes, 1960), est

clairement favorisé par le régime alimentaire de l'Homme. En effet, le saccharose ingéré par l'hôte favorise la formation de glucanes insolubles par la bactérie. L'acide généré *via* la voie de biosynthèse de ces polysaccharides déminéralise l'émail des dents et rend la bactérie plus compétitive au sein de l'écosystème buccal lui permettant ainsi de coloniser la plaque dentaire à travers la formation d'un biofilm. La différenciation phénotypique des bactéries présentes dans le biofilm engendre, par ailleurs, des changements d'expression et de régulation de plusieurs gènes contribuant à la formation d'une matrice extracellulaire solide liant les cellules à la surface dentaire (Nicolas and Lavoie, 2011). La communication cellulaire qui s'établit entre les espèces au sein d'un biofilm participe, par ailleurs, à son bon développement et favorise l'échange d'informations génétiques au sein d'une ou entre plusieurs populations.

La présence d'une capsule contribue également à la colonisation des streptocoques pathogènes en empêchant l'interaction des immunoglobulines (présentes à la surface des cellules bactériennes) avec les récepteurs des cellules phagocytaires, en protégeant les bactéries de l'action des antibiotiques et en limitant leur autolyse (Mitchell and Mitchell, 2010). La capsule polysaccharidique présente à la surface de *S. pneumoniae* est considérée comme le facteur de virulence le plus important chez cette bactérie et actuellement 91 sérotypes sont décrits dans la littérature (Park *et al.*, 2007). Par ailleurs, deux vaccins présentant des anticorps dirigés contre les polysaccharides capsulaires sont actuellement disponibles pour prévenir des infections à ce streptocoque (Pneumovax ® et Prevnar ®) (Kreikemeyer *et al.*, 2011).

Enfin, des déterminants plus spécifiques tels que la production de toxines peuvent également contribuer à la virulence d'une espèce (Mitchell and Mitchell, 2010; Nobbs *et al.*, 2009; Rajagopal, 2009). Ces toxines pouvant également être libérées suite à l'autolyse de la bactérie, ceci souligne l'importance du choix des traitements utilisés lors des infections. A titre d'exemple, la pneumolysine, toxine produite par *S. pneumoniae*, contribue à la formation de pores au niveau des cellules bronchiques et alvéolaires de l'hôte favorisant ainsi la dissémination de la bactérie dans le sang (Mitchell and Mitchell, 2010). D'autres protéines dites de « dégradation » sont également caractérisées. Parmi celles-ci l'énolase (enzyme glycolytique présente dans un bon nombre de streptocoques) contribue à la dégradation de la matrice extracellulaire de l'hôte facilitant ainsi l'invasion bactérienne et leur diffusion à travers la barrière épithéliale (Nobbs *et al.*, 2009).

L'utilisation des antibiotiques pour le traitement des infections liées aux streptocoques est couramment préconisée. Néanmoins, l'apparition de souches résistantes oriente actuellement les recherches sur ces pathogènes vers le développement de nouvelles techniques thérapeutiques. Aussi, de par leurs forts niveaux d'expression au cours des

infections à streptocoques et la présence de caractéristiques antigéniques intéressantes, les pili constitueraient des cibles prometteuses pour la production de vaccins (Kreikemeyer *et al.*, 2011; Mora *et al.*, 2005). Cependant, la grande variabilité des antigènes présents sur les pili des streptocoques présage qu'une protection optimale vis-à-vis d'un sérotype donné sera difficile (Falugi *et al.*, 2008).

Parmi les espèces commensales et pathogènes du genre *Streptococcus*, *S. thermophilus* est le seul streptocoque reconnu comme GRAS (**G**enerally **R**ecognized **A**s **S**afe) et dont les capacités homofermentaires sont utilisées dans l'industrie agro-alimentaire. En tant que bactérie modèle pour mes travaux de thèse, la suite de ce chapitre lui sera plus particulièrement consacrée. Son rôle en tant que ferment industriel sera tout d'abord abordé avant de se focaliser sur les caractéristiques évolutives et génétiques de l'espèce.

2.3 *Streptococcus thermophilus*

2.3.1 *S. thermophilus*, un streptocoque d'intérêt industriel

Isolé du lait, *S. thermophilus* se présente sous la forme de coques ovoïdes allant de 0,7 à 0,9 µm. Cette espèce thermotolérante présente une température optimale de croissance de 42-43°C et est capable de se multiplier sur une échelle de température allant de 15 à 52°C. Par ailleurs, *S. thermophilus* survit à un chauffage de 30 min à 65°C, se développe jusqu'à une concentration de 4% en NaCl mais ne se multiplie pas en milieu basique (pH > 9,6) (Moschetti *et al.*, 1998; Sherman, 1937; Stiles and Holzapfel, 1997).

Ces caractéristiques générales confèrent à *S. thermophilus* de bonnes propriétés en tant que ferment dans la fabrication de produits laitiers tels que le yaourt ou certains fromages où elle se retrouve associée à *Lactobacillus bulgaricus* et *Lactococcus lactis*. Au cours de la fermentation, *S. thermophilus* acidifie le lait par production d'acide lactique et contribue à la formation d'un gel appelé caillé. *S. thermophilus* joue également un rôle dans la production d'arômes et le développement de saveurs caractéristiques en favorisant la synthèse de composés aromatiques tels que l'acétaldéhyde, des cétones, du diacétyle et de l'acétoine. La bactérie agit également sur la consistance et la texture du produit fini *via* la production de métabolites tels que les exopolysaccharides (De Vuyst and Degeest, 1999) mais aussi sur leur conservation *via* la production de bactériocines, inhibant la croissance d'espèces phylogénétiquement proches et indésirables telles que les listeria ou les entérobactéries (Villani *et al.*, 1995).

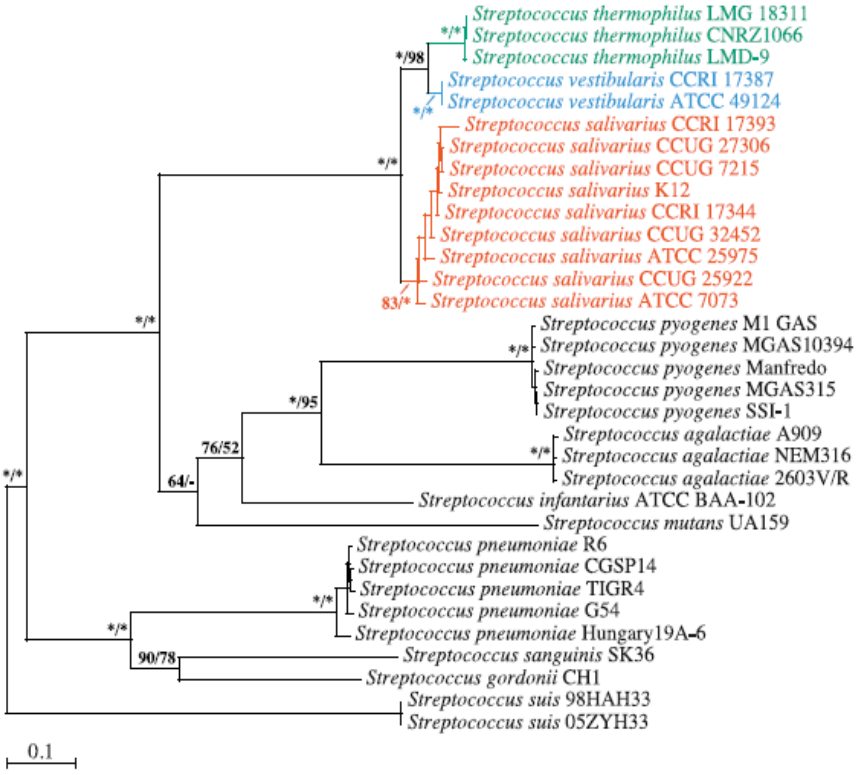


Figure 3 – Ramifications au sein du groupe salivarius établies par inférence phylogénétique sur la séquence concaténée de l'ARNr 16S et des gènes *recA*, *secA* et *secY* (Pombert *et al.*, 2009)**.**

En plus de contribuer à une bonne fermentation des produits laitiers, *S. thermophilus* confère un intérêt santé à l'hôte consommateur en stimulant ses défenses immunitaires et améliorant la digestion du lactose (Guarner *et al.*, 2005) mais aussi en participant à l'équilibre de la flore intestinale (Menard *et al.*, 2004). Enfin, *S. thermophilus* étant retrouvée vivante après passage dans le tube digestif, son rôle éventuel en tant que probiotique a également été soulevé (Elli *et al.*, 2006).

2.3.2 Taxonomie et identification de *S. thermophilus*

S. thermophilus appartient au groupe des streptocoques oraux regroupant les groupes anginosus, mitis, mutans et salivarius (Farrow and Collins, 1984). Bien que la cavité orale soit l'environnement naturel (principal) de ces streptocoques, *S. thermophilus* ne fait que transiter au sein de celle-ci. De par son appartenance au groupe salivarius, *S. thermophilus* se retrouve associée phylogénétiquement à *S. vestibularis* et *S. salivarius* dont l'implication dans des cas d'endocardites et de méningites a été démontrée (Conte *et al.*, 2006; Doyuk *et al.*, 2002; Idigoras *et al.*, 2001; Partridge, 2000). Plusieurs analyses phylogénétiques appuient les relations liant ces trois espèces et décrivent une divergence de l'espèce *salivarius* à la base du clade (Figure 3) (Kawamura *et al.*, 1995b; Pombert *et al.*, 2009; Poyart *et al.*, 1998).

Aussi, malgré la (ré-)habilitation officielle de *S. thermophilus* au rang d'espèce en 1994 (Bergey's manual of determinative bacteriology, Williams et Wilkins, 1994) et son statut reconnu comme GRAS, la proximité de cette espèce avec des pathogènes couplée à son importance technologique a nécessité le développement de techniques pour sa détection et son identification. Les techniques d'hybridation, utilisant des sondes oligonucléotidiques spécifiques à une région de l'ARNr 16S, 23S mais aussi de la partie 5' du gène *lacZ* codant la β-galactosidase, ont tout d'abord été développées (Ehrmann *et al.*, 1994; Lick *et al.*, 1996). Les techniques d'amplification de l'ADN telle que la PCR (**P**olymerase **C**hain **R**eaction) ont ensuite permis d'identifier rapidement les souches de *S. thermophilus* par amplification d'un fragment localisé dans le gène *sodA* codant une superoxide dismutase (Poyart *et al.*, 1998). L'utilisation des méthodes de typage PFGE (**P**ulsed-**Fi**eld **G**el **E**lectrophoresis) et RAPD-PCR (**R**andom **A**mplified **P**olymorphic **D**NA-PCR) a ensuite permis d'explorer la diversité microbienne au sein de l'espèce et de distinguer (de manière quasi-exhaustive) les souches de *S. thermophilus* les unes des autres (Giraffa, 2001; Mora, 2002; Moschetti *et al.*, 1998; O'Sullivan and Fitzgerald, 1998).

La diversité génétique au sein du groupe salivarius a également été étudiée par MLST (**M**ulti**l**ocus **S**equence **T**yping) sur la base de 7 gènes constitutifs (*ilvC*, *pepO*, *pyrE*, *glcK*,

Tableau 2 – Caractéristiques générales de l'ensemble des génomes séquencés appartenant à l'espèce *Streptococcus thermophilus*.

	CNRZ1066	LMG18311	LMD-9	ND03	JIM8232
Taille	1,796 Mb	1,797 Mb	1,856 Mb	1,832 Mb	1,930 Mb
G+C%	39	39	35	39	39
CDS	1915	1890	1709	1919	2145
Plasmide	Absence	Absence	pSTER1 (4449 pb) pSTER2 (3361 pb)	Absence	Absence
Provenance	Yaourt français	Yaourt anglais	Yaourt ou fromage	Lait de Yak, Chine	Lait
Référence	Bolotin *et al.*, 2004	Bolotin *et al.*, 2004	Makarova *et al.*, 2006	Sun *et al.*, 2011	Delorme *et al.*, 2011

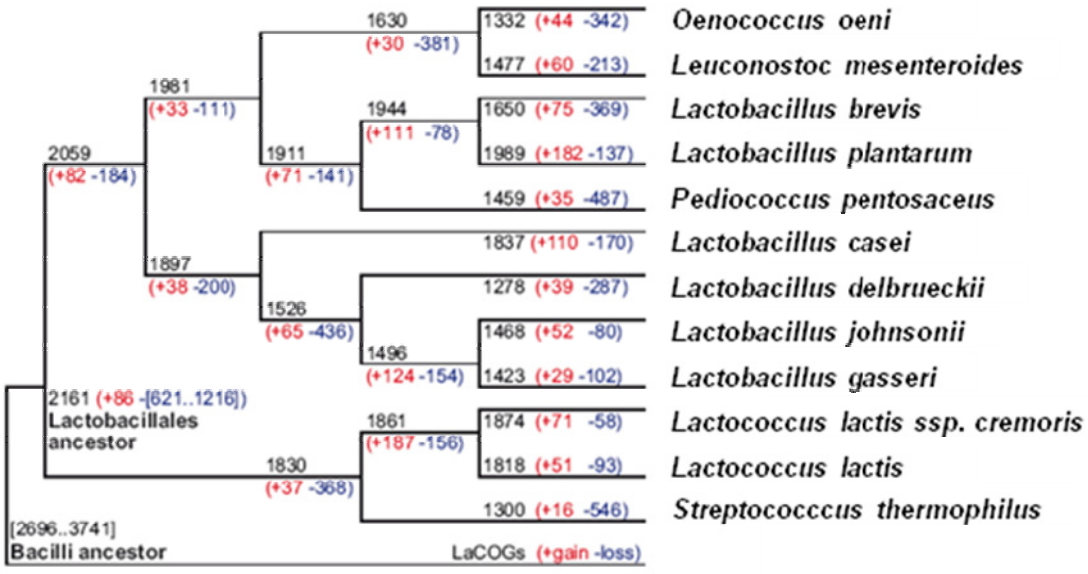


Figure 4 – Dendrogramme représentant les évènements évolutifs au sein de l'ordre des *Lactobacillales* (Makarova *et al.*, 2006)**.**
Au niveau de chaque évènement évolutif le nombre de gènes orthologues « LaCOGs » présents (noir), gagnés (rouge) ou perdus (bleu) pour la branche menant au nœud est indiqué.

ddlA, *thrS* et *dnaK*). L'analyse des profils alléliques des 3 espèces présentes dans ce taxon, a révélé un faible polymorphisme (0,19%) au sein de l'espèce *thermophilus* en comparaison avec ceux des espèces *salivarius* et *vestibularis*, respectivement de 6,6 et de 3,6% (Delorme *et al.*, 2007; Delorme, 2008). Cette faible variabilité génétique de *S. thermophilus* a permis aux auteurs de suggérer une récente émergence de l'espèce de nature clonale. Par ailleurs, le polymorphisme génétique intra-spécifique a fortement été étudié pour l'intérêt industriel qu'il pouvait susciter dans la sélection et le choix des ferments lactiques.

2.3.3 Histoire évolutive de *S. thermophilus*

Cinq génomes sont actuellement séquencés au sein de l'espèce *S. thermophilus.* Les caractéristiques générales de ces génomes sont présentées dans le Tableau 2.

La reconstruction d'un arbre phylogénétique utilisant le principe de parcimonie, et par conséquent conçu de manière à impliquer le minimum d'évènements évolutifs, a permis de restituer l'évolution des *Lactobacillales*, et par conséquent celle de *S. thermophilus*, à partir de l'ancêtre commun des *Bacilli*[1] (Figure 4) (Makarova *et al.*, 2006). Aussi, l'émergence des *Lactobacillales* résulterait d'une perte génétique de 25 à 30% du génome ancestral – associée principalement à la fonction de sporulation et au stress oxydant – et relaterait d'une adaptation anticipée de ces bactéries aux environnements riches en nutriments. D'un point de vue évolutif, la perte de gènes joue également un rôle majeur et prépondérant chez les *Lactobacillales* et en particulier pour l'émergence de l'espèce *S. thermophilus* où elle s'élève à 546 gènes. De manière originale, l'émergence de l'espèce *L. lactis* résulterait majoritairement d'un gain de gènes et serait le seul cas d'évolution positive au sein de cet ordre bactérien.

Bien que *S. thermophilus* et les streptocoques pathogènes partagent une partie non négligeable de leur physiologie et métabolisme global (Figure 5A), l'analyse comparative des génomes des souches séquencées de *S. thermophilus* CNRZ1066, LMG18311 et LMD-9 relate environ 200 gènes non fonctionnels (dû au frameshift, mutations non-sens, délétion et troncation) au sein de cette espèce décrivant ainsi la plus forte proportion de pseudogènes (10%) au sein du genre *Streptococcus* (Bolotin *et al.*, 2004; Makarova *et al.*, 2006). L'innocuité de *S. thermophilus* en serait la principale résultante de par la perte ou l'inactivation de la majorité des déterminants reconnus importants dans le pouvoir pathogène de certains streptocoques (Bolotin *et al.*, 2004). A titre d'exemple, la majorité des 28

[1] La classe des *Bacilli* regroupe l'ordre des *Lactobacillales* et celui des *Bacillales, comprenant les genres Listeria, Bacillus, Staphylococcus*.

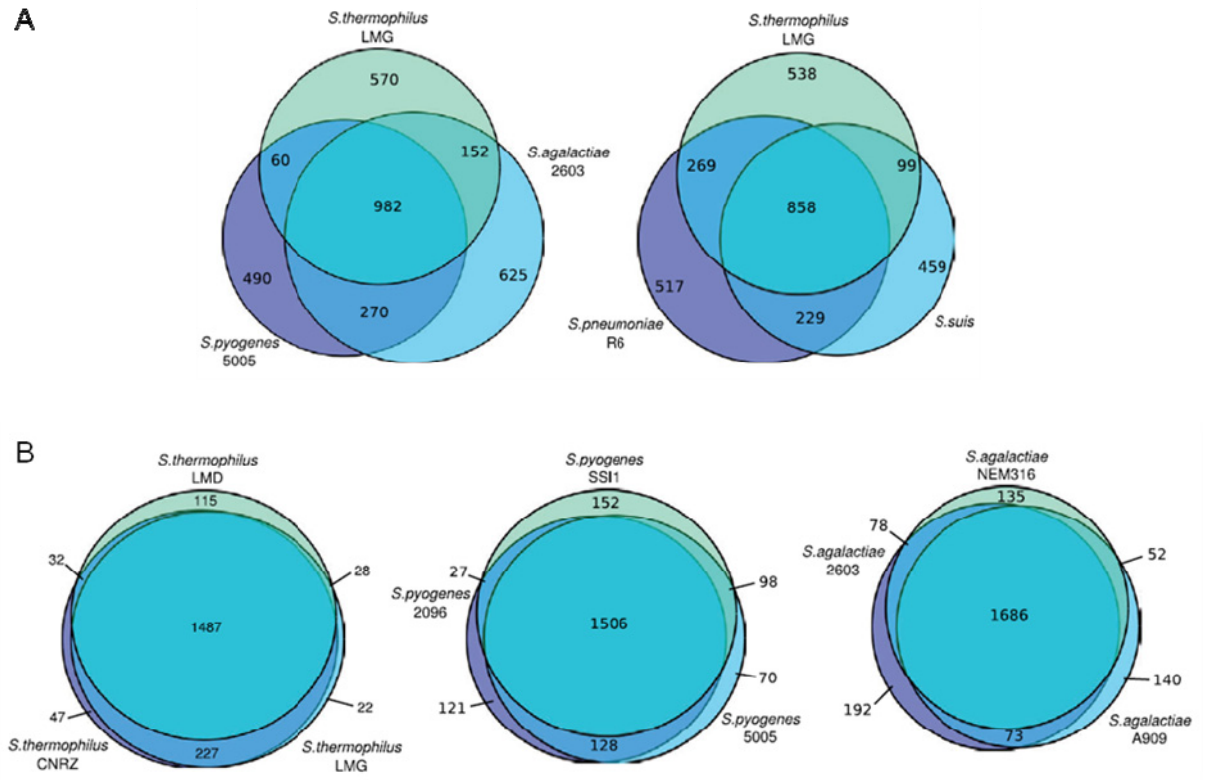


Figure 5 – Diagrammes de Venn représentant le pourcentage de gènes partagés au niveau de différentes espèces (A) ou au sein d'une même espèce (B) de streptocoques (Lefébure and Stanhope, 2007)**.**

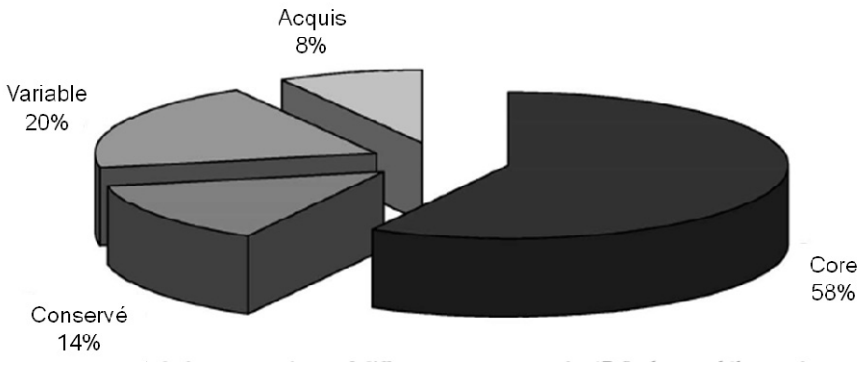


Figure 6 – Composition du pan-génome de *Streptococcus thermophilus* (Rasmussen *et al.*, 2008)**.**

Le pan-génome de l'espèce *thermophilus* présente 58% de gène commun à tous les individus de l'espèce auquel 14% de gènes conservés sont associés malgré leur absence (ou leur perte) dans un nombre restreint de souches. Huit pourcents constituent un pool de gènes récemment acquis présents dans un petit nombre de souches. Enfin 20% du pan-génome correspond aux gènes dits variables présents et représentés dans la majorité des souches.

déterminants de surface importants dans la virulence de *S. pneumoniae* (sortases, adhésines, protéases, etc.) sont absents ou inactivés chez *S. thermophilus* (Bolotin *et al.*, 2004). Néanmoins, certaines souches de *S. thermophilus* auraient conservé certains gènes similaires à des gènes de virulence (protéase de paroi PrtS, gènes impliqués dans la biosynthèse des exopolysaccharides et opéron uréase) (Hols *et al.*, 2005). Cependant, la fonction de ces derniers aurait évolué avec le temps et participerait à l'adaptation de *S. thermophilus* à l'environnement laitier. Par ailleurs, le phénomène de perte de gènes touche également certains transporteurs, des protéines impliquées dans le métabolisme énergétique et les éléments génétiques mobiles (phages, transposons, séquences d'insertion).

Malgré une dégénérescence évidente du génome de *S. thermophilus* et un faible polymorphisme au sein de l'espèce, de nombreux îlots génomiques seraient, quant à eux, acquis par transfert horizontal (Hols *et al.*, 2005). L'ensemble des gènes présents chez tous les streptocoques ou « core-génome » a été évalué à 611 et représente environ 25% du génome d'un streptocoque (Figure 5A) alors qu'au niveau de l'espèce il est en moyenne de 70-75% (Figure 5B) (Lefébure and Stanhope, 2007). Sur la base de 47 isolats, Rasmussen et ses collaborateurs ont défini un core-génome équivalent à 1271 séquences codantes pour l'espèce *thermophilus*. L'analyse approfondie du génome « dispensable » (gènes non représentés dans l'ensemble de la population), représentant 42% du pan-génome de l'espèce (Figure 6), a ensuite permis d'identifier 431 gènes dits « variables » pouvant résulter d'une acquisition par transfert horizontal au vu de leur représentation aléatoire au sein des génomes de la population (Rasmussen *et al.*, 2008). Les analyses *in silico* basées sur des biais de composition du génome permettent également de générer des pools de gènes potentiellement acquis par transfert horizontal et appelés « atypiques ». Une première étude, basée sur les génomes séquencés de *S. thermophilus* a ainsi révélé 197 gènes potentiellement acquis par transfert horizontal parmi lesquels 60% sont associés à des gènes impliqués dans des mécanismes de transfert codant des transposases, bactériophages et l'ARN de transfert (ARNt) (Liu *et al.*, 2009). Dans une seconde étude, le pool de gènes « atypiques » fut défini sur la base des 47 isolats utilisés lors de l'étude de Rasmussen *et al.*, et comparé aux 431 gènes « variables » identifiés dans cette même étude. Un total de 54 gènes dits « atypiques et variables », localisés dans de nombreux îlots génomiques, a alors été décrit comme acquis par transfert horizontal (Eng *et al.*, 2011). Bien que la majorité des protéines codées par ces gènes soient associées à des fonctions hypothétiques (41%) ou tronquées (13%), les produits des gènes annotés interviennent quant à eux dans des voies de biosynthèse d'exopolysaccharides (9) ou de bactériocines (1), dans des systèmes de transport (5) ou de restriction/modification (2), dans le

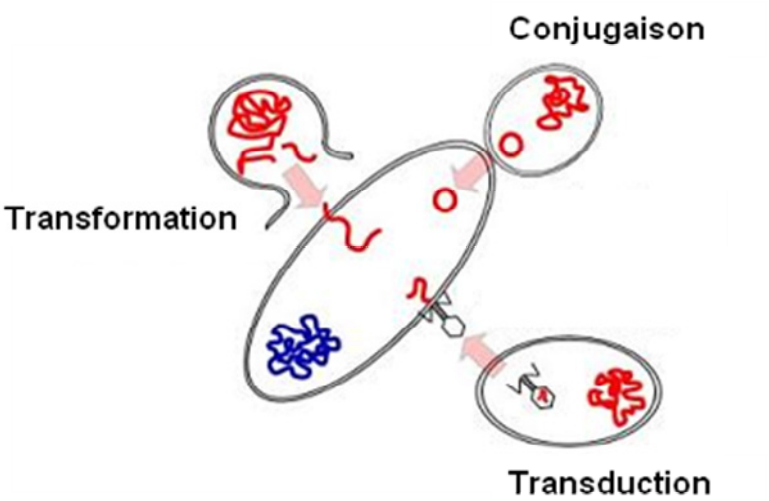


Figure 7 – Les mécanismes de transfert horizontal chez les bactéries.

métabolisme des sucres (3) ou codent également des régulateurs transcriptionnels appartenant à la famille des Rgg (**R**egulator **G**ene of **G**lucosyltransferase). Enfin, trois gènes coderaient pour des protéines d'origine phagique.

La perte de gènes ou leur dégénérescence jouent un rôle majeur dans l'adaptation de *S. thermophilus* à son environnement mais l'évolution de son génome résulte également d'une acquisition de gènes par transfert horizontal. Les trois mécanismes de transfert connus pour être les phénomènes clés de l'évolution bactérienne – conjugaison, transduction et transformation – sont par ailleurs effectifs chez ce streptocoque et contribuent sûrement de manière coordonnée aux flux de gènes de cette espèce. Ils seront donc abordés dans la partie suivante.

2.3.4 Le transfert de l'information génétique chez *S. thermophilus*

Par opposition au transfert vertical, le transfert horizontal de gènes permet aux bactéries d'échanger du matériel génétique provenant d'un autre organisme sans en être le descendant. Trois mécanismes de transfert horizontal sont distingués chez les bactéries : la transduction, la conjugaison et la transformation (Figure 7). A l'heure actuelle, 1,6 à 32,6% des génomes microbiens auraient été acquis par transfert horizontal (Koonin *et al.*, 2001).

- La transduction :

Le phénomène de transduction permet le transfert d'un fragment d'ADN d'une bactérie donneuse à une bactérie receveuse par l'intermédiaire d'un phage. Deux voies de transduction sont décrites : la transduction généralisée et la transduction spécialisée. La transduction généralisée est liée à une erreur de spécificité de l'encapsidation résultant en l'intégration d'ADN bactérien au sein de la particule phagique. Aussi, au cours du cycle d'infection suivant, cette séquence d'ADN sera transférée au sein d'une autre bactérie et pourra être recombinée au niveau de son génome. Dans le cas de la transduction spécialisée, l'excision impropre du prophage, intégré dans le chromosome bactérien, résulte en l'export de l'ADN bactérien contigu au site d'intégration. Lors du cycle d'infection suivant, l'intégration du génome du phage dans le chromosome bactérien aboutira à l'apport de gène(s) modifiant ainsi le patrimoine génétique de la cellule (Sørensen *et al.*, 2005). Chez *S. thermophilus*, bien que le rôle des phages dans le transfert de gènes au sein de l'espèce ait été décrit en 1988 (Mercenier *et al.*, 1988), ce n'est que récemment que l'infection de phages entre deux ferments lactiques a été mise en évidence. Aussi, l'évolution et

l'acquisition de modules génétiques entre *S. thermophilus* et *L. lactis* seraient en partie médiées par les phages (Ammann *et al.*, 2008).

- La conjugaison :

Mis en évidence chez *Escherichia coli* par Lederberg et Tatum en 1946, le transfert par conjugaison de l'ADN nécessite un contact physique entre une cellule « donneuse » et une cellule « receveuse » et se traduit généralement par l'acquisition d'un plasmide conjugatif possédant l'information génétique lui permettant de se transférer entre bactéries et de se répliquer de manière autonome (Thomas and Nielsen, 2005). L'acquisition de ces plasmides apporte aux bactéries des caractères adaptatifs avantageux tels que la résistance aux antibiotiques ou la capacité à utiliser de nouvelles sources de carbones (Sørensen *et al.*, 2005). Le transfert entre les organismes donneur et accepteur se fait en plusieurs étapes et nécessite en premier lieu une étape de reconnaissance qui s'établit soit par la présence de pili chez les bactéries à Gram négatif ou soit à travers la production de protéines d'agrégation chez les bactéries à Gram positif (Sørensen *et al.*, 2005). Chez les bactéries lactiques, cette méthode de transfert d'ADN a été largement observée (Teuber *et al.*, 1999; Toomey *et al.*, 2009). Chez les entérocoques, le transfert par conjugaison répond à un mécanisme de communication cellulaire. Aussi la production de molécules de signalisation par les cellules receveuses va induire une réponse chez les cellules donneuses, déclenchant la synthèse de protéines membranaires favorisant l'agrégation des cellules donneuses et receveuses entre elles (Dunny *et al.*, 2001; Thomas and Nielsen, 2005) (Chapitre 2, § 2.2.4). Des éléments chromosomiques pouvant être excisés et transférés à une autre cellule ont également été identifiés et constituent un groupe d'éléments mobiles appelé ICE (**I**ntegrative and **C**onjugative **E**lement). Ces éléments intégratifs et conjugatifs font appel à un système de recombinaison site-spécifique codant les éléments nécessaires à leur excision, à leur transfert et à leur intégration au niveau de deux sites, l'un chromosomique et l'autre présent dans l'élément (Wozniak and Waldor, 2010). Ce processus de conjugaison a par ailleurs été décrit chez *S. thermophilus* (Burrus *et al.*, 2002).

- La transformation :

Le phénomène de transformation naturelle a été découvert *chez S. pneumoniae* (Avery *et al.*, 1944; Griffith, 1928), et est aujourd'hui caractérisé dans plusieurs phyla bactériens – tels que les firmicutes, les protéobactéries, les cyanobactéries et les archaea (Johnsborg *et al.*, 2007). Pour qu'une bactérie soit naturellement transformable, elle doit être dans un état physiologique appelé « compétence » qu'elle acquiert via l'expression d'un pool de gènes dit « précoces », en réponse à la détection de signaux environnementaux (Claverys *et al.*,

2009). Cet état physiologique contrôle par ailleurs l'expression de gènes dit « tardifs » codant la machinerie de compétence nécessaire à l'internalisation et à la recombinaison homologue de l'ADN dans le chromosome bactérien (Johnsborg *et al.*, 2007; Johnsborg and Håvarstein, 2009; Tortosa and Dubnau, 1999). Aussi, la transformation génétique est définie comme l'acquisition d'un fragment d'ADN exogène dans une cellule, entrainant éventuellement une modification phénotypique de l'organisme receveur. Ce phénomène est décrit comme transitoire dans la plupart des espèces bactériennes naturellement transformables (Johnsborg *et al.*, 2007).

Bien que les mécanismes d'acquisition de la compétence soient spécifiques à chaque espèce (*e.g.* Chapitre 2 § 2.1.3 ; Chapitre 3 § 3.2.2 et 3.2.3), la machinerie d'absorption de l'ADN forme, quant à elle, un complexe protéique conservé entre les espèces et appelé le transformasome (Claverys *et al.*, 2009). De manière simplifiée, les protéines ComGA-F constituent un pseudo-pilus nécessaire à la liaison de l'ADN double brin (ADNdb) à la surface des cellules compétentes et assurent la formation d'un pore d'entrée permettant la reconnaissance de l'ADNdb par un récepteur membranaire appelé ComEA. Ce récepteur fixe par ailleurs l'ADNdb sans spécificité de séquence. L'ADNdb est ensuite dégradé en ADN simple brin (ADNsb) par une endonucléase (EndA ou NucA). L'ADNsb formé est alors internalisé à travers un canal de translocation appelé ComEC. Enfin, une protéine de liaison de l'ATP, ComFA, apporte l'énergie nécessaire au transport de l'ADNsb (Chen and Dubnau, 2004; Claverys *et al.*, 2009). Une fois internalisé, l'ADNsb est ensuite inclus au sein d'un complexe protéique, composé des protéines SSB (**S**ingle **S**trand **B**inding), RecA et DprA, appelé complexe d'éclipse assurant la protection de ce dernier et sa restauration en ADNdb (Claverys *et al.*, 2009). La recombinaison de l'ADN exogène est ensuite prise en charge par la protéine RecA et, la formation d'un hétéroduplex entre l'ADNsb exogène et chromosomique induit la substitution du brin donneur au brin receveur (Claverys *et al.*, 2009). Cette étape d'intégration de l'ADN monocaténaire est effective en présence d'un degré d'homologie suffisant entre l'ADN exogène et chromosomique et peut soit résulter en l'apport de nouvelles fonctions physiologiques ou soit éventuellement en l'inactivation de certains gènes par insertion ou délétion (Thomas and Nielsen, 2005).

De nombreux streptocoques, comme *S. thermophilus*, ont longtemps été considérés comme « non naturellement compétents ». Des techniques, permettant de perforer les enveloppes bactériennes des bactéries non compétentes ont alors été développées afin de permettre leur manipulation génétique. Chez *S. thermophilus*, l'électroporation – consistant à soumettre les bactéries à des impulsions électriques ayant pour but de créer des pores dans les enveloppes bactériennes – a longtemps été utilisée comme moyen le plus efficace pour transformer les cellules (Marciset and Mollet, 1994; Somkuti and Steinberg, 1988) avant que le phénomène de compétence naturelle ne soit décrit au sein de cette espèce (Blomqvist *et*

al., 2006a; Fontaine *et al.*, 2010a; Gardan *et al.*, 2009) (Chapitre 3 § 3.2.2). Récemment, la mise en évidence de transfert de fragments d'ADN entre différentes souches de *S. thermophilus* par transformation naturelle, a ouvert des perspectives intéressantes concernant l'échange d'informations génétiques au sein de l'espèce et l'acquisition de nouvelles fonctions (*e.g.* résistance aux phages) pertinentes pour les industries agro-alimentaires (Fontaine *et al.*, 2010b).

Bien que ces mécanismes soient fréquents chez les bactéries, l'ADN transféré aux bactéries par transfert horizontal doit procurer un avantage sélectif afin d'être acquis de manière durable (Thomas and Nielsen, 2005). Par ailleurs, les éléments génétiques mobiles, tels que les séquences d'insertions, transposons, intégrons et îlots génomiques, échangés entre les bactéries contribuent également à la plasticité des génomes bactériens. Enfin, le transfert horizontal n'est pas restreint aux bactéries. Il est également décrit au sein d'autres domaines et possible entre archées et bactéries mais aussi entre eucaryotes et bactéries (Boto, 2010).

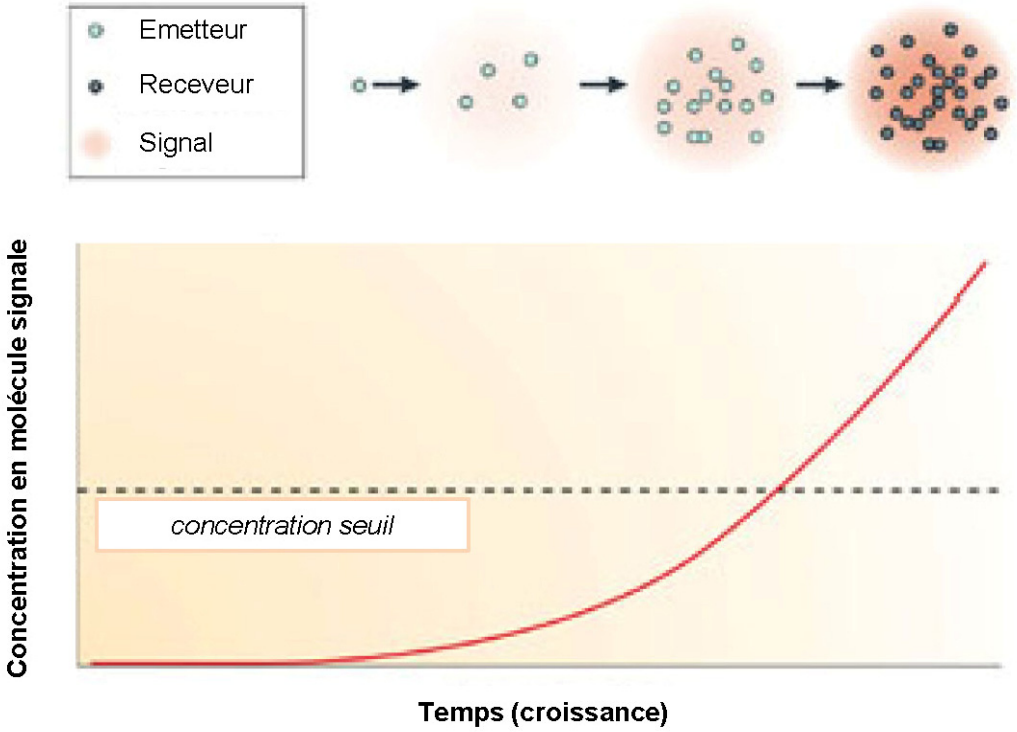


Figure 8 - Représentation schématique du déclenchement d'un système de quorum-sensing (d'après Keller and Surette, 2006).

Chapitre 2 : La communication cellulaire de type quorum-sensing

1. Concepts généraux

Mise en évidence par les travaux de Tomasz sur la compétence *chez S. pneumoniae* (Tomasz, 1965) et Hasting chez les deux bactéries marines luminescentes *Vibrio fischeri* et *Vibrio harveyi* (Nealson and Hasting, 1970 et 1979), la communication cellulaire de type quorum-sensing (QS), procurant aux bactéries une meilleure résistance, survie ou adaptation à leur environnement, se définit comme un mécanisme de régulation de l'expression des gènes en réponse à l'augmentation de la densité cellulaire (Figure 8). Ceci se traduit par la capacité des bactéries à percevoir leurs congénères à travers la production, la sécrétion puis la détection de molécules de signalisation propres à chaque espèce bactérienne. A cette notion de détection appelée « sensing », s'ajoute une notion de seuil ou « quorum » amenant les bactéries à percevoir les molécules de signalisation émises dans le milieu environnant, lorsque ces dernières atteignent une certaine concentration critique. Aussi, la mise en évidence de l'action simultanée de ces deux notions de « détection » et de « seuil » conduit l'ensemble d'une population bactérienne à modifier de manière coordonnée l'expression de gènes lui conférant ainsi de nouvelles propriétés.

Au cours des dernières décennies, il a été admis que les bactéries sont capables de répondre à des signaux qu'elles synthétisent et sécrètent. Ces petites molécules de signalisation ont été appelées autoinducteurs (AI) ou phéromones. Cette découverte de l'existence d'interactions au sein du monde des procaryotes, a profondément bouleversé la notion qu'on en avait. En effet, les bactéries ne fonctionnaient plus de manière autonome mais, avaient la capacité d'initier un comportement collectif en réponse à leur environnement (Atkinson and Williams, 2009).

Bien que le mécanisme général de fonctionnement des systèmes de communication de type QS réponde à un schéma précis, leur spécificité rend compte de la nature de la molécule de signalisation utilisée. En effet, alors que chez les bactéries à Gram négatif, les *N*-acyl homoserine lactones constituent la famille de molécules de signalisation la mieux caractérisée, les bactéries à Gram positif utilisent quant à elles des peptides autoinducteurs issus du clivage d'une molécule « précurseur » subissant, dans certains cas, des

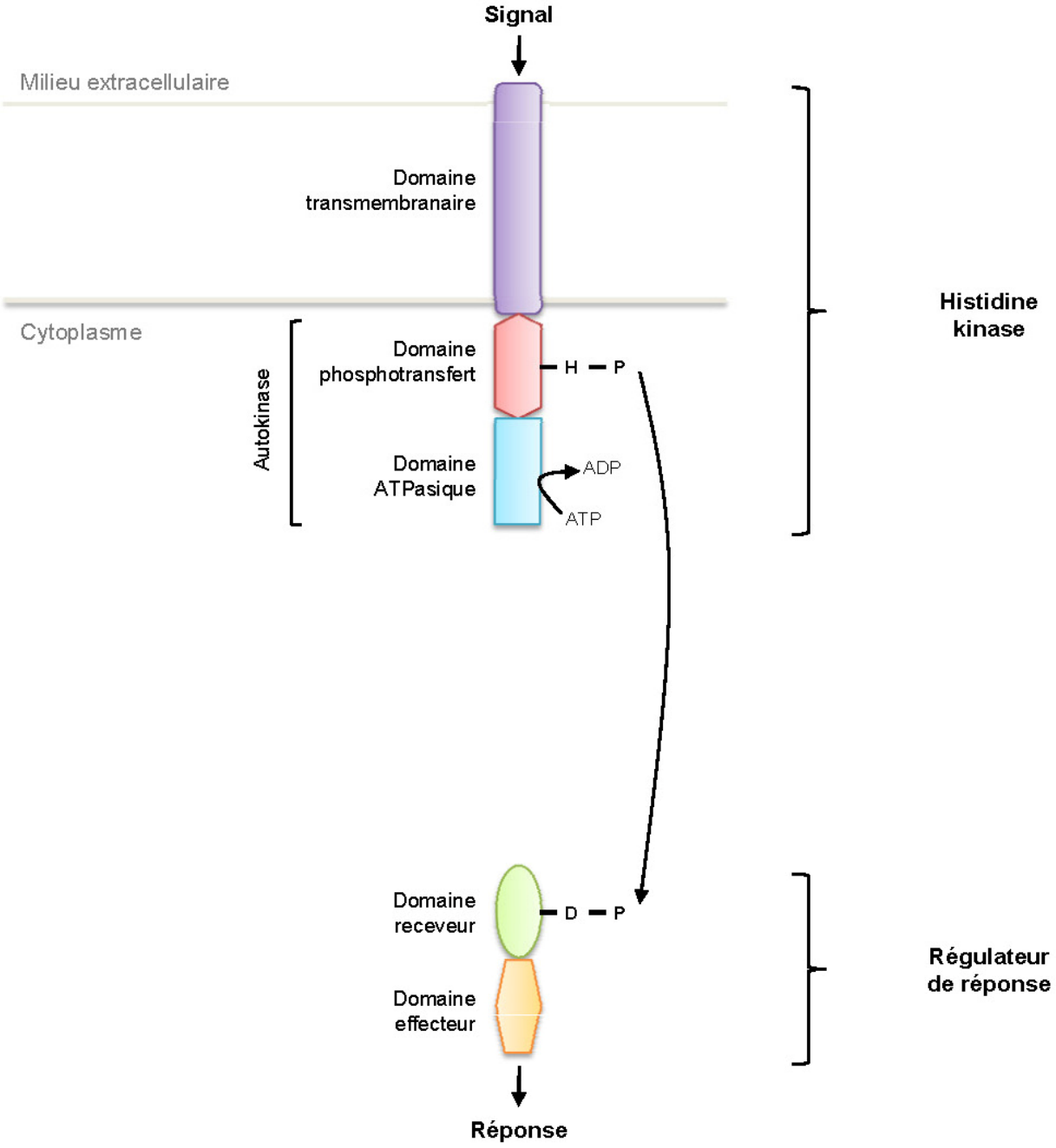


Figure 9 – Représentation schématique d'un système à deux composants.

modifications post-traductionnelles (Dunny and Leonard, 1997). Aussi, en fonction de la nature chimique de la molécule de signalisation mise en œuvre, les bactéries utilisent différentes voies de signalisation pour assurer leur détection.

En lien avec mes travaux de thèse, ce chapitre sera consacré aux systèmes de QS chez les bactéries. Dans une première partie seront présentés les principaux systèmes de QS chez les bactéries à Gram positif. Puis, dans une deuxième et troisième partie seront décrits, de manière plus succincte, le QS chez les bactéries à Gram négatif et au niveau inter-espèces.

2. Le QS chez les bactéries à Gram positif

Chez les bactéries à Gram positif, les molécules de signalisation impliquées dans des phénomènes de QS correspondent à des peptides autoinducteurs (AIP, **A**uto**i**nducing **p**eptide) pouvant présenter ou non des modifications. Deux voies de signalisation de ces peptides sont décrites dans la littérature. Alors que la première assure la reconnaissance puis la transduction du signal *via* l'utilisation d'un système à deux composants ; la deuxième nécessite dans un premier temps la réimportation de la molécule de signalisation dans le cytoplasme de la cellule *via* un transporteur d'oligopeptides puis dans un second temps son interaction avec un effecteur intracellulaire appartenant à la famille des RNPP (**R**ap, **N**prR, **P**lcR, **P**rgX) (Miller and Bassler, 2001; Rocha-Estrada *et al.*, 2010).

Sur ces observations générales, une première partie sera consacrée aux systèmes impliquant une détection extracellulaire des AIP, c'est-à-dire utilisant un système à deux composants et une seconde partie présentera les systèmes de QS en lien avec mes travaux de thèse et nécessitant un système de transport d'oligopeptides pour l'import des phéromones dans le cytoplasme.

2.1 Détection extracellulaire des peptides autoinducteurs

2.1.1 Les systèmes à deux composants

Les systèmes à deux composants (TCS, **T**wo **C**omponent **S**ystems) sont largement utilisés par les bactéries pour répondre à différents signaux et stimuli de l'environnement (Hoch, 2000). Ces systèmes sont composés d'une histidine kinase et d'un régulateur de réponse (Figure 9) (Casino *et al.*, 2010).

La plupart des histidines kinases contiennent un domaine transmembranaire couplé à un domaine autokinase. Aussi, la détection du signal au niveau du domaine transmembranaire

provoque l'activation du domaine autokinase, *via* la phosphorylation d'un résidu histidine. Le groupe phosphoryl est ensuite transféré sur un résidu aspartate localisé dans le domaine receveur du régulateur de réponse apparenté, entraînant *in fine* une modulation du domaine effecteur de ce dernier. Cette modulation, provoquant un changement conformationnel du régulateur de réponse, module l'activité du domaine effecteur, représentant dans 63% des cas un domaine de liaison à l'ADN. Le régulateur transcriptionnel peut alors se fixer sur les régions promotrices des gènes cibles et induire la réponse physiologique apparentée (Casino *et al.*, 2010; Gao and Stock, 2009). Le régulateur de réponse représente généralement le dernier composant de la cascade de phosphorylation et peut être soit phosphorylé (comme décrit ci-dessus), soit déphosphorylé par l'histidine kinase jouant alors le rôle d'une phosphatase (Casino *et al.*, 2010).

De nombreuses fonctions biologiques sont régulées par des TCS chez les bactéries. Parmi les TCS impliqués dans des mécanismes de QS, les systèmes *agrBDCA* de *Staphylococcus aureus* et *comCDE* de *S. pneumoniae* ont été particulièrement étudiés. Ils seront présentés ci-dessous.

2.1.2 La virulence chez *Staphylococcus aureus* : le système *agrBDCA*

S. aureus est une bactérie pathogène opportuniste responsable de nombreuses infections, telle que l'endocardite, liées en partie à l'invasion des tissus de l'hôte. Pour contrôler l'expression de la majorité des déterminants impliqués dans son pouvoir pathogène, *S. aureus* utilise un mécanisme de communication cellulaire connu sous le nom de système *agr* (**a**ccessory **g**ene **r**egulator). Aussi, à forte densité cellulaire, l'expression des protéines de surface (adhésine, protéine A, coagulase, etc.) assurant l'adhésion et la colonisation de la bactérie sera inhibée en faveur de facteurs de virulence (exotoxines, hémolysines, enzymes dégradatives, etc.) nécessaires à la dégradation des tissus et à la dissémination de la bactérie dans l'hôte (Novick, 2003).

Le locus *agr* consiste en deux unités de transcription divergentes, RNAII (3500 nt) et RNAIII (514 nt), placées sous le contrôle de deux promoteurs appelés respectivement P2 et P3 (Figure 10). L'opéron P2, RNAII ou *agrBDCA*, code quatre protéines requises pour la synthèse de l'AIP, sa détection et la transduction du signal généré par ce dernier. Les gènes *agrD* et *agrB* codent respectivement un propeptide autoinducteur et une endopeptidase transmembranaire requise pour la maturation de l'AIP. En effet, AgrD est codé sous la forme d'un précurseur de 45 acides aminés contenant une séquence signale en N-terminal lui assurant un adressage à la membrane. La protéine AgrB intervient ensuite dans la

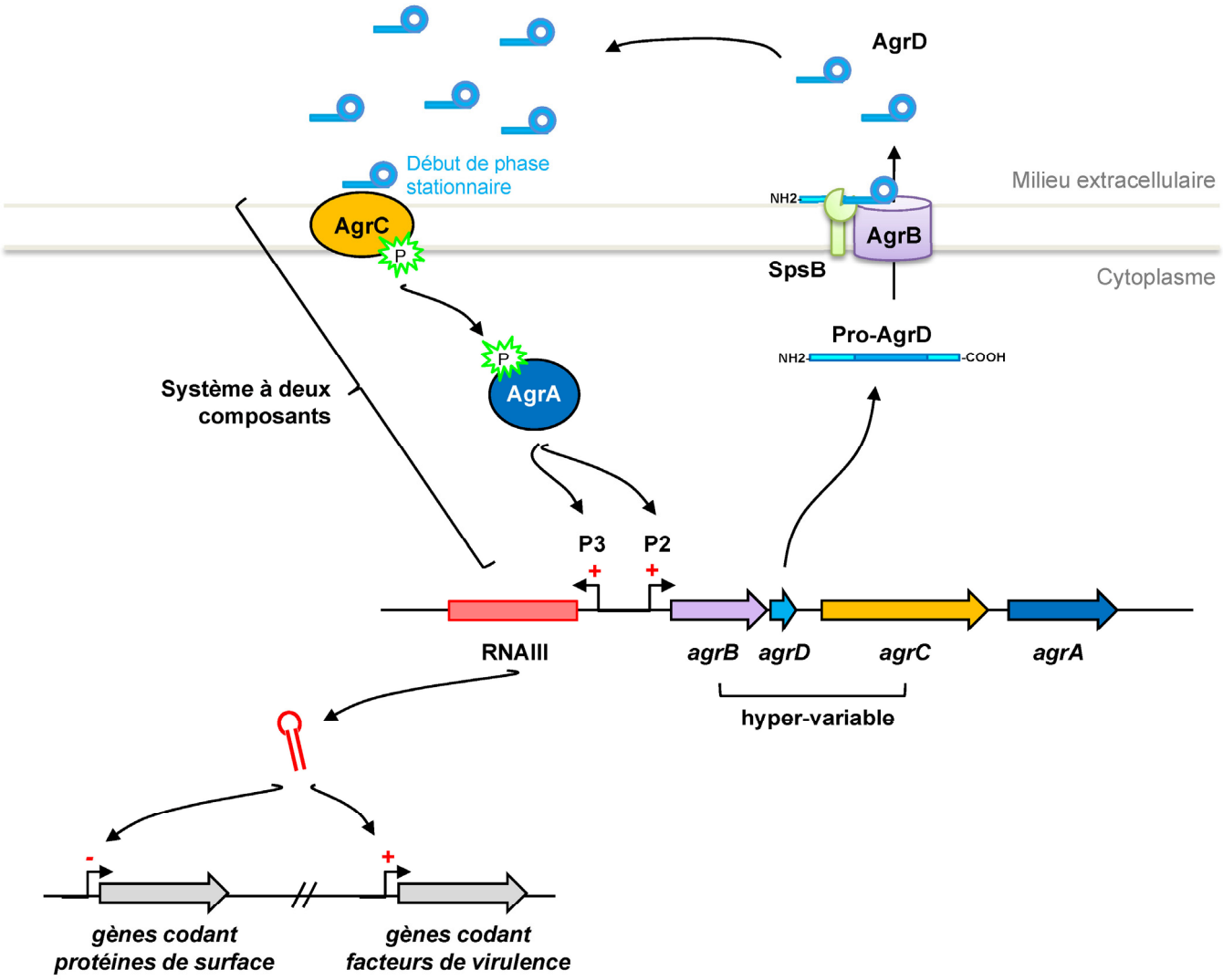
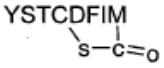
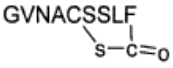
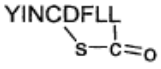
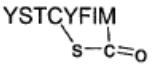


Figure 10 – Représentation schématique du système *agrBDCA* chez *Staphylococcus aureus* (inspiré de Novick and Geisinger, 2008 ; Thoendel and Horswill, 2009)**.**
La phéromone codée par le gène *agrD* est maturée et modifiée, par une endopeptidase membranaire AgrB et une signal peptidase SpsB, au cours de son export. En début de phase stationnaire, AgrD est détecté par son récepteur membranaire AgrC. Cette reconnaissance assure l'autophosphorylation de AgrC et le transfert du phosphate formé au régulateur de réponse AgrA. AgrA~P assure alors l'expression de l'opéron *agrBDCA* et d'un ARN régulateur RNAIII – contrôlant positivement l'expression de facteurs de virulence et négativement l'expression de protéines de surfaces – en se fixant au niveau des régions promotrices P2 et P3.

maturation du peptide en assurant un clivage côté C-terminal et la formation d'une liaison thioesther entre une cystéine conservée et le groupement carboxyle de la partie C-terminale du peptide. Cette structure une fois obtenue représente la signature du peptide et est essentielle à son activité (Novick and Geisinger, 2008). Plus récemment, le rôle d'une signal peptidase, appelée SpsB, a également été démontré dans la maturation du peptide. Elle assure le relargage du peptide mature dans le milieu extracellulaire suite au clivage de sa séquence signale en N-terminal (Figure 10) (Thoendel and Horswill, 2009). Les gènes *agrA* et *agrC* forment quant à eux un système à deux composants où AgrC représente l'histidine kinase et AgrA, le régulateur de réponse. Aussi, à forte densité cellulaire, la liaison du peptide AgrD mature à son récepteur AgrC déclenchera l'autophosphorylation de ce dernier (Mayville 1999). Le phosphate ainsi formé est ensuite transféré au régulateur de réponse AgrA, modulant *in fine* son état d'oligomérisation. Cette étape requise pour l'activation des promoteurs P2 et P3 assurera la fixation de AgrA dans leur région promotrice sur des régions répétées directes (consensus [TA][AC][CA]GTTN[AG][TG] séparées par une région de 12 à 13 pb) (Koenig *et al.*, 2004; Novick and Geisinger, 2008). L'opéron *agrBDCA* assure donc son auto-induction tout en régulant positivement l'activité du promoteur P3 codant l'effecteur de réponse du système *agr*, une molécule d'ARN appelée RNAIII, responsable de la répression des protéines de surfaces et de l'induction des facteurs de virulence (Figure 10) (Novick, 2003). Cet ARN régulateur présente une structure secondaire complexe composée de 14 motifs tige boucle et interfère sur l'efficacité de traduction des protéines cibles (α-hémolysine, protéine A, etc.) ; soit en modifiant la stabilité de leur ARN messager (ARNm), soit en libérant ou masquant les séquences de Shine-Dalgarno et par conséquent le site de fixation des ribosomes. RNAIII agit également sur le régulon Rot (**R**epressor **o**f **t**oxins) en contrôlant négativement la transcription de ce facteur (Thoendel *et al.*, 2011). Néanmoins, bien que RNAIII soit l'effecteur central du système *agr*, certains gènes décrits comme importants dans la pathogénicité de la bactérie sont régulés positivement et de manière directe par AgrA. Ces gènes, connus sous le nom de *psm* (**p**henol-**s**oluble **m**odulins), codent de petits peptides amphipathiques présentant des propriétés pro-inflammatoires, leucocytolytiques ou encore chimiotactiques (Thoendel *et al.*, 2011). Aussi, la mise en place de la boucle rétro-positive de l'opéron *agrBDCA* assure et induit le passage de la population à un état de virulence. Son rôle dans la formation de biofilm a également été décrit (Novick and Geisinger, 2008). Enfin, le niveau de transcription de RNAIII est également contrôlé par d'autres régulateurs (tel que SarA) ou par des TCS en réponse à des signaux environnementaux (tels que ArlRS, SrrAB) (Yarwood and Schlievert, 2003).

L'analyse des différents systèmes *agr* chez les staphylocoques a révélé une variation allélique de la région *agrBDC* (Figure 10). Ce polymorphisme a conduit à l'identification de

Tableau 3 –Peptides autoinducteurs AgrD connus chez *Staphylococcus aureus*.

AgrD-1	AgrD-2	AgrD-3	AgrD-4
			

trois groupes (I/IV, II et III) de systèmes *agr* chez *S. aureus* et assure, pour un groupe donné, une spécificité de séquence du peptide AgrD et sa reconnaissance par son récepteur apparenté. A l'exemple des peptides AgrD produits par *S. aureus* (Tableau 3), les peptides AgrD sont relativement hydrophobes, composés de 7 à 9 résidus et présentent des caractéristiques structurales conservées (liaison thioester). Un phénomène de cross-inhibition a par ailleurs été décrit entre ces peptides au niveau intra ou inter-espèces. Aussi, chaque peptide bien qu'activateur des systèmes *agr* du groupe auquel il appartient, inhibe de manière compétitive les autres systèmes *agr*. Ce phénomène a été décrit entre les différents peptides AgrD de *S. aureus* (Wright *et al.*, 2005) et également entre différentes espèces de staphylocoques tels que *S. epidermidis* et *S. aureus* (Otto, 2009). Ce phénomène assure par conséquent un avantage compétitif lors de l'infection et la colonisation de l'hôte aux bactéries possédant le même allèle et ayant atteint les premières leur densité critique de population nécessaire au déclenchement de leur système *agr*. Très récemment, un dipeptide cyclique produit par *Lactobacillus reuteri* a été décrit comme inhibiteur des systèmes *agr* chez *S. aureus* (Li *et al.*, 2011).

Enfin, même s'ils ne seront pas développés, des systèmes orthologues au système *agr* du genre *Staphylococcus* ont été décrits chez *E. faecalis* (système *fsr*), *Lactobacillus plantarum* (système *lam*) et *Listeria monocytogenes* (système *agr*). Bien que ces systèmes interviennent tous dans la formation de biofilm, le système *fsr* et *lam* ont également été décrits comme impliqués, respectivement, dans la virulence de la bactérie *via* la production de facteurs de virulence tels que la gélatinase GelE et l'adhérence de la bactérie aux surfaces (Thoendel and Horswill, 2010). Par ailleurs, des analyses de Psi-Blast ont permit d'identifier des systèmes homologues au sein du genre *Clostridium* et chez quelques espèces appartenant au genre *Bacillus* (Wuster and Babu, 2008).

2.1.3 La compétence chez *Streptococcus pneumoniae* : le système *comCDE*

La régulation de la compétence pour la transformation génétique, chez les streptocoques du groupe mitis et anginosus, implique un système de communication cellulaire, résumé dans la Figure 11. Ce système, étudié en détails chez *S. pneumoniae*, requiert la production d'un peptide phéromone appelé CSP (**C**ompetence-**S**timulating **P**eptide). CSP est codé sous la forme d'un précurseur de 41 acides aminés par le gène *comC* et possède dans sa partie N-terminale une séquence signale lui conférant un adressage à la membrane et contenant un motif double glycine (GG) (Håvarstein *et al.*, 1995). La maturation et la sécrétion de CSP sont prises en charge par un ABC-transporteur, appelé ComAB, décrit comme essentiel pour

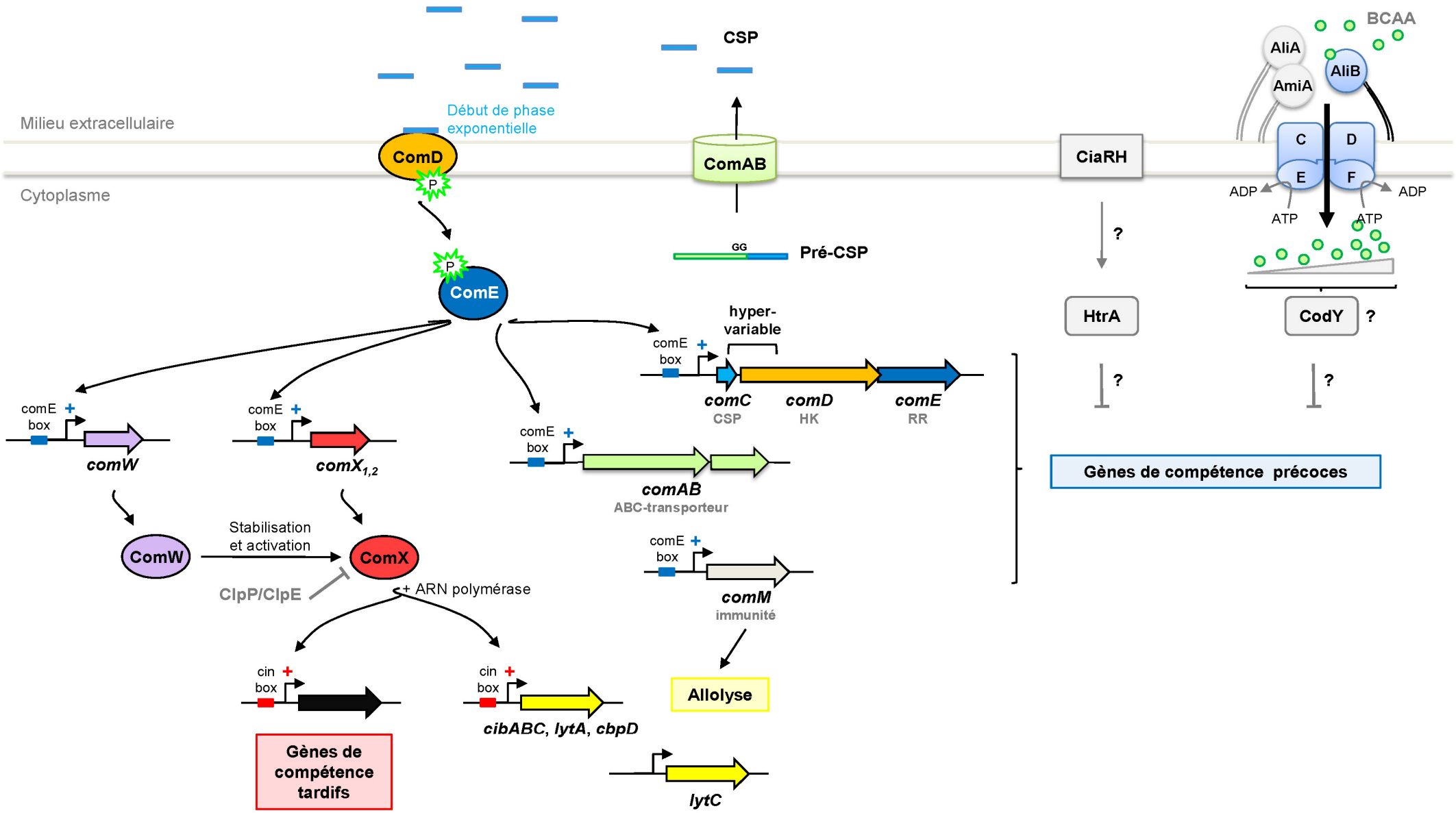


Figure 11 – Représentation schématique du système *comCDE* chez *Streptococcus pneumoniae.*

La phéromone CSP est codée par le gène *comC* et maturée au cours de son export par ComAB. Sa reconnaissance par le senseur membranaire ComD (histidine kinase d'un système à deux composants) assure la phosphorylation du régulateur de réponse ComE qui, en se fixant sur un motif appelé « comE box », assurera la transcription des gènes de compétences précoces. Aussi, l'expression de ComX, régulateur central de la compétence, permettra la transcription des gènes tardifs de compétence et d'allolyse, en se fixant sur un motif appelé « cin box » localisé dans la région promotrice de ces derniers.
BCAA, **B**ranched **C**hain **A**mino **A**cids, Acides aminés à chaîne ramifiée.

le développement de la compétence (Hui *et al.*, 1995). Par ailleurs, ComA, possèdant un domaine protéolytique conservé identifié pour le clivage des peptides contenant un motif GG, assure la maturation de CSP et ComB sa sécrétion (Håvarstein *et al.*, 1995). Aussi CSP correspond à un heptadecapeptide non modifié et présente une charge globale positive (Tableau 4) (Håvarstein *et al.*, 1995). La détection de CSP est réalisée par une histidine kinase d'un système à deux composants, ComD, dont la partie N-terminale, variable sur 80 résidus, est proposée pour la reconnaissance et la liaison de CSP (Håvarstein *et al.*, 1996). La liaison de CSP à son récepteur membranaire ComD provoquerait l'autophosphorylation de ce dernier puis le transfert du phosphate produit au régulateur de réponse ComE (Martin *et al.*, 2010; Pestova *et al.*, 1996b). Une fois phosphorylé, ComE active la transcription des gènes de compétence précoces en se fixant au niveau de deux séquences répétées directes de 9 pb, localisées dans la région promotrice des gènes cibles (Ween *et al.*, 1999). Aussi, ComE induit l'expression des opérons *comAB* et *comCDE* dans un mécanisme de rétrocontrôle positif. ComE est également requis pour l'expression du gène *comX* (facteur sigma alternatif) décrit comme régulateur central de la compétence et nécessaire à l'activation des gènes de compétence tardifs (Lee and Morrison, 1999). ComE contrôle également l'expression du gène *comW*, dont le produit intervient dans la stabilisation (prévention contre la protéolyse) et l'activation de ComX (Luo *et al.*, 2004; Sung and Morrison, 2005). Cette protéine est donc requise, au même titre que ComX, pour l'expression des gènes tardifs de compétence. Par ailleurs, ComX régule l'expression de ses gènes cibles en se fixant sur une séquence régulatrice conservée, appelée cin box (**c**ompetence **in**duiced, ou com box), localisée en position -10 (Figure 11) (Campbell *et al.*, 1998). Le déclenchement de la compétence est un phénomène en deux temps où l'expression de *comCDE* se fait en amont des gènes tardifs (régulon « cin »). Ceci provoque une induction décalée des gènes « cin » 5 min après l'expression du régulon ComE (Alloing *et al.*, 1998). La compétence est par ailleurs décrite comme un phénomène transitoire qui se déclenche en début de phase exponentielle de croissance et s'arrête 20 min après son induction (Alloing *et al.*, 1998). Des analyses de Northern blot ont suggéré la présence d'un/de mécanisme(s) impliqué(s) dans l'arrêt de la compétence. En effet, il a été montré que la transcription de *comCDE* et des gènes « cin » était diminuée malgré la présence de CSP (Alloing *et al.*, 1998). Ce(s) mécanisme(s) reste(nt) à ce jour inconnu(s) bien que le système à deux composants CiaRH ait été décrit comme jouant un rôle dans l'arrêt de la compétence en impactant de manière indirecte (et probablement via HtrA) l'expression de l'opéron *comCDE* (Figure 11) (Johnsborg and Håvarstein, 2009). Enfin, la compétence semble également être modulée par le transporteur d'oligopeptides AmiCDEF. En effet, il a été montré qu'un mutant des 3 protéines affines (AmiA, AliA et AliB) associées à ce transporteur altère le profil d'induction de la compétence (Alloing *et al.*, 1994; Alloing *et al.*,

Tableau 4 – Le peptide CSP de *Streptococcus pneumoniae*.

Précurseur			**Forme mature**	
Codé par	Nom	Taille	Séquence	Taille
comC	Pré-CSP	41 aa	EMRLSKFFRDFILQRKK	17 aa

1996). De plus, une déficience complète de la transformation est observée avec un mutant simple du gène *aliB* (Alloing *et al.*, 1996). Aussi, bien que le transporteur Ami ne soit pas requis pour l'induction de la compétence, la détection de signaux environnementaux *via* celui-ci semble néanmoins pouvoir influencer l'expression de l'opéron *comCDE* (Figure 11).

En parallèle de la régulation des gènes requis pour la transformation génétique, ComX contrôle également un phénomène d'allolyse (Figure 11), également appelé « pneumococcal fratricide », provoquant une lyse des cellules non-compétentes par les cellules compétentes d'une même souche (Claverys and Håvarstein, 2007). Ce phénomène aboutissant à la libération d'ADN dans le milieu extracellulaire serait à l'origine d'un meilleur échange génétique et pourrait asservir un phénomène de compétition intra-souches. Il n'est, par ailleurs, pas exclu que le relargage de facteurs de virulence lors de l'allolyse contribue à la pathogénicité des souches (Claverys and Håvarstein, 2007). Plusieurs gènes et un opéron ont été décrits comme impliqués dans ce phénomène et impliquent un système codant des bactériocines (*cibABC*) et trois gènes codant des hydrolases de paroi (*lytA*, *cbpD* et *lytC* codant respectivement une autolysine, une muréine hydrolase et du lysozyme) (Guiral *et al.*, 2005; Kausmally *et al.*, 2005). Les bactériocines CibAB ont été décrites comme étant le facteur déclenchant de l'allolyse en s'intégrant dans la membrane des cellules non-compétentes, les rendant alors plus sensibles et vulnérables à l'action des différentes hydrolases (Guiral *et al.*, 2005). L'expression de ces gènes de lyse (mis à part *lytC*) est par ailleurs contrôlée par ComX. Deux gènes codant des facteurs d'immunité, *cibC* et *comM*, protègent les cellules compétentes des « killing » facteurs qu'elles produisent (Guiral *et al.*, 2005; Håvarstein *et al.*, 2006). Bien que *cibC* soit régulé par ComX, l'expression de *comM* est quant à elle dépendante de ComE et assure par conséquent en amont, une protection des cellules compétentes contre les « killing » facteurs (Claverys *et al.*, 2006). Enfin, l'induction d'une troisième classe de gènes dits « retardés » et impliquant 14 gènes de réponses au stress suggère que la compétence induirait un stress temporaire aux cellules bactériennes (Claverys *et al.*, 2006). Aussi, l'induction de la compétence pouvant être générée par l'utilisation d'antibiotiques connus pour engendrer une réponse de type SOS chez les bactéries (telles que *E. coli*), il a été proposé que CSP ne soit pas un effecteur du QS à proprement parler mais une alarmone transmettant un stimulus de stress aux cellules bactériennes (Claverys *et al.*, 2006; Prudhomme *et al.*, 2006).
L'état physiologique transitoire des cellules où s'entremêle l'expression des gènes de compétence, d'allolyse et de réponse au stress a été appelé le « X-state » (Claverys *et al.*, 2006).

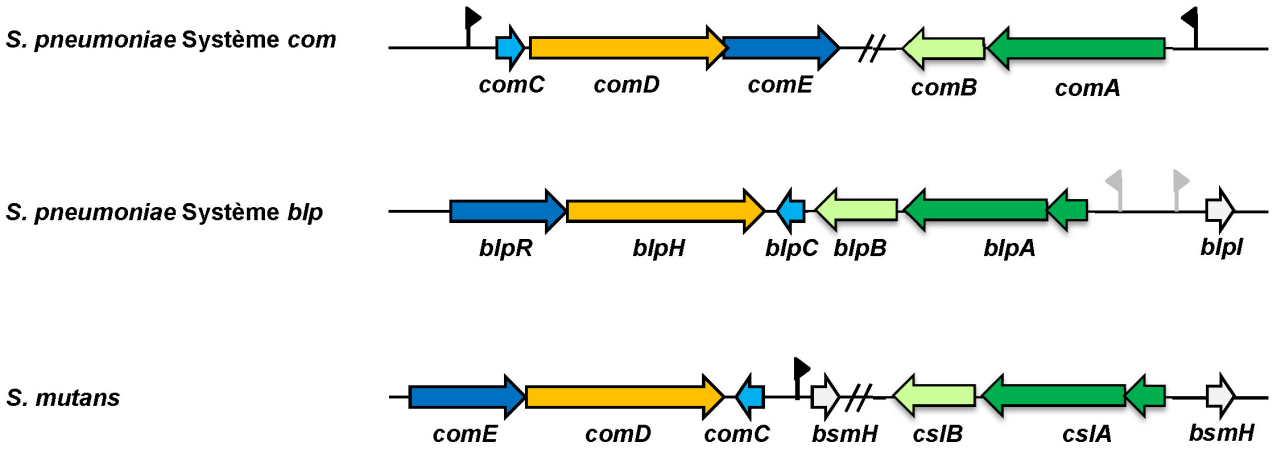


Figure 12 – Organisation génétique des systèmes Com et Blp chez *Streptococcus pneumoniae* et *Streptococcus mutans* (d'après Martin *et al*., 2006)**.**

Aparté :

En 2006, la distribution phylogénétique des gènes de compétences a souligné la conservation universelle de *comX* chez les streptocoques et les Lactobacillales et une distribution limitée du système *comCDE* (décrit ci-dessus) aux streptocoques du groupe mitis et anginosus (Martin *et al.*, 2006). De plus, les gènes *com* décrits comme contrôlant la compétence chez *S. mutans* ont constitué une deuxième classe de systèmes de compétence paralogues à *comABCDE* chez *S. pneumoniae* (Figure 12) (Martin *et al.*, 2006). En effet, Martin et ses collaborateurs ont montré que le système *com* décrit chez *S. mutans* présente une plus forte homologie et une organisation génétique similaire à celle d'un locus impliqué dans la biosynthèse d'une bactériocine chez *S. pneumoniae* (locus *blp*) (Figure 12).

Bien que de nombreuses études proposent que la compétence soit contrôlée *via* le système *com* chez *S. mutans*, l'analyse des résultats proposés et la caractérisation de certains éléments de régulation suggèrent *a contrario* l'existence d'un autre mécanisme de régulation pour ce phénomène. En effet, les gènes *comCDE* de *S. mutans* décrits comme impliqués dans le phénomène de compétence chez cette bactérie, ne semblent néanmoins pas essentiels à l'établissement de ce phénomène. En effet, seulement une diminution du nombre de transformants a été obtenue dans des mutants simples des gènes *comC*, *comD* et *comE* (Li *et al.*, 2001; Petersen *et al.*, 2005). De plus, l'identification d'un motif com box dans la région promotrice des gènes tardifs de compétence révèle une régulation de ces gènes *via* ComX (Li *et al.*, 2001). Néanmoins, l'absence d'un motif ComE box dans la région promotrice de ComX renforce l'hypothèse d'un autre mécanisme nécessaire à l'activation de ce facteur de transcription (Van der Ploeg, 2005). Enfin, alors que l'induction des gènes tardifs de compétence chez *S. pneumoniae* est rapide, celle des gènes de *S. mutans* nécessite un délai de 2h malgré l'ajout de CSP dans les surnageants de culture (Kreth *et al.*, 2005). Aussi, l'ensemble de ces résultats souligne premièrement que la phéromone CSP n'est pas le signal requis au déclenchement du phénomène et deuxièmement l'absence de lien direct entre ComE et ComX. Par ailleurs, aucun homologue *de comW* n'a été détecté chez *S. mutans* (Martin *et al.*, 2006). Aussi, la transcription de *comX* est en faveur d'un autre mécanisme de régulation, lequel sera mis en évidence en 2011 par Federle et ses collaborateurs. Ce mécanisme impliquant un régulateur transcriptionnel appartenant à la famille des Rgg est décrit dans le chapitre 3 de cette introduction bibliographique. Aussi, le mécanisme de QS décrit par CSP chez *S. mutans* serait principalement impliqué dans la production de bactériocines, la formation de biofilm et la tolérance au stress acide (Li *et al.*, 2002; Van der Ploeg, 2005).

2.2 Détection intracellulaire des peptides autoinducteurs

La détection intracellulaire des AIP nécessite la présence de deux partenaires : un transporteur d'oligopeptides impliqué dans l'import de l'AIP à l'intérieur de la cellule et un effecteur cytoplasmique interagissant de manière directe avec l'AIP. Après une présentation rapide de ces deux partenaires, deux mécanismes seront détaillés : le système PapR/PlcR décrit chez *Bacillus cereus* et le système peptides phéromones/PrgX chez *E. faecalis*.

2.2.1 Les transporteurs d'oligopeptides

2.2.1.1 Une appartenance à la famille des ABC-transporteurs

Les transporteurs d'oligopeptides (Opp, **O**ligo**p**eptide **p**ermease) appartiennent à la famille des ABC-transporteurs (**A**TP-**b**inding **c**assette). Par conséquent, ils utilisent l'hydrolyse d'un ATP comme source d'énergie pour le transport de molécules à travers la membrane plasmique (Higgins, 2007). L'organisation structurale de ces systèmes est définie par quatre protéines principales : deux protéines transmembranaires, formant le site de liaison et de translocation du substrat et deux protéines cytoplasmiques, couplant l'énergie de la catalyse de l'ATP au transport (Linton and Higgins, 2007).

2.2.1.2 Des transporteurs à protéines affines

A la différence des cellules eucaryotes qui sont capables de transporter leurs substrats du cytoplasme à l'extérieur de la cellule ou à l'intérieur d'une organelle, les bactéries ont acquit une protéine additionnelle extracellulaire, appelée SBP (**S**ubstrate **B**inding **P**rotein), leur permettant l'importation de substrats à partir du milieu extérieur (Saurin *et al.*, 1999). Les SBP associées aux Opp sont capables de fixer un ou généralement plusieurs substrats avec une grande affinité avant de le présenter au complexe ABC. Les substrats sont soit des di- ou tri-peptides ou soit, des oligopeptides pouvant présenter une grande variété de peptides dont certains ont la particularité d'être très hydrophobes (Garault *et al.*, 2002; Higgins, 2001).

Les SBP sont des lipoprotéines associées à la membrane cytoplasmique par une liaison thioesther entre un résidu cystéine en N-terminal et un diacylglycérol chez les bactéries à Gram positif (Monnet, 2003). Ces lipoprotéines sont par ailleurs synthétisées sous la forme d'un précurseur dans le cytoplasme et transloquées à travers la membrane *via* la machinerie de sécrétion Sec ou Tat (Hutchings *et al.*, 2009). La présence d'un motif lipobox (L-[A/S/T]-[G/A]-C) dans leur partie C-terminale leur assure un adressage vers la machinerie de

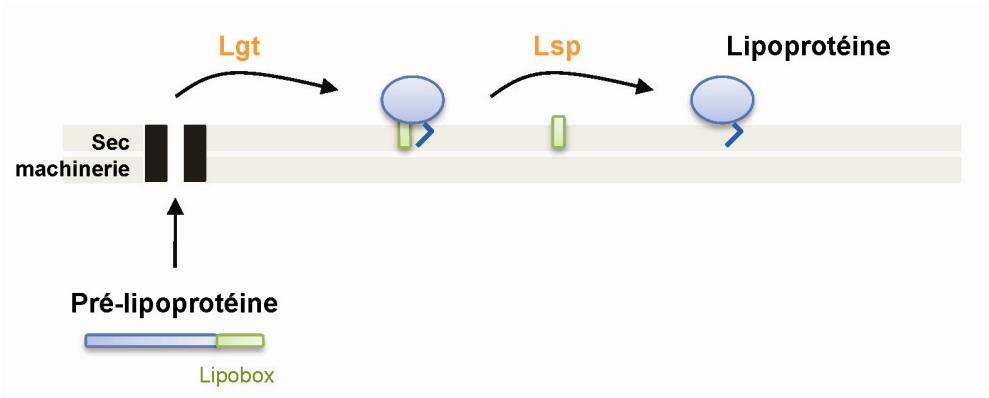


Figure 13 – Voie de biosynthèse des lipoprotéines chez les bactéries à Gram positif (d'après Hutchings *et al.*, 2009).

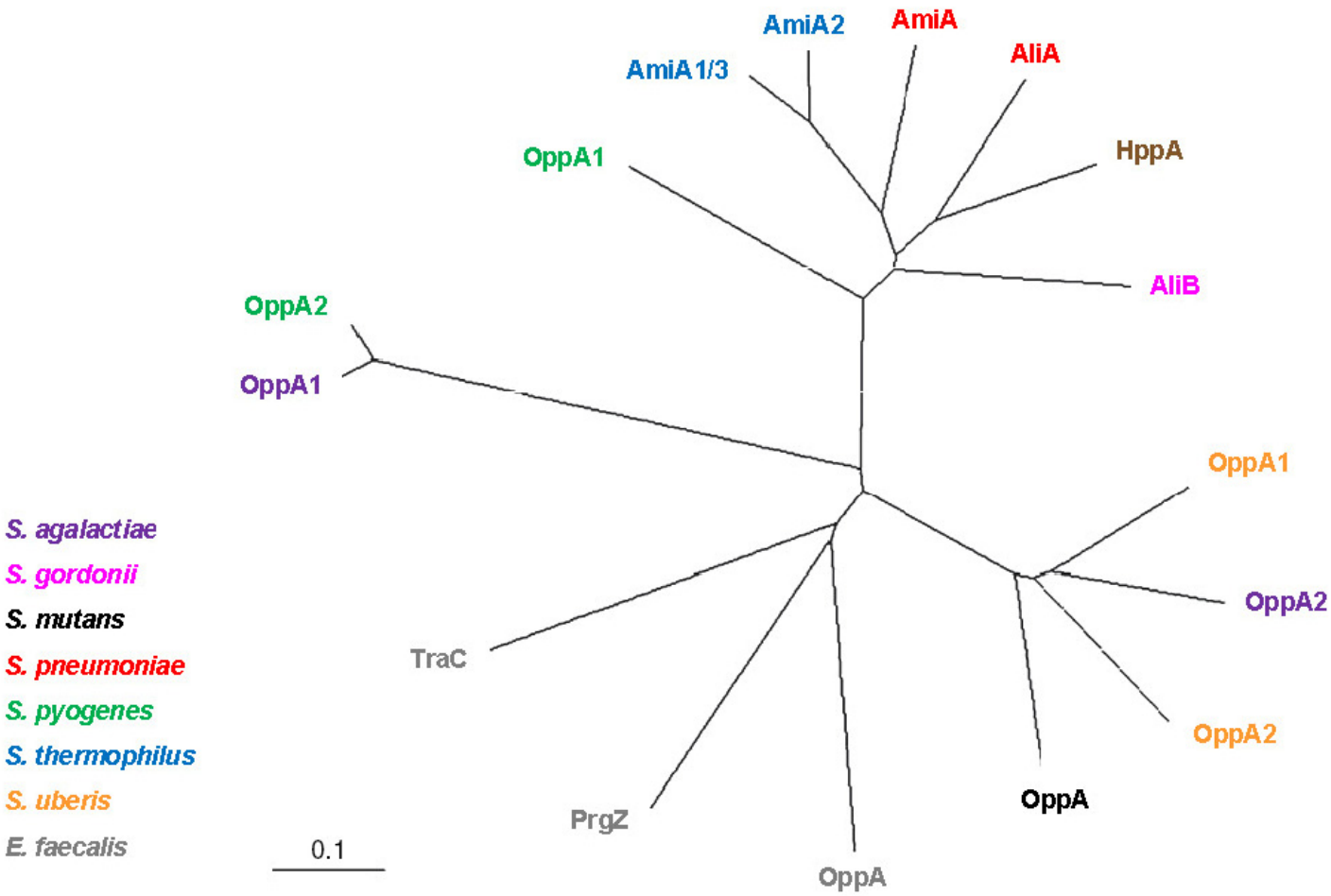


Figure 14 – Arbre phylogénétique des protéines affines OppA chez différents Firmicutes (d'après Nepomuceno *et al.*, 2007).
Les branches diffèrent lorsque la divergence entre les séquences atteint 10%. Les protéines TraC et PrgZ chez *E. faecalis* sont spécifiques des molécules de signalisation.

biogénèse des lipoprotéines. Chez les bactéries à Gram positif, une modification lipidique par une pro-lipoprotéine diacylglycéryl transférase (Lgt) puis le clivage de la séquence signale (lipobox) par une lipoprotéine signal peptidase (Lsp) sont les étapes assurant la biosynthèse des lipoprotéines (Figure 13) (Hutchings *et al.*, 2009). D'autres peptidases, telles que Eep (**E**nhanced **e**xpression of **p**heromone), participent également à la production de certaines lipoprotéines (An *et al.*, 1999; Denham *et al.*, 2008).

Chez les bactéries à Gram positif, les systèmes Opp sont généralement codés en opéron et plusieurs lipoprotéines peuvent être associées à un même ABC-transporteur (Monnet, 2003). Les lipoprotéines présentent par ailleurs des niveaux d'identités différents et variables entre les espèces, suggérant ainsi une évolution en terme de spécificité de substrats pour une espèce donnée (Figure 14). Aussi, chaque lipoprotéine présentant des affinités de substrats différentes, leur accumulation procure un avantage nutritionnel et physiologique aux bactéries (Garault *et al.*, 2002; Juille *et al.*, 2005). *S. thermophilus* représenterait alors le streptocoque ayant le plus exploité cet avantage. En effet, ce dernier possède trois SBP capables de réimporter des peptides contenant de 2 à 23 résidus (Garault *et al.*, 2002). Le nombre de lipoprotéines peut cependant s'avérer variable entre deux souches d'une même espèce. Ceci a également été décrit chez *S. thermophilus* où bien que certaines souches contiennent jusqu'à 3 lipoprotéines (Garault *et al.*, 2002), d'autres en revanche n'en n'auraient accumulé que deux (Bolotin *et al.*, 2004; Juille *et al.*, 2005). Enfin, la présence de transposases dans l'environnement génétique proche des opérons *opp* dans la quasi-totalité des génomes des streptocoques, suggère que ces derniers auraient été acquis par transfert horizontal (Nepomuceno *et al.*, 2007).

2.2.2 Les effecteurs de la famille RNPP

Une fois réimportés par les systèmes Opp, la reconnaissance de l'AIP à son effecteur intracellulaire est requise pour assurer la régulation des gènes cibles des systèmes de QS. Deux types d'effecteurs sont décrits dans la littérature, les phosphatases et les régulateurs transcriptionnels. En 2007, des analyses phylogénétiques ont suggéré que ces effecteurs dérivent d'un ancêtre commun ; ils ont alors été rassemblés au sein d'une famille dénommée RNPP, pour Rap, NprR, PlcR et PrgX (Declerck *et al.*, 2007). En effet, par une analyse de type Psi-Blast, les auteurs ont montré que ces quatre effecteurs partageaient des caractéristiques communes, telles que :

(i) Leur appartenance au phylum des Firmicutes et plus précisément aux classes des *Bacilli* et *Clostridia*,

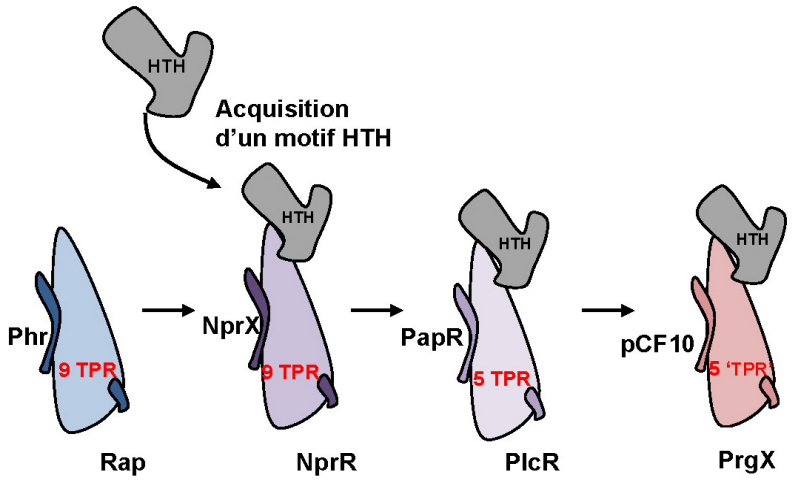


Figure 15 – Relations évolutives des protéines de la famille RNPP (Thèse L. Bouillaut, 2007)**.**

(ii) Leur interaction avec une molécule de signalisation, importée dans le cytoplasme par un système Opp, dont la forme mature est issue du clivage d'un précurseur,

(iii) La présence d'un domaine régulateur en C-terminal caractérisé par un nombre variable de domaines TPR (**t**etratrico**p**eptide **r**epeat). Ces domaines ont été décrits comme impliqués dans les interactions protéine/protéine ou peptide/protéines (Goebl and Yanagida, 1991).

Aussi, sur la base de leur séquence, du nombre de domaines TPR et la présence d'un domaine hélice-tour-hélice (HTH) appartenant à la famille des XRE (**X**enobiotic **R**egulatory **E**lement, PFAM01381) dans leur partie N-terminale, les relations évolutives liant ces effecteurs ont été proposées (Figure 15). La phosphatase Rap des *Bacilli*, décrite plus proche du régulateur transcriptionnel NprR mais ne contenant pas de domaine HTH, a alors été proposée comme effecteur le plus ancestral. S'en suit, le régulateur transcriptionnel NprR présentant le même nombre de domaines TPR que les phosphatases Rap (équivalent à 9) mais ayant acquis un domaine HTH au cours de l'évolution. Arrive ensuite le régulateur transcriptionnel PlcR possédant 5 domaines TPR et, le régulateur transcriptionnel PrgX structuralement très proche de PlcR (Declerck *et al.*, 2007).

Par ailleurs, ces analyses de type Psi-Blast ont également fait ressortir une homologie entre PlcR et PrgX et d'autres régulateurs transcriptionnels de la famille des Firmicutes. Ces régulateurs sont composés d'un domaine HTH en N-terminal et d'un domaine TPR-like (11 hélices) prédit comme plus proche de celui de PrgX (Declerck *et al.*, 2007). Parmi ces régulateurs, on retrouve les protéines Rgg (Chapitre 3). Aussi, cette partie de mon introduction bibliographique se concentrera sur les systèmes de QS impliquant les régulateurs PlcR et PrgX.

2.2.3 Le système PapR/PlcR chez *Bacillus cereus* : virulence

Les espèces de *Bacillus* produisent une grande quantité d'enzymes dégradatives et de toxines en fin de phase exponentielle ou en début de phase stationnaire de croissance. En 1996, l'expression du gène *plcA*, codant la phospholipase C, a été décrite sous le contrôle d'un régulateur transcriptionnel nommé PlcR (**P**hospho**l**ipase **C** **r**egulator) (Lereclus *et al.*, 1996). Depuis, plusieurs études ont décrit de nombreux gènes cibles de ce régulateur définissant aujourd'hui un régulon composé de plus de 100 gènes, aléatoirement distribués sur le chromosome. Ce régulon, composé de gènes et d'opérons, est constitué majoritairement de facteurs de virulences (enzymes dégradatives, protéines de surfaces,

toxines, etc.) mais également de senseurs et de transporteurs (Agaisse *et al.*, 1999; Gohar *et al.*, 2002; Gohar *et al.*, 2008; Ivanova *et al.*, 2003; Økstad *et al.*, 1999).

2.2.3.1 L'histoire d'un motif de fixation

Le régulateur transcriptionnel PlcR est un régulateur pléiotrope spécifique du groupe *cereus* regroupant *B. anthracis*, *B. cereus*, *B. mycoides*, *B. pseudomycoides*, *B. thuringiensis* et *B. wiehenstephanensis* (Agaisse *et al.*, 1999). En 1999, l'analyse des régions promotrices des 14 gènes connus régulés par PlcR a permis d'identifier une séquence palindromique conservée (TATGNA-N_4-TNCATA). Cette séquence, dénommée PlcR box, a été proposée comme site putatif de fixation du régulateur à l'ADN (Agaisse *et al.*, 1999). Ce palindrome permet, par ailleurs, la transcription de deux gènes transcrits de manière divergente, lorsqu'il est localisé dans la région intergénique de ces derniers (Agaisse *et al.*, 1999; Økstad *et al.*, 1999). En outre, la distance entre le motif et la région -10 du promoteur du gène cible est variable et peut aller de 20 à plus de 200 pb (Agaisse *et al.*, 1999; Slamti and Lereclus, 2002). Plus récemment, une étude a permis de discriminer les gènes réellement régulés par PlcR dans un pool de gènes contenant la PlcR box dans leur région promotrice (Gohar *et al.*, 2008). L'analyse des régions promotrices des gènes régulés par PlcR a démontré que les séquences en amont et en aval du motif étaient significativement plus riches en AT, suggérant un rôle non négligeable de cet environnement génétique dans la liaison de PlcR à l'ADN et/ou sur l'activité transcriptionnelle du promoteur (Gohar *et al.*, 2008).

2.2.3.2 L'expression de PlcR est contrôlée par un mécanisme de QS

L'utilisation de fusions transcriptionnelles *lacZ* en système hétérologue chez *B. subtilis*, a permis de suggérer que l'expression du gène *plcA*, régulé par PlcR était également dépendante du produit codé par le gène localisé en aval de *plcR* et nommé *orf2* (Lereclus *et al.*, 1996). L'identification de la PlcR box en amont du gène *orf2* a alors suggéré que ce dernier soit sous le contrôle de PlcR. De plus, la présence d'une séquence signale dans la partie N-terminale du produit du gène *orf2* suggère la sécrétion de ce dernier (Agaisse *et al.*, 1999). Une collection de mutants obtenue par mutagénèse aléatoire a ensuite conduit à l'identification d'un transporteur d'oligopeptides essentiel à l'expression du régulon PlcR (Gominet *et al.*, 2001). Le rôle du transporteur d'oligopeptides a par ailleurs été caractérisé de manière indépendante à l'effet répresseur de SpoA~P (Lereclus *et al.*, 2000). L'ensemble de ces travaux a alors suggéré que la régulation de PlcR pouvait être contrôlée par un mécanisme de QS.

Tableau 5 – Le peptide PapR7 de *Bacillus cereus*.

Précurseur			Forme mature	
Codé par	Nom	Taille	Séquence	Taille
papR	Pré-PapR	48 aa	ADLPFEF	7 aa

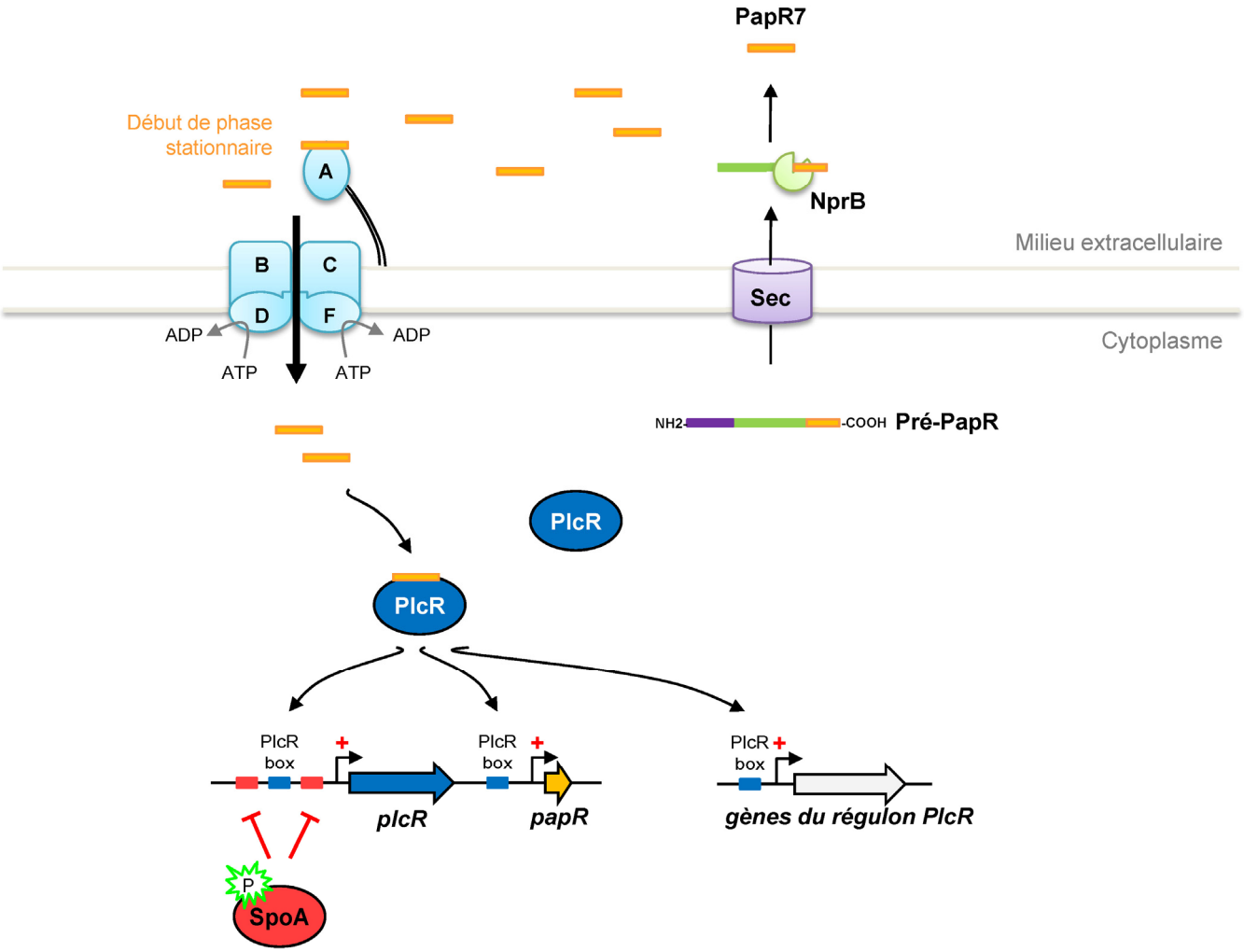


Figure 16 – Représentation schématique du système PapR/PlcR chez *Bacillus cereus*.
Le peptide PapR7, codé par le gène *papR*, est maturé et sécrété dans le milieu extracellulaire avant d'être détectée (en début de phase stationnaire) et importé dans la cellule par le système de transport d'oligopeptides OppBCDF. Dans le cytoplasme, l'interaction de PapR7 avec le régulateur transcriptionnel PlcR induit *in fine* la liaison de ce dernier à l'ADN sur un motif appelé « PlcR box », localisé dans la région promotrice de ces gènes cibles. Le régulateur de réponse SpoA~P réprime quant à lui la transcription du gène *plcR* en se fixant sur ses séquences cibles localisées dans la région promotrice de ce dernier.

Aussi, en 2002, la caractérisation du produit codé par le gène *orf2* a été entreprise sur l'expression du régulon PlcR, qui s'est avérée être abolie dans un mutant Δ*orf2* (Slamti and Lereclus, 2002). Le produit du gène *orf2* présentant des caractéristiques d'export, des complémentations ont alors été réalisées avec des surnageants de culture. Seulement les surnageants de la souche wt (vs Δ*orf2*) étant capable de restaurer l'expression du régulon PlcR, les auteurs ont conclu que le produit du gène *orf2* était d'une part sécrété et d'autre part essentiel à l'activation de PlcR. Sur la base de ces résultats, le gène *orf2* a été nommé *papR* pour **P**lcR **a**ctivating **p**eptide (Slamti and Lereclus, 2002). Le gène *papR* code un peptide de 48 acides aminés, dont la forme active est probablement libérée suite à un clivage du précurseur lors de son export. En 2009, le rôle de la protéase de paroi NprB dans la maturation de PapR sera démontré (Pomerantsev *et al.*, 2009).

Par l'utilisation de peptides synthétiques, les auteurs ont montré que l'extrémité C-terminale du peptide est nécessaire à l'activité du régulon PlcR et que des tailles variables du peptide, composées de 5, 7 ou 9 acides aminés induisaient de manière équivalente le régulon PlcR (Slamti and Lereclus, 2002). En 2008, Bouillaut et ses collaborateurs identifient l'heptapeptide, PapR7, comme étant la forme physiologique majoritaire de ce peptide (Tableau 5) (Bouillaut *et al.*, 2008).

Afin d'étudier le rôle du transporteur d'oligopeptides Opp et de mettre en évidence la réimportation de PapR à l'intérieur de la cellule *via* ce système de transport, Slamti et Lereclus ont introduit un plasmide inductible surproduisant PapR5 dans une souche présentant un contexte génétique délété du gène *oppB*. En comparaison avec la souche sauvage, ils ont étudié l'activité hémolytique du mutant et montré que celle-ci était restaurée dans le mutant Δ*oppB* seulement lors de l'induction du plasmide surproducteur. De même, l'activité de la fusion transcriptionnelle *plcA::lacZ* a été restaurée dans un mutant Δ*papR* lors de l'induction de ce même plasmide surproducteur. Aussi, l'activation du régulon PlcR a été décrite comme effective par la présence de PapR dans le cytoplasme de la bactérie (Slamti and Lereclus, 2002). Pour définir plus précisément le rôle de PapR, un plasmide inductible surproduisant PlcR a ensuite été introduit dans deux souches présentant la fusion transcriptionnelle *plcA::lacZ* dans un contexte génétique sauvage et délété du gène *papR*. Les auteurs travaillant dans un milieu ne permettant pas l'expression du régulon PlcR, l'induction du plasmide surproduisant PlcR est alors nécessaire pour en assurer son expression. Aussi, une complémentation étant visible dans la souche sauvage seulement après induction du plasmide, les auteurs ont montré que PlcR requiert la présence de PapR pour être fonctionnel et jouer son rôle d'activateur (Slamti and Lereclus, 2002). Enfin, des expériences d'empreinte à la DNase ont démontré une fixation de PlcR sur la région promotrice du gène *plcA* PapR-dépendante (Slamti and Lereclus, 2002). L'ensemble de ces résultats a alors conduit au modèle suivant, présenté dans la Figure 16. Le gène *papR* est

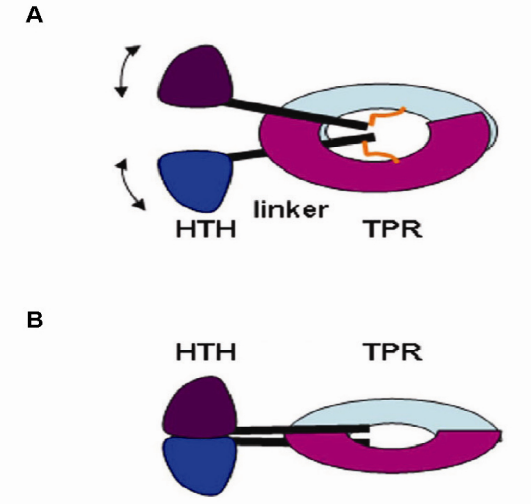


Figure 17 – Représentation schématique de l'organisation du domaine HTH du régulateur PlcR en présence (A) ou absence (B) du peptide PapR (Declerck *et al.*, 2007)**.**

exprimé à un niveau basal en phase exponentielle. La forme précurseur du peptide phéromone est ensuite sécrétée puis maturée par NprB. En fin de phase exponentielle / début de phase stationnaire, PapR atteint sa concentration seuil induisant sa détection puis sa réimportation à l'intérieur de la cellule par le transporteur d'oligopeptides Opp. L'expression des gènes du régulon PlcR est alors assurée par la formation d'un complexe PlcR/PapR, permettant la fixation de l'activateur PlcR dans la région promotrice des gènes cibles.

2.2.3.3 Structure et interaction PapR/PlcR

PlcR est composé d'un domaine N-terminal de liaison à l'ADN et d'un domaine C-terminal comprenant 5 domaines TPR. La structure de ce régulateur a été déterminée en présence ou en absence du peptide PapR5 (Declerck *et al.*, 2007). Dans les deux cas, PlcR présente une structure dimérique et un réarrangement du domaine HTH est visible en présence du peptide. Ce changement structural nécessaire à l'activation du régulateur a été à la base d'un modèle où la liaison de PapR aux domaines TPR de PlcR entrainerait l'ouverture des hélices du domaine HTH (Figure 17). En fonction de la concentration en phéromone, le complexe PapR/PlcR aurait, par ailleurs, tendance à s'oligomériser (Declerck *et al.*, 2007). Bien qu'une région (résidus 95 à 144) ait été décrite comme impliquée dans la multimérisation de PlcR, aucune substitution au sein de celle-ci n'a encore clairement été démontrée comme impliquée dans ce phénomène (Bouillaut *et al.*, 2008).

La comparaison des séquences de 29 protéines PlcR a souligné que seul l'acide aminé en position 278 différait de manière spécifique selon les groupes (Bouillaut *et al.*, 2008). La construction de protéines chimères à partir de ce résidu a ensuite montré que l'extrémité C-terminale des protéines PlcR est nécessaire à l'activation du régulon PlcR et interfère sur le niveau d'expression de ce dernier (Bouillaut *et al.*, 2008).

2.2.3.4 PapR/PlcR : un système de régulation souche-spécifique

La comparaison de la séquence peptidique de PapR parmi 15 souches de *Bacillus* a mis en évidence, contrairement à la partie N-terminale hautement conservée, une divergence de la partie C-terminale contenant la forme mature du peptide (Slamti and Lereclus, 2002). Des tests de complémentation sur milieu gélosé ont montré que le premier acide aminé des pentapeptides (leucine, méthionine ou valine) joue un rôle dans la spécificité de réponse. Des expériences d'empreintes à la DNase ont par ailleurs spécifié le rôle de cet acide aminé

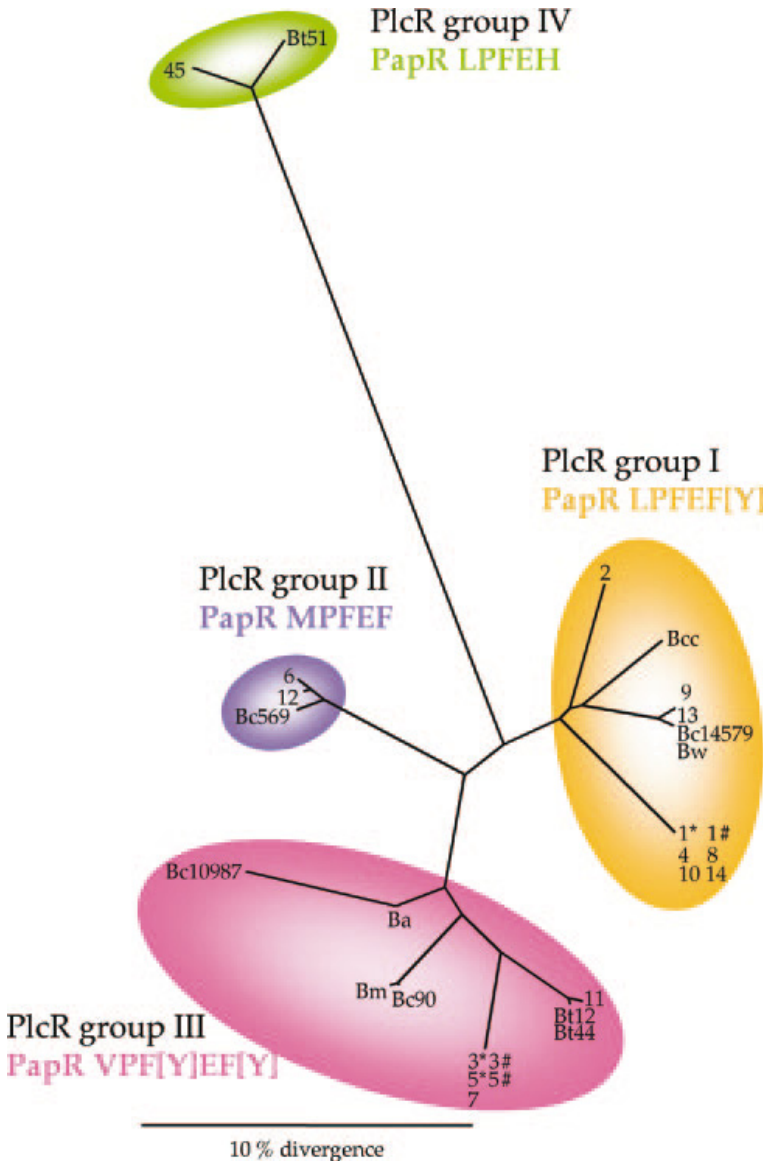


Figure 18 – Arbre phylogénétique des protéines PlcR et des peptides PapR apparentés (Slamti and Lereclus, 2005)**.**

dans l'interaction PlcR/PapR (Slamti and Lereclus, 2002). En 2005, l'alignement de 29 séquences de PapR couplé à des tests de complémentation a par ailleurs rendu compte que cette spécificité était également due au dernier acide aminé du pentapeptide. Ces résultats ont alors permis de définir quatre groupes de peptides phéromones. La construction d'un arbre phylogénétique des protéines PlcR a ensuite souligné une coévolution des protéines PlcR et des formes matures pentapeptidiques identifiées (Figure 18) (Slamti and Lereclus, 2005). Aussi, en 2008, suite à l'identification de la forme physiologique majoritaire du peptide PapR (correspondant à l'heptapeptide, PapR7), la comparaison de l'activité des formes penta et heptapeptidiques a tout d'abord été entreprise. A l'exception du groupe II où l'heptapeptide est 6,5 fois plus actif, les penta et heptapeptides présentent un niveau d'activité similaire (Bouillaut *et al.*, 2008). Afin d'étudier la spécificité des systèmes PapR/PlcR, une souche de chaque groupe, contenant une fusion transcriptionnelle *plcA::lacZ* dans son génome, a alors été complémentée avec les quatre heptapeptides considérés au sein des groupes identifiés. Bien que chaque paire légitime PapR/PlcR se soit avérée être plus spécifique, la présence d'un cross-talk plus ou moins important entre les groupes I, II et III a pu être démontré. Néanmoins, les niveaux d'activité de deux peptides activant les régulons PlcR des groupes apparentés ne sont pas forcément identiques. A titre d'exemple, bien que le peptide PapRI active le régulon PlcR du groupe III, le peptide PapRIII en revanche n'induit que faiblement le régulon PlcR du groupe I. Par ailleurs, aucune cross-activation n'a été décrite pour le groupe IV, phylogénétiquement plus éloigné (Bouillaut *et al.*, 2008).

2.2.3.5 Un régulon impliqué dans la virulence

L'implication du régulon PlcR dans la virulence de *B. cereus* ou *B. thuringiensis* a été démontrée *in vivo*. En effet, une mutation dans le gène *plcR* conduit à une forte diminution de la virulence de ces deux bactéries chez l'insecte ou des mammifères tels que le lapin ou la souris (Callegan *et al.*, 2003; Salamitou *et al.*, 2000). Par ailleurs, la virulence de la bactérie chez l'insecte est également diminuée dans des mutants des gènes *papR* et *oppB* (Gominet *et al.*, 2001; Slamti and Lereclus, 2002).

2.2.4 Le système peptides phéromones / PrgX chez *Enterococcus faecalis*: conjugaison

Bactéries communément impliquées dans les infections nosocomiales, les entérocoques sont largement connus pour leurs propriétés de conjugaison leur permettant d'acquérir et de

Tableau 6 – Les deux principaux plasmides conjugatifs décrits chez *Enterococcus faecalis*.

	pCF10	**pAD1**
Taille du plasmide	54 kb	60 kb
Sex-phéromone	cCF10	cAD1
Anti-phéromone	iCF10	iAD1
Effecteur intracellulaire	PrgX	TraA
Protéine affine	PrgZ	TraC
Résistances	Tétracycline	Hémolysine
		Bactériocine
		Résistance aux UV

disséminer de nombreux déterminants impliqués dans la virulence et la résistance aux antibiotiques. Ce phénomène, répondant à un mécanisme de communication cellulaire, diffère néanmoins de ceux précédemment décrits où la molécule de signalisation produite est détectée par l'organisme émetteur. En effet, dans le cas de la conjugaison, l'information initiée par la molécule de signalisation traduit la présence de deux populations bactériennes ; la première dite donneuse et possédant le plasmide conjugatif, la seconde dite receveuse et produisant la molécule de signalisation requise au déclenchement du processus de conjugaison par les cellules donneuses.

Il a été montré que les souches d'*E. faecalis* ne possédant pas de plasmide conjugatif secrètent de multiples sex-phéromones, chacune spécifique d'un groupe de plasmides conjugatifs (Clewell *et al.*, 2000). De plus, après acquisition par la cellule receveuse d'un plasmide conjugatif donné, la production de la sex-phéromone apparentée est arrêtée et / ou masquée par la production d'un peptide inhibiteur compétiteur (anti-phéromone) empêchant les cellules donneuses d'être (auto-)induites par la production de phéromones endogènes (Clewell *et al.*, 2000).

De nombreux plasmides conjugatifs sont décrits dans la littérature (Tableau 6) et les gènes de conjugaison les composant présentent de fortes similarités (Dunny and Leonard, 1997). Dans le cadre de cette introduction bibliographique, ce processus sera détaillé pour l'exemple le plus documenté : le plasmide pCF10 conférant aux cellules receveuses une résistance à la tétracycline.

2.2.4.1 Le mécanisme de transfert conjugatif

L'induction du transfert conjugatif est basée sur la détection puis la réimportation de sex-phéromones exogènes dans le cytoplasme de cellules donneuses. Aussi, dans le but d'acquérir de nouveaux plasmides conjugatifs, tel que pCF10, les cellules receveuses sécrètent dans le milieu extracellulaire la sex-phéromone apparentée, dans ce cas cCF10. Les gènes de conjugaison codés par le plasmide pCF10 sont appelés *prg* pour **p**heromone **r**esponsive **g**ene. Le mécanisme décrit dans cette partie est illustré par la Figure 19.

De manière simplifiée, la détection de la sex-phéromone cCF10 par les cellules donneuses est assurée par la protéine affine PrgZ. Cette reconnaissance est effective à de faible concentration en sex-phéromone (2.10^{-11} M) dans le milieu extracellulaire (Dunny and Leonard, 1997; Leonard *et al.*, 1996) et initie un mécanisme de signalisation qui résulte en l'activation de l'expression des fonctions de conjugaison. En effet, une fois réimportée à

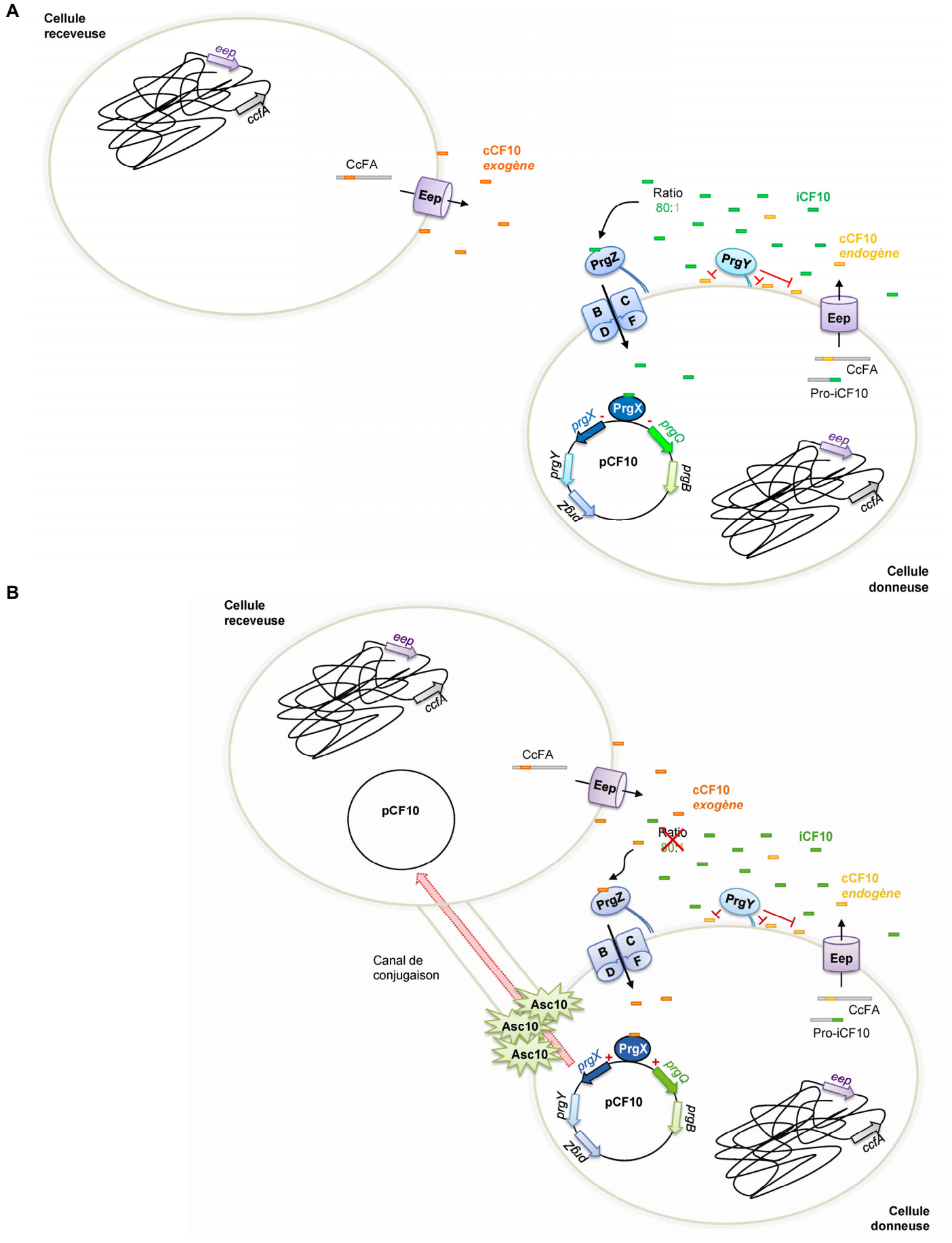


Figure 19 – Représentation schématique du transfert conjugatif pCF10 chez *Enterococcus faecalis*.
(A) La sex-phéromone cCF10 est exprimée par les cellules donneuses et receveuses. L'anti-phéromone iCF10 produite par les cellules donneuses est sécrétée et prévient leur auto-induction. **(B)** La présence d'une cellule receveuse à proximité d'une cellule donneuse, modifie le ratio iCF10/cCF10 et permet la détection de cCF10 par la protéine affine PrgZ. Une fois importée dans le cytoplasme par le transporteur d'oligopeptides OppBCDF, cCF10 induit l'expression des gènes de transfert, comprenant le gène *prgB* codant une adhésine de surface médiant l'agrégation des cellules donneuse et receveuse. La formation d'un canal dit de conjugaison permet ensuite le transfert, sous forme simple brin, du plasmide pCF10 à la cellule receveuse.

l'intérieur de la cellule *via* le système de transport d'oligopeptides OppBCDF (Dunny and Leonard, 1997; Leonard *et al.*, 1996), cCF10 interagit avec son effecteur intracellulaire, le régulateur transcriptionnel PrgX (Bae *et al.*, 2000). Cette interaction abolissant la fonction répresseur de ce dernier permet alors la transcription de nombreux déterminants impliqués dans le transfert conjugatif (Dunny *et al.*, 2001). En premier lieu, ceci induit la production d'une adhésine de surface Asc10 codée par le gène *prgB*. L'expression de cette adhésine à la surface de la cellule donneuse facilite l'attachement des cellules donneuses induites à une substance de liaison appelée EBS (**e**nterococcal **b**inding **s**ubstance). L'agrégation formée par l'association adhésine-EBS génère un canal dit de conjugaison qui permettra le transfert du plasmide conjugatif à la cellule receveuse (Dunny and Leonard, 1997). Par ailleurs, le plasmide est transféré sous la forme d'ADNsb avant d'être recircularisé et de servir de matrice pour la synthèse du brin complémentaire (Dunny *et al.*, 2001).

Il a été montré que les cellules donneuses et receveuses produisent une quantité similaire de sex-phéromones, résultant de l'expression d'un gène chromosomique, *ccfA* (Chandler and Dunny, 2004). Néanmoins, afin d'assurer le transfert du plasmide conjugatif à travers un apport exogène de phéromones (cellules receveuses), les cellules donneuses ont développé une stratégie assurant la non détection des phéromones endogènes. Aussi, une fois acquis par les cellules initialement receveuses, le plasmide conjugatif pCF10 code deux déterminants génétiques impliqués dans un mécanisme de contrôle d'auto-induction des sex-phéromones endogènes (Dunny and Leonard, 1997). La protéine membranaire PrgY a été décrite comme impliquée dans ce phénomène et affecte la production des sex-phéromones endogènes par les cellules donneuses, sans modifier le niveau de cCF10 exogènes produites par les cellules receveuses. En effet, bien que ce mécanisme ne soit pas encore élucidé, PrgY jouerait un rôle au niveau de la paroi des cellules donneuses en séquestrant ou dégradant les sex-phéromones endogènes et par conséquent, induirait une diminution de leur quantité associée à la paroi des cellules donneuses (Buttaro *et al.*, 2000). Le second déterminant correspond à un peptide inhibiteur (anti-phéromone) secrété dans le milieu extracellulaire et qui neutraliserait les sex-phéromones endogènes excrétées par les cellules donneuses et qui échappent au contrôle de PrgY (Chandler and Dunny, 2004; Dunny *et al.*, 2001). Ce peptide, codé par le gène *prgQ*, est appelé iCF10. Ces peptides inhibiteurs sont également proposés comme peptides compétitifs à la liaison des sex-phéromones endogènes à la protéine affine PrgZ (Bensing *et al.*, 1997; Chandler and Dunny, 2004). Aussi, l'induction des cellules donneuses résulte d'une régulation fine leur permettant d'être extrêmement sensibles aux apports exogènes en cCF10 modifiant ainsi le ratio extracellulaire iCF10/cCF10 (Figure 19) (Kozlowicz *et al.*, 2006a) (Dunny and Johnson, 2011).

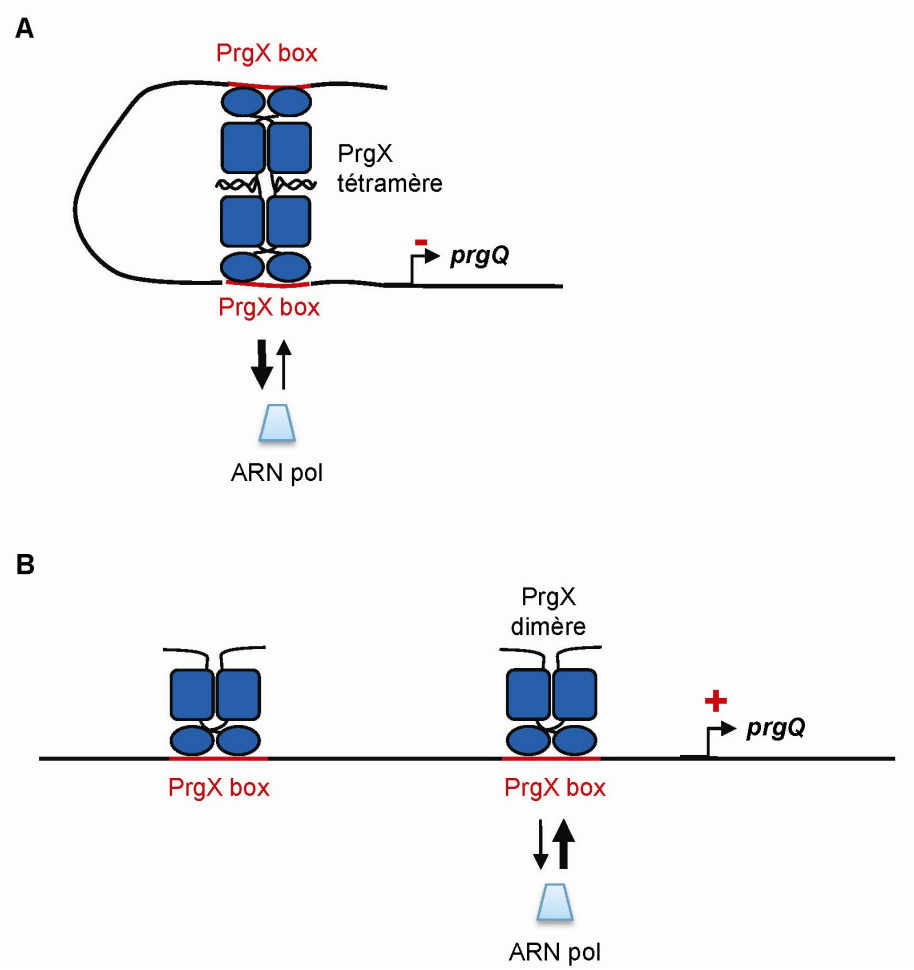


Figure 20 – Influence de l'oligomérisation du régulateur transcriptionnel PrgX sur l'expression des gènes de conjugaison (adapté de Dunny, 2007)**.**
A – Cellule donneuse non-induite ; B – Cellule donneuse induite.

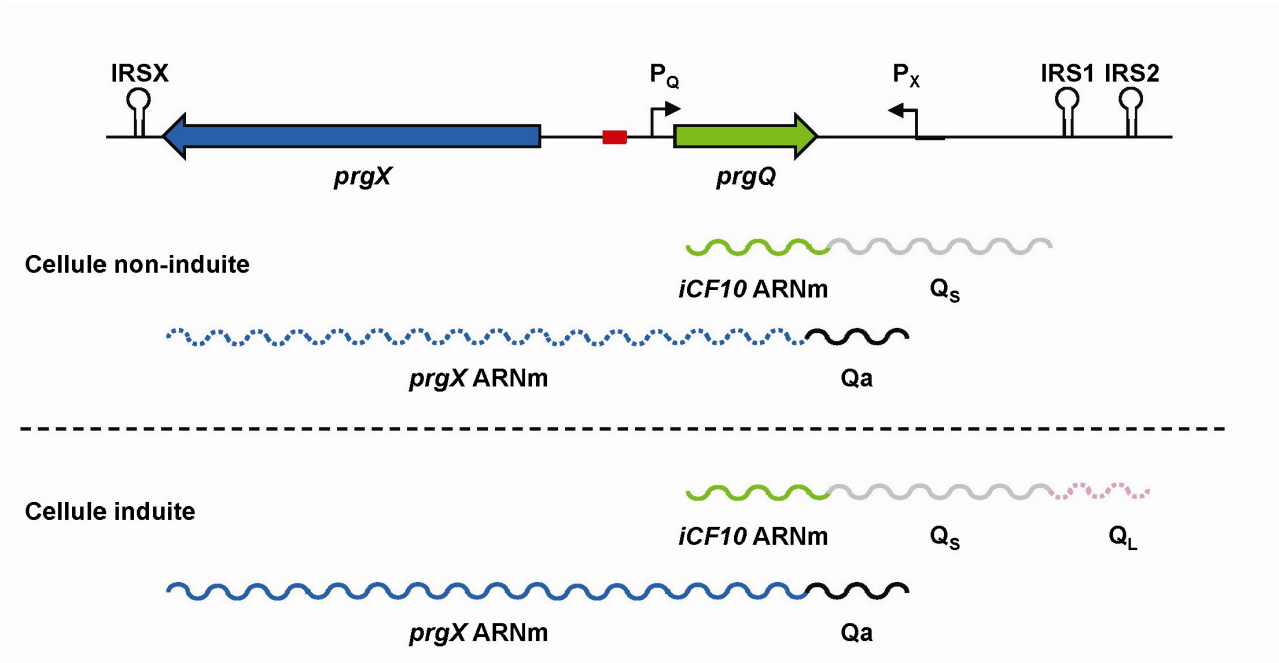


Figure 21 – Contrôle post-transcriptionnel de l'expression des gènes de conjugaison (adapté de Dunny and Johnson, 2011 et Bae *et al.*, 2002)**.**

2.2.4.2 Le régulateur transcriptionnel PrgX : mode d'action

Le mécanisme de régulation de la conjugaison est un phénomène extrêmement complexe impliquant de nombreux contrôles au niveau transcriptionnel et traductionnel (Dunny *et al.*, 2001).

PrgX est le régulateur central de ce mécanisme et des études structurales ont montré que son activité est modifiée en fonction du peptide phéromone auquel il se lie, *i.e.* la sex-phéromone cCF10 ou l'anti-phéromone iCF10. Néanmoins, quelque soit le peptide phéromone qu'il lie, PrgX est toujours lié à l'ADN au niveau de deux séquences localisées dans la région intergénique *prgX-Q* (Figure 20) (Bae *et al.*, 2002). En absence d'induction, PrgX interagit avec l'anti-phéromone iCF10 internalisée. Cette interaction assurant la répression de la machinerie de conjugaison est corrélée à une structure tétramérique du régulateur impliquant la formation d'une boucle au niveau de l'ADN (Kozlowicz *et al.*, 2006b; Shi *et al.*, 2005). L'affinité de PrgX à iCF10 étant plus faible que celle de la sex-phéromone cCF10 (ration 80 :1), l'induction des cellules donneuses par cette dernière entraine la liaison de PrgX à cCF10. Aussi, l'interaction de PrgX avec la sex-phéromone provoque un changement d'oligomérisation du régulateur, ce dernier passant alors d'une forme tétramérique à une forme dimérique. Ce changement de structure rompt la boucle d'ADN formée en présence du tétramère et par conséquent entraine une levée de la répression de PrgX et l'activation du promoteur P_Q (gène *prgQ*) (Figure 20) (Kozlowicz *et al.*, 2006b; Shi *et al.*, 2005).

Lors de l'induction des cellules donneuses, le taux de transcrits de plusieurs ARN régulateurs est également modifié (Figure 21) (Kozlowicz *et al.*, 2006a). En effet, en plus de coder l'anti-phéromone iCF10, différents transcrits sont produits à partir du promoteur P_Q :

- un transcrit Q_S, de 430 nt, essentiel à l'activation des gènes impliqués dans la machinerie de transfert et,
- un transcrit Q_L, de 100 nt, stabilisant la transcription du gène *prgB*.

Par ailleurs, l'augmentation de la transcription du gène *prgQ* diminue l'effet répresseur d'un ARN antisens Qa, transcrit de manière divergente à partir de l'extrémité 5' de l'ARNm du gène *prgX* (promoteur P_X) (Figure 21) (Johnsborg and Håvarstein, 2009). Cet effet répresseur est à la base de l'augmentation du taux de transcrits de *prgX* et assure ainsi l'auto-induction du régulateur et par conséquent l'activation de la machinerie de transfert (Kozlowicz *et al.*, 2006a). Enfin, en absence de sex-phéromone cCF10, l'interaction entre l'ARN antisens Qa et le transcrit Qs prévient l'initiation des transcrits au niveau du promoteur P_Q, ces deux transcrits étant complémentaires. Aussi, la présence de l'ARN régulateur Qa

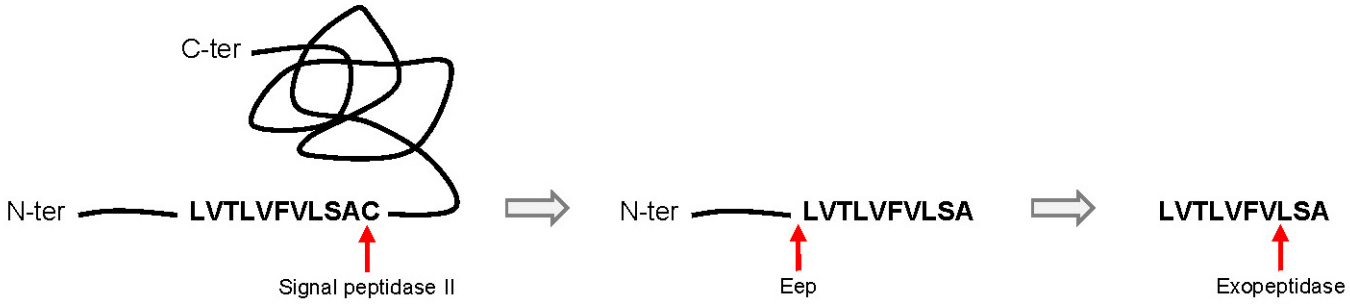


Figure 22 – Biosynthèse de la sex-phéromone cCF10 à partir du produit du gène *ccfA* chez *Enterococcus faecalis* (d'après Antiporta and Dunny, 2002 ; Chandler *et al.*, 2008)**.**

Tableau 7 – Précurseurs des sex-phéromones (adapté de Clewell, 2000).

	Précurseur	Sex- phéromone	Séquence[a]	Taille[b]
pCF10	pCF10p	cCF10	VKKYKRLLLMAGLVTLVFV**L**SA**C**GTAPVS…	275 aa
pAD1	pAD1p	cAD1	MKVNKFVKGFAAIALFSLV**L**AG**C**GADKK…	143 aa
pPD1	pPD1p	cPD1	MRKLNRWLYGSGLLFLVMF**L**SG**C**VKTGA…	234 aa

[a] La forme mature des sex-phéromones est soulignée.
[b] La taille indiquée en acides aminés (aa) représente la taille globale de la séquence du précurseur, seulement la partie contenant la forme mature des sex-phéromones étant représentée.

Tableau 8 – Précurseurs des anti-phéromones (adapté de Clewell, 2000)**.**

	Précurseur	Anti-phéromone	Séquence
pCF10	pCF10p	iCF10	MKTTLKKLSRYIAVVIAITLIFI
pAD1	pAD1p	iAD1	MSKRAMKKIIPLITLFVVTLVG
pPD1	pPD1p	iPD1	MKQQKKHIAALLFALILTLVS

bloque l'expression des gènes de conjugaison contrôlés par PrgX et transcrits à partir du promoteur P_Q, mais ne diminue que partiellement l'expression de *prgQ* et par conséquent la production des anti-phéromones iCF10 (Chandler and Dunny, 2004; Kozlowicz *et al.*, 2006a).

2.2.4.3 Maturation des peptides phéromones

- Les sex-phéromones :

La voie de biosynthèse des sex-phéromones se distingue de celle des autres molécules de signalisation. En effet, les formes matures des sex-phéromones ont la particularité d'être des fragments de séquences signales de précurseurs de lipoprotéines (Clewell *et al.*, 2000). Aussi, pour acquérir leur forme mature, les sex-phéromones subissent trois clivages : un premier en C-terminal au niveau d'un motif lipobox (chapitre 2 § 2.2.1.2), un deuxième en N-terminal, réalisé dans la majorité des cas par le déterminant chromosomique Eep et, un troisième clivage par une exopeptidase non identifiée au niveau de trois résidus localisés en C-terminal (Figure 22) (Thoendel and Horswill, 2010). En effet, il a été montré qu'un mutant du gène *eep* entraine une diminution de la production de nombreuses sex-phéromones (An *et al.*, 1999). Les caractéristiques structurelles de la protéine Eep suggèrent qu'elle joue un rôle de protéase membranaire. En outre, le rôle des lipoprotéines à partir desquelles les sex-phéromones sont produites, telle que CcFA, n'est pas encore élucidé et ne semble pas être en lien avec la conjugaison (Chandler and Dunny, 2004).

A l'exemple des plasmides conjugatifs les plus connus, les sex-phéromones correspondent à des hepta ou octapeptides et ont la particularité d'être très hydrophobes (Tableau 7). L'extrémité N-terminale des sex-phéromones a été décrite comme essentielle à leur activité. En effet, des peptides analogues à cCF10 modifiés en N-terminal (VATLVFV et LATLVFV) ne présentent pas d'activités biologiques lorsqu'ils sont présents à des concentrations physiologiques (Dunny *et al.*, 2001). Des expériences similaires ont été décrites pour les sex-phéromones cPD1 et cAD1 (Dunny *et al.*, 2001). De plus, des études portant sur la construction de sex-phéromones hybrides associant l'extrémité N-terminale d'une phéromone à l'extrémité C-terminale d'une autre phéromone, ont montré que seulement le plasmide conjugatif apparenté aux résidus côté N-terminal était induit (Dunny *et al.*, 2001). Aussi, la partie N-terminale des sex-phéromones est décrite comme intervenant dans l'interaction peptide / régulateur alors que leur partie C-terminale jouerait probablement un rôle dans la stabilisation de cette interaction (Dunny *et al.*, 2001).

- Les anti-phéromones :

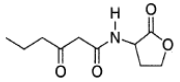
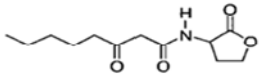
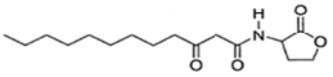
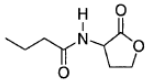
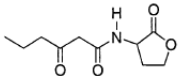
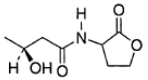
Les anti-phéromones sont issues de peptides codés par des petits gènes, présentant des caractéristiques de séquences signales, et sont maturées suite à un clivage côté C-terminal par la protéase membranaire Eep (Thoendel and Horswill, 2010). La sécrétion et la maturation des anti-phéromones se produiraient de manière simultanée au cours de leur export (Chandler and Dunny, 2004). Néanmoins, ce mécanisme n'est pas connu à ce jour.

Les anti-phéromones correspondent également à des hepta ou octapeptides hydrophobes, présentant de fortes homologies entre elles (Tableau 8), avec les sex-phéromones et en particulier avec celle du plasmide apparenté (Dunny *et al.*, 2001). Aussi, les anti-phéromones étant des antagonistes des sex-phéromones, les résidus composant leur séquence semblent jouer un rôle crucial à leur activité biologique (Chandler and Dunny, 2008).

2.2.4.4 Spécificité des plasmides conjugatifs

Comme vu dans la partie précédente, les sex- et anti-phéromones présentent de fortes similarités au niveau de leur séquence peptidique. De plus, les différents gènes codant les protéines et les régulateurs nécessaires au transfert des différents plasmides conjugatifs caractérisés présentent également de fortes homologies. Aussi, l'absence de cross-activation entre ces différents éléments conjugatifs souligne une grande spécificité de réponse de chaque système avec ses partenaires apparentés (Dunny *et al.*, 2001). En effet, il a été montré qu'une cellule donneuse, contenant les plasmides conjugatifs pCF10 et pPD1, répond de manière identique à la présence de cCF10 et ceci bien qu'elle soit présentée seule ou en mélange avec la sex-phéromone cPD1. Des résultats identiques ont également été démontrés entre les plasmides pAD1 et pPD1 (Dunny *et al.*, 2001). Aussi, malgré leur similarité, les cellules induisent spécifiquement les plasmides conjugatifs apparentés et toute possibilité de cross-activation semble être abolie. Cette spécificité semble être effective (i) dans le milieu extracellulaire au niveau de la reconnaissance et de la détection de la sex-phéromone par sa protéine affine et, (ii) de manière intracellulaire avec le régulateur transcriptionnel. De manière complémentaire, les mécanismes de contrôle de l'auto-induction des cellules donneuses par les phéromones endogènes seraient également très spécifiquement contrôlés (Dunny *et al.*, 2001).

Tableau 9 – Caractéristiques des principales molécules de signalisation appartenant à la famille des *N*-AHL chez les bactéries à Gram négatif.

	Nom	Structure	Synthèse	Senseur	Fonction régulée
Vibrio fischeri	*N*-(3-oxohexanoyl)-HSL		LuxI	LuxR	Bioluminescence
Agrobacterium tumefaciens	*N*-(3-oxooctanoyl)-HSL		TraI	TraR	Transfert conjugatif
Pseudomonas aeruginosa	*N*-(3-oxododecanoyl)-HSL		LasI	LasR	Virulence Formation de biofilm
	N-butyryl-HSL		RhlI	RhlR	
Erwinia stewartii	*N*-(3-oxohexanoyl)-HSL		EsaI	EsaR	Virulence et biosynthèse de la capsule polysaccharidique
Vibrio harveyi	*N*-(3-hydroxybutanoyl)-HSL		LuxM	LuxN	Bioluminescence

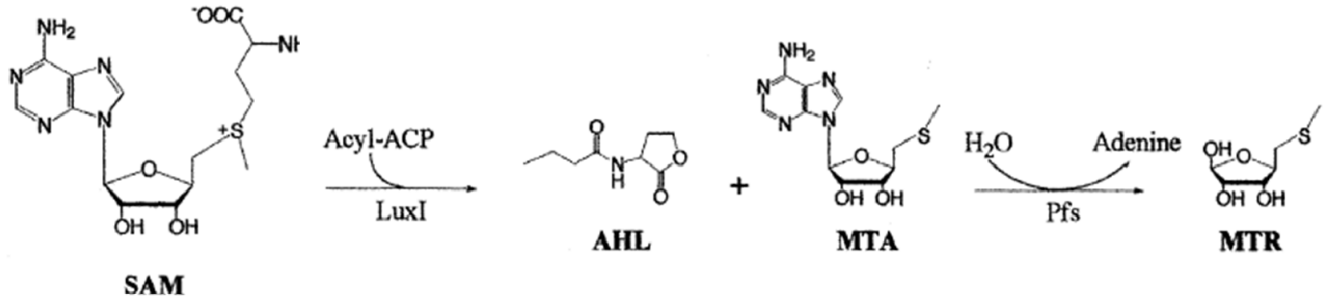


Figure 23 – Exemple d'une voie de biosynthèse des *N*-acyl homoserines lactones par les enzymes LuxI (Bassler, 2002)**.**
SAM, *S*-adenosylmethionine; AHL, acyl homoserines lactones; MTA, méthylthioadenosine; MTR, méthylthioribose.

3. Le QS chez les bactéries à Gram négatif

Les *N*-acyl homosérine lactones (*N*-AHL) – ou autoinducteur 1 – représentent la classe majeure d'autoinducteurs utilisés dans les mécanismes de QS chez les bactéries à Gram négatif. Une variété considérable de *N*-AHL est produite par les bactéries ; et bien que certaines bactéries ne soient capables de produire qu'un seul type de *N*-AHL, d'autres possèdent en revanche la capacité d'en produire plusieurs types différents (Tableau 9). Bien que la structure de base de ces molécules de signalisation soit conservée, les *N*-AHL diffèrent par la taille de leur chaine latérale *N*-liée (comprise entre 4 et 16 carbones), par la nature de la substitution à la position carbone C3 et par la présence d'un ou plusieurs ponts insaturés au niveau de leurs chaines latérales (Fuqua and Greenberg, 2002; Ng and Bassler, 2009).

3.1 Biosynthèse des *N*-AHL : les enzymes de la famille LuxI et LuxM

Les *N*-AHL peuvent être issues de deux voies de biosynthèse et les *N*-acyl synthétases catalysent leur production. Parmi celles-ci, les protéines de la famille LuxI sont le plus couramment utilisées.

Les *N*-AHL dérivent des *S*-adenosylmethionines (SAM) par acylation et lactonisation de ces dernières. Plus précisément, l'enzyme LuxI couple le groupement acyl d'un acyl-ACP (**A**cyl **c**arrier **p**rotein) sur le groupement méthionine de la SAM. La lactonisation de l'intermédiaire formé résulte ensuite en la production d'une *N*-AHL et de méthylthioadenosine (MTA), molécule toxique pour les cellules. Aussi, une nucléotidase, appelée Pfs, assurera par la suite le clivage de la MTA en deux molécules non toxiques : de l'adénine et du méthylthioribose (MTR) dont le devenir n'est pas connu à ce jour (Figure 23) (Ng and Bassler, 2009). De manière simplifiée, la partie « acyl » dérive d'un acide gras précurseur conjugué à une acyl-ACP alors que la partie « homosérine lactone » provient quant à elle d'une SAM. Ce métabolite étant un précurseur commun reconnu par l'ensemble des protéines de la famille LuxI, la diversité des *N*-AHL synthétisées est alors assurée par la reconnaissance d'un acyl-ACP spécifique apportant un groupe acyl caractéristique qui sera transféré lors de l'étape d'acylation de la SAM. Ce mécanisme biochimique est démontré pour plusieurs protéines LuxI-like provenant de différentes bactéries telles que LuxI de *V. fischeri*, TraI de *Agrobacterium tumefaciens* et RhlI de *Pseudomonas aeruginosa* (Fuqua and Greenberg, 2002).

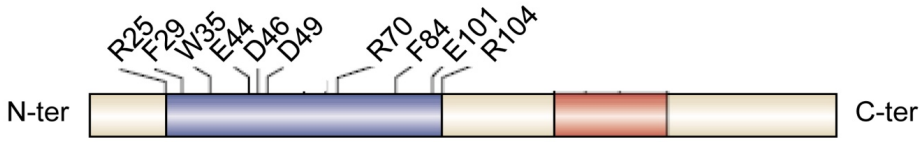


Figure 24 – Représentation schématique des *N*-acyl synthétases de la famille LuxI (adapté de Fuqua and Greenberg, 2002)**.**
Les résidus conservés au sein de toutes les protéines LuxI sont indiquées sur la figure. En bleu est représentée le domaine catalytique, en rouge le domaine de spécificité de liaison au substrat. Les positions indiquées correspondent à celles identifiées sur la séquence de la protéine LuxI.

Enfin, bien qu'elles catalysent les mêmes réactions, certaines bactéries à Gram négatif possèdent des *N*-AHL synthétases ne présentant pas d'homologie avec LuxI. Parmi celles-ci, LuxM de *V. harveyi*, AinS de *V. fischeri* et HdtS de *Pseudomonas fluorescens* ont été identifiées. L'étude de la voie de biosynthèse des *N*-AHL par l'enzyme AinS a alors montré que ces *N*-AHL synthétases utilisent également la SAM et l'acyl-CoA ou acyl-ACP comme substrat à la production des *N*-AHL (Hanzelka *et al.*, 1999).

3.2 Caractéristiques des protéines de la famille LuxI

De nombreuses protéines LuxI sont identifiées à ce jour. Ces protéines présentent une taille comprise de 190 à 230 acides aminés et partagent 30 à 35% d'identité (Fuqua and Greenberg, 2002). Alors que la partie C-terminale de ces protéines semble être impliquée dans la spécificité de liaison au substrat, leur partie N-terminale représente leur domaine catalytique. L'alignement de 16 séquences de protéines LuxI-like a révélé la présence de 10 acides aminés conservés localisés au niveau du site actif (Figure 24). Des études de mutagénèse dirigée sur la protéine RhlI chez *P. aeruginosa,* ont par ailleurs indiqué que 7 des 10 acides aminés conservés étaient essentiels à la catalyse de la réaction, les 3 restants affectant seulement son efficacité (Fuqua and Greenberg, 2002).

Malgré l'absence de corrélation entre le niveau d'identité des protéines LuxI et la nature des *N*-AHL produites, les protéines LuxI catalysant la synthèse d'une même *N*-AHL peuvent néanmoins présenter des similarités. Aussi, un résidu thréonine conservé en position 143 et localisé dans le domaine de liaison au substrat, a été identifié chez toutes les protéines LuxI-like synthétisant le 3-oxo-acyl-HSL (Fuqua and Greenberg, 2002). Des études cristallographiques ont par ailleurs permis de définir la structure de deux *N*-acyl synthétases complexées à leur substrat : EsaI de *Pantoea stewarii* et LasI de *P. aeruginosa* (Fuqua and Greenberg, 2002). Deux configurations différentes ont alors été observées au niveau des poches de liaison du substrat, l'une présentant une cavité fermée (EsaI), l'autre un tunnel allongé (LasI). Ces caractéristiques structurelles semblent alors conférer aux protéines LuxI une spécificité liée à l'encombrement stérique des substrats qu'elles peuvent lier ; EsaI ne pouvant accueillir des substrats présentant une chaine latérale *N*-liée relativement courte.

3.3 Sécrétion des *N*-AHL

A la suite de leur synthèse, les *N*-AHL diffusent librement à travers la membrane, c'est-à-dire vers l'intérieur et vers l'extérieur de la cellule. Néanmoins, certains efflux transporteurs

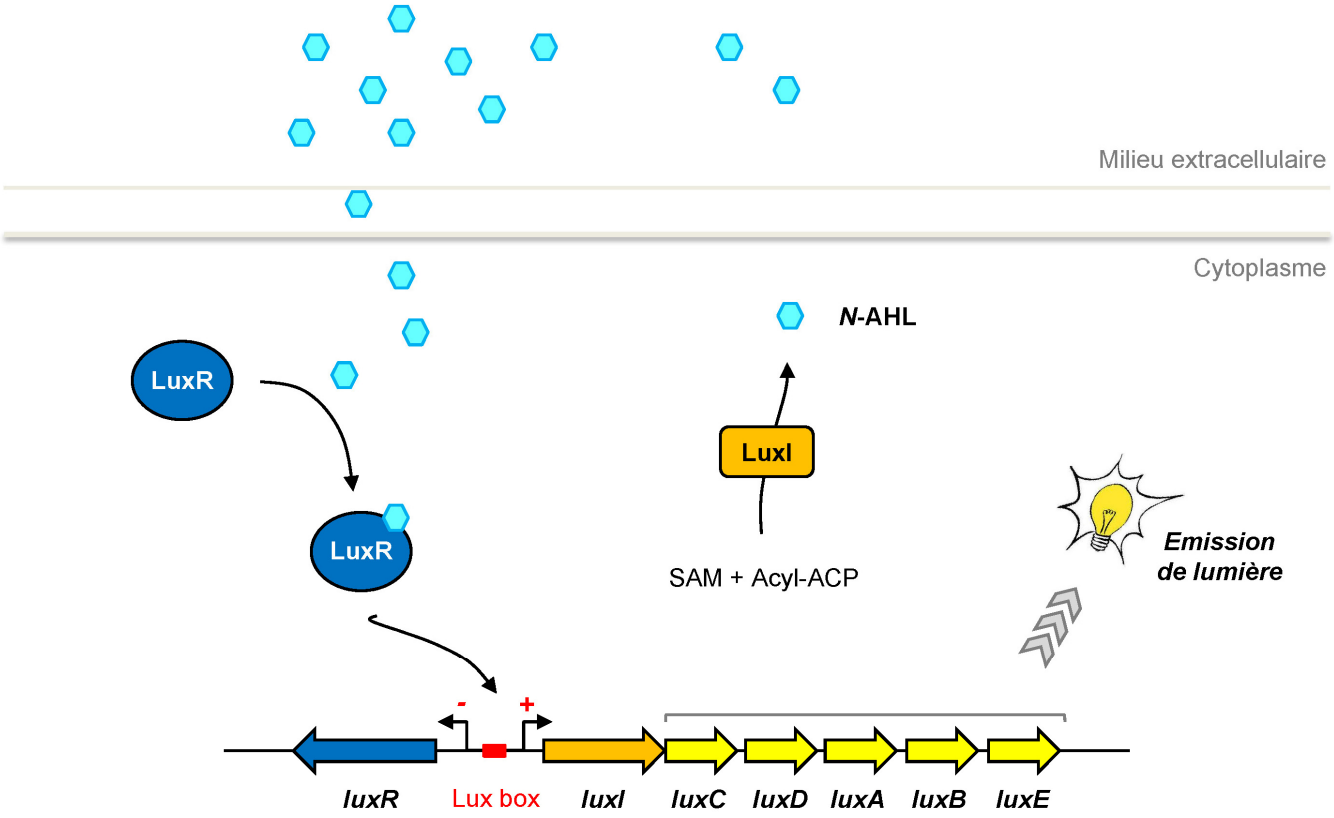


Figure 25 – Représentation schématique du système *luxIR* contrôlant la bioluminescence chez *Vibrio fischeri* (inspiré de Miller and Bassler, 2001)**.**
La protéine LuxI est responsable de la synthèse de la *N*-AHL capable de diffuser librement à travers la membrane plasmique. Une fois sa concentration seuil critique atteinte, la *N*-AHL est détecté par la protéine régulatrice LuxR. Aussi, la fixation du complexe LuxR / *N*-AHL à l'ADN sur un motif appelé « Lux box » activera la transcription de l'opéron *luxICDABE*, responsable de la production de lumière et réprimera la transcription du gène *luxR*. Ce contrôle négatif compense la boucle positive de régulation mis en place au niveau du promoteur de l'opéron *luxICDABE*.

pourraient être impliqués dans le transport de certaines *N*-AHL à travers la membrane et en particulier, lorsque ces dernières présentent une longue chaine latérale *N*-liée pouvant freiner leur diffusion à travers la bicouche lipidique (Fuqua and Greenberg, 2002).

3.4 Détection des *N*-AHL

3.4.1 Les systèmes *luxIR* :

Mécanisme de QS largement représenté au sein des bactéries à Gram négatif (70 analogues identifiés), les systèmes *luxIR* ont été caractérisés en détails chez plusieurs bactéries telles que *V. fischeri*, *P. aeruginosa*, *A. tumefaciens* ou encore *P. stewartii*. Ces systèmes assurent la régulation de nombreuses fonctions telles que la bioluminescence, le transfert conjugatif, la formation de biofilm ou encore la production de facteurs de virulence pour les pathogènes des plantes et des animaux (Miller and Bassler, 2001). Bien que leur rôle physiologique soit différent, tous ces systèmes fonctionnent néanmoins de la même manière et la molécule de signalisation générée par une *N*-AHL synthétase de la famille LuxI est perçue par un *N*-AHL récepteur appartenant à la famille LuxR (régulateur transcriptionnel) (Figure 25). En effet, suite à leur production, les *N*-AHL diffusent librement à travers la membrane. Cette caractéristique assure un équilibre entre leur concentration extracellulaire et intracellulaire. Aussi, dès lors que la molécule de signalisation atteint un seuil de concentration critique, proportionnel à la densité cellulaire de la population, elle sera liée par son récepteur localisé dans le cytoplasme. La liaison de l'autoinducteur à son régulateur contrôle son activité transcriptionnelle et résulte *in fine* en la régulation de gènes cibles *via* la fixation de LuxR dans la région promotrice de ces derniers. Par ailleurs, les gènes cibles incluent généralement le gène *luxI*, assurant ainsi une boucle de régulation positive de la synthèse des *N*-AHL et le passage de l'ensemble de la population bactérienne concernée à la production d'une fonction physiologique donnée (Miller and Bassler, 2001).

A l'heure actuelle, le système *luxIR* de *V. fischeri* est considéré comme LE modèle pour l'étude des systèmes de QS chez les bactéries à Gram négatif. Aussi, une fois lié à sa molécule de signalisation, LuxR induira la production de lumière *via* l'activation de l'opéron luciférase *luxICDABE* (Figure 25) (Ng and Bassler, 2009).
A l'instar de *V. fischeri*, certaines bactéries à Gram négatif ont la particularité d'accumuler plusieurs systèmes homologues de type LuxIR dans leur génome. Tel est le cas chez *P. aeruginosa* où deux systèmes appelés LasIR et RhlIR ont été caractérisés et contrôlent la production de plusieurs facteurs de virulence et la formation de biofilm (Tableau 9) (Waters

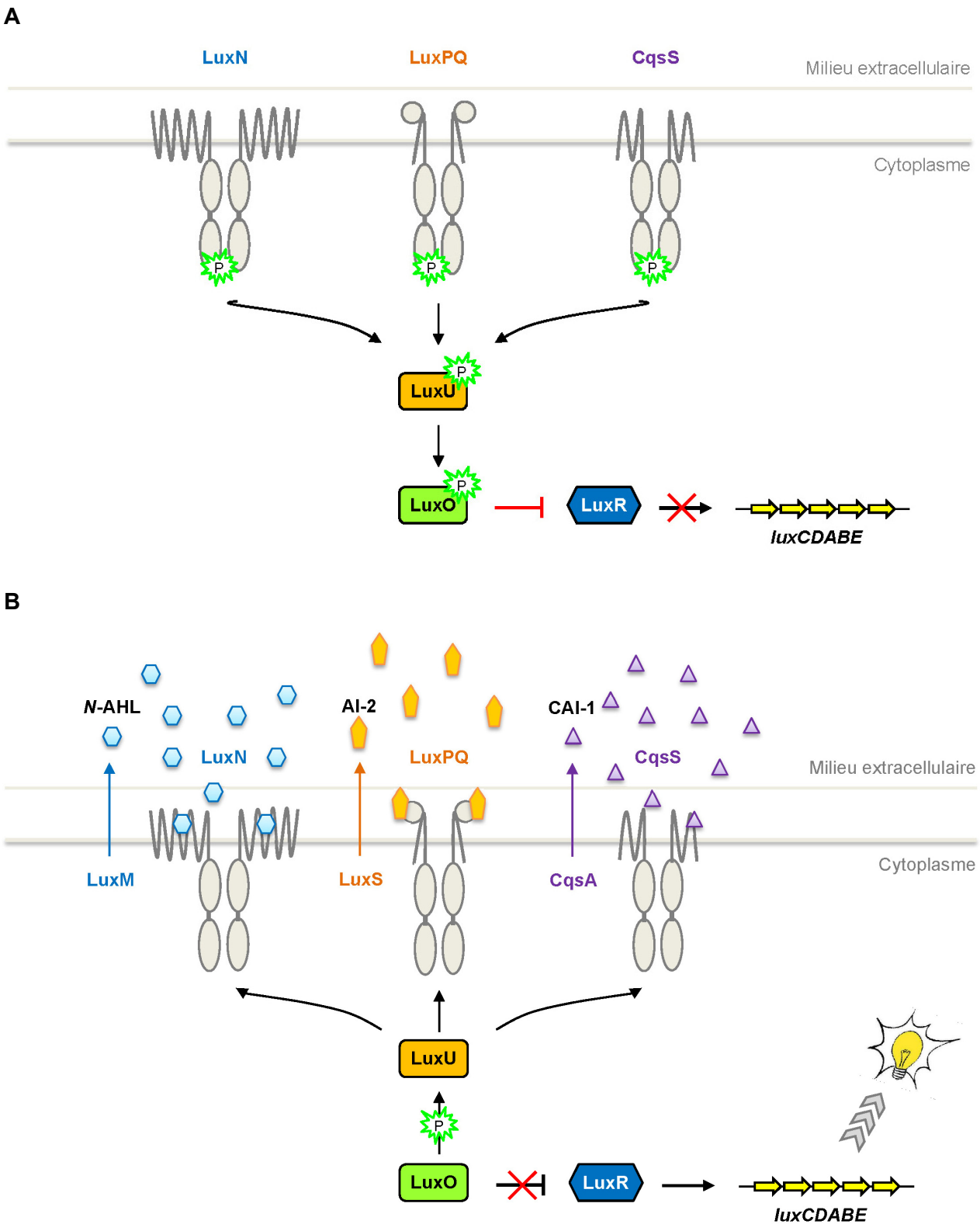


Figure 26 – Représentation schématique des systèmes de quorum-sensing contrôlant la bioluminescence chez *Vibrio harveyi* (adapté de Ng and Bassler, 2009)**.**
(A) Faible densité cellulaire. La phosphorylation de LuxO suite au transfert du groupe phosphoryle de LuxU apporté par les histidines kinases LuxN, LuxPQ et CqsS traduit la répression de LuxR et l'absence de transcription de l'opéron *luxCDABE*. **(B)** Haute densité cellulaire. La détection des molécules de signalisation par les senseurs histidine kinase (LuxN, LuxPQ et CqsS), provoque la déphosphorylation de LuxU et par conséquent, le retour du groupe phosphoryle de LuxO. Ceci entraîne une levée de la répression sur LuxR et ainsi la transcription de l'opéron *luxCDABE* et la production de lumière.

and Bassler, 2005). Ces deux systèmes fonctionnent en tandem ; le système LasIR étant nécessaire à l'activation du second.

3.4.2 Les systèmes *luxMN* :

V. harveyi, au même titre que *V. fischeri*, possède un système de QS lui permettant de produire de la lumière et médié par une *N*-AHL. Néanmoins, la synthèse de cette dernière est dirigée par une *N*-AHL synthétase appartenant à la famille LuxM. Aussi, bien que cette molécule de signalisation diffuse librement à travers la membrane, sa perception implique l'intervention d'un récepteur de type kinase localisé dans la membrane et appelé LuxN (Figure 26). LuxN appartient à la classe des systèmes à deux composants hybrides et est composé d'un domaine senseur et d'un domaine régulateur de réponse.
A faible densité cellulaire, l'autophosphorylation de LuxN déclenche une cascade de phosphorylation assurant *in fine* le transfert d'un phosphate à un régulateur de réponse appelé LuxO. Ce dernier est alors responsable de la répression des gènes de l'opéron *luxCDABE*. En revanche, à forte densité cellulaire, l'interaction de la molécule de signalisation avec LuxN entraine un reflux du phosphate aboutissant à la déphosphorylation et à l'inactivation de LuxO. Aussi, un activateur transcriptionnel appelé LuxR (non similaire aux protéines des systèmes *luxIR*) induit la transcription des gènes *luxCDABE* et par conséquent la production de lumière par la bactérie (Miller and Bassler, 2001; Waters and Bassler, 2005).

Un système homologue à LuxMN est également retrouvé chez *V. fischeri*. Ce dernier, appelé AinS/AinR, fonctionnerait en tandem avec le système LuxIR précédemment décrit. De la même manière que pour le système de *V. harveyi*, l'interaction de AinR avec sa molécule de signalisation inactive LuxO permettant ainsi à un activateur, cette fois ci LitR, d'induire l'expression de plusieurs gènes cibles tels que *ropS* ou des gènes en lien avec la mobilité de la bactérie et également d'assurer le déclenchement du deuxième système de QS régulant le gène *luxI* (Studer *et al.*, 2008).

4. La communication inter-espèces

L'autoinducteur 2 (AI-2) ou furanosyl borate diester est la troisième grande famille de molécules de signalisation répandue dans le monde bactérien. De par le caractère généralisé de l'AI-2 et de l'enzyme assurant sa synthèse (LuxS) chez les bactéries à Gram positif et à Gram négatif, l'AI-2 a été proposé comme molécule de signalisation inter-espèces

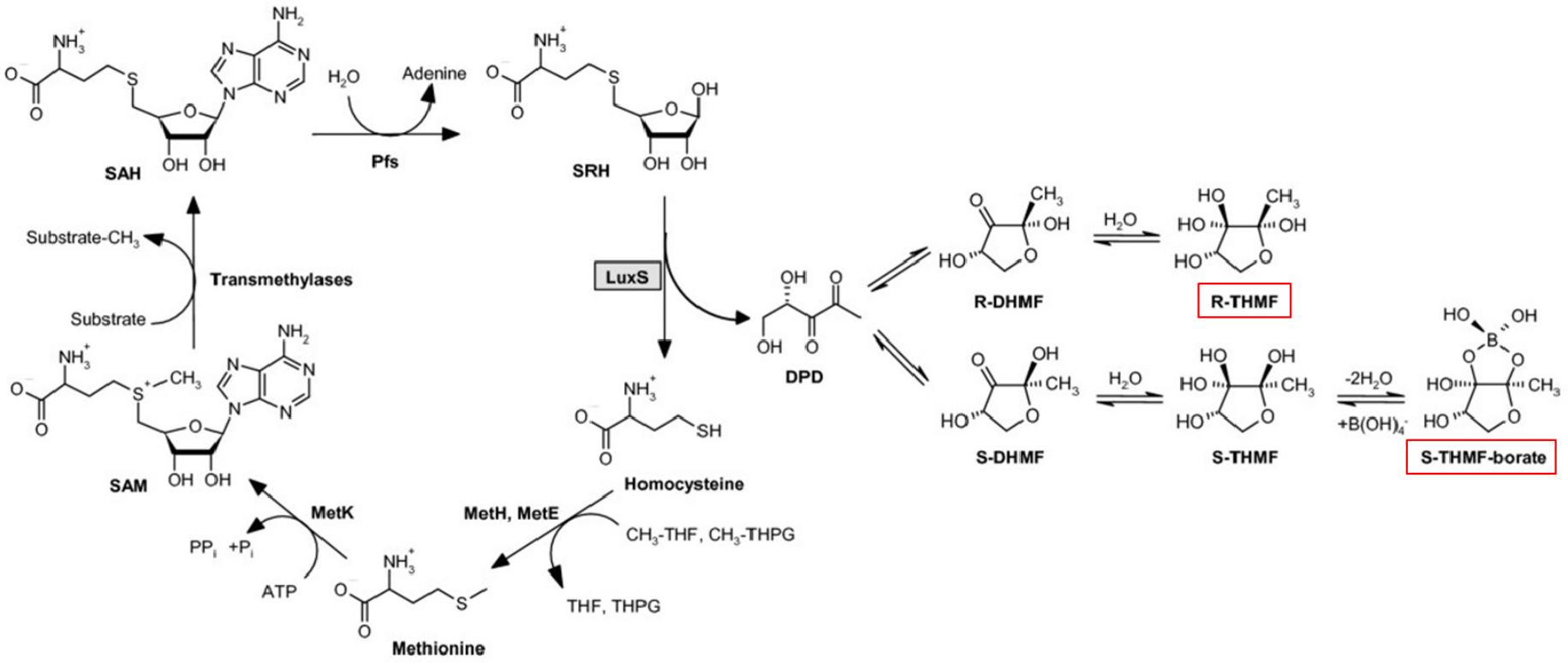


Figure 27 – Voie de biosynthèse de l'autoinducteur 2 (Rezzonico and Duffy, 2008)**.**
SAM, *S*-adenosylmethionine ; SAH, *S*-adenosylhomocysteine ; SRH, *S*-ribosylhomocysteine ; DPD, 4,5-dihydroxy-2,3 pentanedione. En fonction de la présence ou non de borate, l'AI-2 correspond soit à du (2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran-borate (*S*-THMF-borate), soit à du (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran (*R*-THMF).

chez les bactéries (Schauder and Bassler, 2001). A l'heure actuelle, l'AI-2 est la seule molécule de signalisation commune entre les bactéries à Gram positif et à Gram négatif.

De manière similaire aux *N*-AHL, l'AI-2 est produit à partir des SAM et dérive d'un précurseur commun appelé le 4,5-dihydroxy-2,3 pentanedione (DPD), produit par l'enzyme LuxS. Plus précisément, le DPD est produit à partir de la SAM suite à l'action de 3 enzymes (Figure 27). Une méthyltransférase catalyse le transfert du groupe méthyle de la SAM induisant la formation de SAH, toxique pour les cellules. Aussi, dans un second temps, l'action de la nucléotidase Pfs permettra le clivage de ce composé en adénine et SRH (*S*-ribosylhomocystéine). Enfin, l'hydrolyse de la SRH par LuxS assurera la formation de DPD par le relargage d'une molécule d'homocystéine, précurseur de méthionine pour la cellule et assurant *in fine* la régénération de la SAM (Figure 27) (Ryan and Dow, 2008). Le DPD ainsi produit (molécule instable) serait alors converti de manière spontanée en différents composés (Schauder and Bassler, 2001). Il subit une cyclisation et différents réarrangements pour acquérir la forme active de l'AI-2 d'une espèce bactérienne en question (Waters and Bassler, 2005). Aussi, en présence de borate, comme dans les environnements marins, le DPD formera du *S*-THMF[2] correspondant aux molécules d'AI-2 produites par *V. harveyi* et *Vibrio cholerae*. *A contrario*, sur l'environnement terrestre où le bore est limité, le DPD formera du *R*-THMF[3] correspondant aux molécules d'AI-2 produites par *E. coli* ou *Salmonella*. Aussi, de par l'instabilité du DPD et son interaction possible avec une large variété de composés pouvant interagir avec du borate ou d'autres espèces réactives, le DPD pourrait induire la formation de molécules de signalisation de structures différentes pouvant être à la base (au même titre que les *N*-AHL ou que les peptides autoinducteurs) d'une réelle famille de molécules de signalisation de type AI-2 (Waters and Bassler, 2005). Néanmoins, le rôle de l'AI-2 dans la communication inter-espèce est contreversé depuis l'appartion des travaux chez *E. coli* et *Salmonella* qui présentent, cette molécule, comme un métabolite intervenant la détoxification de la SAH, plutôt qu'une molécules de signalisation à proprement parler (Winzer *et al.*, 2002 ; Walters et Sperandio, 2006).

Bien que de nombreuses bactéries possèdent un gène homologue à *luxS*, permettant la synthèse de l'AI-2, seulement deux systèmes sont caractérisés en détails et concernent *V. harveyi* (Figure 26) et *V. cholerae*. Chez ces deux espèces, la détection de l'AI-2 requiert une protéine périplasmique appelée LuxP. Le complexe ainsi formé interagit ensuite avec LuxQ, protéine kinase nécessaire à la transduction du signal. L'information est alors

[2] (2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran-borate (*S*-THMF)

[3] (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran (*R*-THMF)

transférée, de la même manière que le système LuxMN, à une phosphotransférase (LuxU) qui transmet *in fine* le signal au régulateur de réponse LuxO (Miller and Bassler, 2001; Waters and Bassler, 2005). De manière intéressante, ce système contrôle également l'opéron *luxCDABE* nécessaire à la production de la lumière chez *V. harveyi*. En revanche, chez *V. cholerae*, il contrôle positivement la production d'une protéase (Hap) et négativement des gènes impliqués dans la formation de biofilm et la production de facteur de virulence (Ng and Bassler, 2009).

Enfin, bien qu'il ne soit pas détaillé, *V. harveyi* possède également un troisième système de QS contrôlant la bioluminescence chez cette bactérie. Ce système, fonctionnant de manière identique aux systèmes LuxMN et LuxSPQ, fait intervenir une molécule de signalisation CAI-1, de nature inconnue, qui une fois détectée par son récepteur membranaire CqsS, induira la réponse physiologique de la cellule (Figure 26) (Ng *et al.*, 2011).

5. Autres molécules de signalisation connues chez les bactéries

Bien que les principales molécules de signalisation aient été décrites dans cette partie, d'autres molécules enrichissent cette classification (Ryan and Dow, 2008).

A titre d'exemple, chez les Streptomycètes (bactéries à Gram positif), les molécules de signalisation appartiennent à la famille des γ-butyrolactones et contrôlent la différenciation (formation des hyphes) et la production d'antibiotiques, tel que la streptomycine. Les γ-butyrolactones sont des analogues structuraux aux *N*-AHL, cependant, aucune cross-activation entre ces molécules de signalisation n'a encore été démontrée (Bassler and Losick, 2006).

La molécule d'AI-3, composé aromatique de structure inconnue (Pacheco and Sperandio, 2009), est produite par la flore intestinale des individus sains, par les bactéries commensales *E. coli*, *Klebsiella pneumoniae* et *Enterobacter cloacae* et par de nombreux pathogènes *E. coli* O157 :H7, *Shigella* spp. et *Salmonella* spp. (Walters and Sperandio, 2006). Chez *E. coli* O157:H7, il a été montré que l'AI-3 active l'expression de la transcription des gènes de l'îlot de pathogénicité LEE (Locus of enterocyte effacement) et du régulon flagella (Hughes and Sperandio, 2008). Par ailleurs, les protéines Qse (senseurs de l'AI-3) et les composants de la cascade de signalisation sont présents dans de nombreuses bactéries suggérant que cette molécule de signalisation peut être impliquée dans des mécanismes de signalisation inter-espèces parmi les bactéries intestinales. Il a par ailleurs été montré que

deux hormones eucaryotes (épinéphrine et norépinéphrine) se fixent également sur le senseur Qse et que les antagonistes de ces hormones bloquent la voie de transduction du signal et par conséquent l'expression des gènes cibles. Aussi, l'AI-3 a été proposée comme molécule de signalisation inter-règne (Pacheco and Sperandio, 2009).

L'indole et les dérivés d'indole, tels que les quinolones, ont également été décrits comme intervenants dans la communication cellulaire de type QS (Lee and Lee, 2010). Chez *P. aeruginosa*, un troisième système de QS a été décrit utilisant la molécule de signalisation PQS pour **P**seudomonas **q**uinolone **s**ignal (Venturi, 2006).

Enfin, d'autres molécules telles que les di-peptides cycliques chez les bactéries à Gram négatif, les dérivés d'acides gras et les antibiotiques sont également décrits comme intervenant dans les mécanismes de communication cellulaire de type QS (Ryan and Dow, 2008).

6. Conclusion

La possibilité pour les bactéries de communiquer entre elles *via* des mécanismes de QS leur permet de coordonner leur comportement à l'échelle d'une population. Dans ce chapitre, nous avons pu voir que cette capacité de communiquer au sein d'une espèce ou entre espèces est capitale pour la survie bactérienne et / ou leurs interactions dans leurs habitats naturels.

6.1 La complexité des réseaux de communication cellulaire

Les avancées dans le domaine montrent que la plupart des bactéries communiquent probablement en utilisant des molécules de signalisation pour coordonner leur comportement à l'échelle d'un groupe. Les continuelles apparitions et mises en évidence de nouvelles molécules de signalisation augmentent la complexité des systèmes de QS décrits dans la littérature. A l'heure actuelle, les *N*-AHL et les AIP représentent les deux principales classes de molécules de signalisation (caractérisées respectivement chez les bactéries à Gram négatif et à Gram positif) et la communication cellulaire *via* ces autoinducteurs suscite des réponses spécifiques et limitées dans la plupart des cas à une espèce bactérienne. A l'inverse, le langage universel décrit sur la base de l'AI-2 peut être détecté indifféremment par une bactérie et ce, que l'AI-2 produit soit de nature endogène ou exogène. De manière originale, il a été montré que certaines bactéries, telle que *P. aeruginosa*, ne possédant pas d'enzyme LuxS et par conséquent ne pouvant produire de l'AI-2, étaient néanmoins aptes à

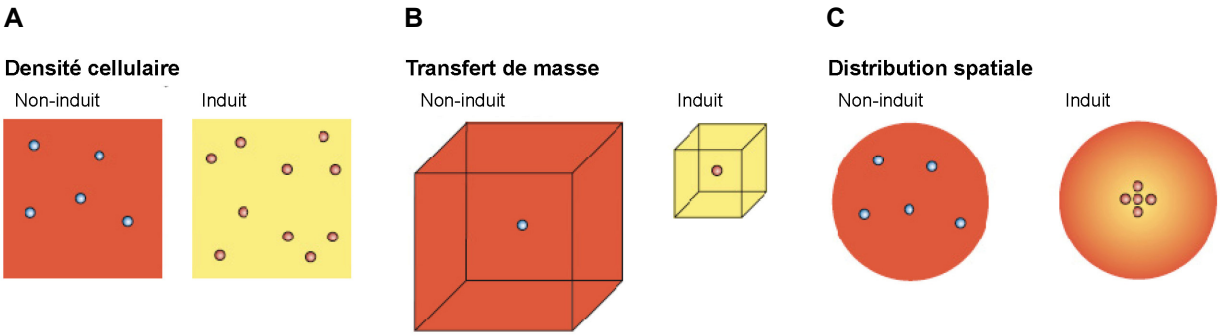


Figure 28 – Paramètres influençant le déclenchement d'une réponse physiologique coordonnée, de type quorum-sensing, chez les bactéries (d'après Hense *et al.*, 2007)**.**
La densité cellulaire critique **(A)**, un espace confiné **(B)** et une distribution spatiale agrégée des cellules **(C)** conduit à une réponse physiologique des cellules concernées. Un fond rouge ou jaune indique, respectivement, une concentration faible ou élevée en autoinducteur. Les bactéries non induites sont de couleur bleue et les bactéries induites sont en rouge.

y répondre (Duan *et al.*, 2003; Rickard *et al.*, 2006). Aussi, les bactéries semblent être capables de percevoir, d'intégrer et de répondre de manière appropriée à un pool de signaux présents dans leur environnement. Ceci est d'autant plus vrai que le langage universel décrit sur la base de l'AI-2 est généralement identifié en parallèle d'un système de QS modulé par les *N*-AHL ou les AIP et, que plusieurs systèmes de QS modulés par des *N*-AHL ou des AIP peuvent s'accumuler au sein d'une bactérie.

6.2 Le QS et les notions de diffusion et efficiency-sensing

Comme vu dans ce chapitre, les bactéries possèdent divers mécanismes leur permettant de coordonner leur physiologie à l'échelle d'une population *via* la production de molécules de signalisation. Aussi, alors que le terme de « quorum-sensing » réfère à la quantité de bactéries présentes dans leur environnement, ce dernier ne décrit pas de manière adéquate l'ensemble des situations où les bactéries emploient des molécules de signalisation diffusibles. Aussi, à l'heure actuelle, cette notion est discutée et de nouvelles expressions, telles que « diffusion-sensing » (Redfield, 2002) ou encore « efficiency-sensing » (Hense *et al.*, 2007) voient le jour. Alors que la première expression s'illustre par la notion de transfert de masse (Figure 28), la seconde attribue le « sensing » d'une molécule de signalisation à l'association d'au minimum deux composantes parmi la densité cellulaire, le transfert de masse et la distribution spatiale de la population (Figure 28). Aussi, bien que la résultante de tous ces phénomènes soit un comportement coordonné de la population bactérienne, différents paramètres semblent être nécessaires à prendre en compte dans la définition ou l'utilisation de la notion de « quorum-sensing ». En 2010, les travaux de Yang *et al.* ont montré l'impact de la viscosité du milieu sur la diffusion de la molécule de signalisation CSP chez *S. pneumoniae* (Yang *et al.*, 2010).

Tableau 10 – Les principaux régulateurs transcriptionnels Rgg caractérisés et leur gène cible proximal associé chez les bactéries à Gram positif.

	Régulateur	Gènes régulés*		Fonctions contrôlées
Streptococcus gordonii	Rgg	*gtfG*	+	Glucosyltransferase
Streptococcus oralis	Rgg	*gtfF*	+	Glucosyltransferase
Streptococcus pyogenes	RopB	*speB*	+	Synthèse de l'exotoxine B
Streptococcus agalactiae	RovS	*fbsA*	-	Adhésion aux cellules épithéliales
Streptococcus thermophilus	RggC	*(orf2)*	?	Réponse au stress oxydatif
Lactococcus lactis	GadR	*gadCB*	+	Antiport glutamate/GABA
Streptococcus mutans	MutR	*mutA-G*	+	Biosynthèse des mutacines I, II et III
Lactobacillus sakei	LasX	*lasA-W*	+	Biosynthèse de la lactocine S

* Les signes « + » et « - » indiquent respectivement que l'expression du gène est positivement ou négativement contrôlée par la protéine régulatrice Rgg.

Chapitre 3 : Les régulateurs transcriptionnels de la famille des Rgg

1. Présentation des protéines régulatrices Rgg

Les protéines Rgg (**R**egulator **g**ene of **g**lucosyltransferase) constituent une famille de régulateurs transcriptionnels identifiés jusqu'à ce jour chez les bactéries à Gram positif à bas GC% (*Listeria*, *Lactobacillus*, *Lactococcus* et *Streptococcus*). La première protéine découverte qui a donné naissance à cette famille de régulateurs transcriptionnels a été mise en évidence chez *Streptococcus gordonii* où elle a été décrite comme contrôlant l'expression d'un gène codant une glucosyltransférase (Sulavik *et al.*, 1992). Sur la base de similarité de séquence, ce nom a ensuite été élargi à toute une famille de protéines contrôlant l'expression de gène(s) cible(s) impliqué(s) dans différentes fonctions physiologiques (Tableau 10).
Ces protéines sont caractérisées par la présence d'un motif HTH appartenant à la famille XRE. Ce motif, localisé dans leur domaine N-terminal, leur assure une fonction de liaison à l'ADN. Par ailleurs, les protéines Rgg présentent un domaine C-terminal conservé (Rgg-Cterm, TIGR01716).

Cette partie reporte les données connues sur ces régulateurs. Dans un premier temps seront traitées les protéines Rgg dites « stand-alones » reconnues comme des régulateurs transcriptionnels autonomes contrôlant l'expression d'un ou plusieurs gènes cibles. Dans un deuxième temps seront exposées les données récentes concernant l'association de protéines Rgg à de petits peptides hydrophobes, décrits comme nécessaires à l'activation de ces dernières.

2. Les protéines Rgg « stand-alone »

2.1 Les protéines Rgg : production de glucosyltransférases

Habitant non-cariogène de la plaque dentaire, *S. gordonii* partage avec d'autres streptocoques (tels que *S. mutans*, *S. pyogenes*, *S. pneumoniae* et *S. salivarius*) la capacité de produire une enzyme extracellulaire, la **g**lucosyl**t**rans**f**érase (GTF), qui hydrolyse le saccharose et les résidus glucosylés pour la synthèse de polysaccharides insolubles tel



Figure 29 – Organisation génétique du locus *gtfG* chez *Streptococcus gordonii* (adapté de Vickerman *et al.*, 2001)**.**

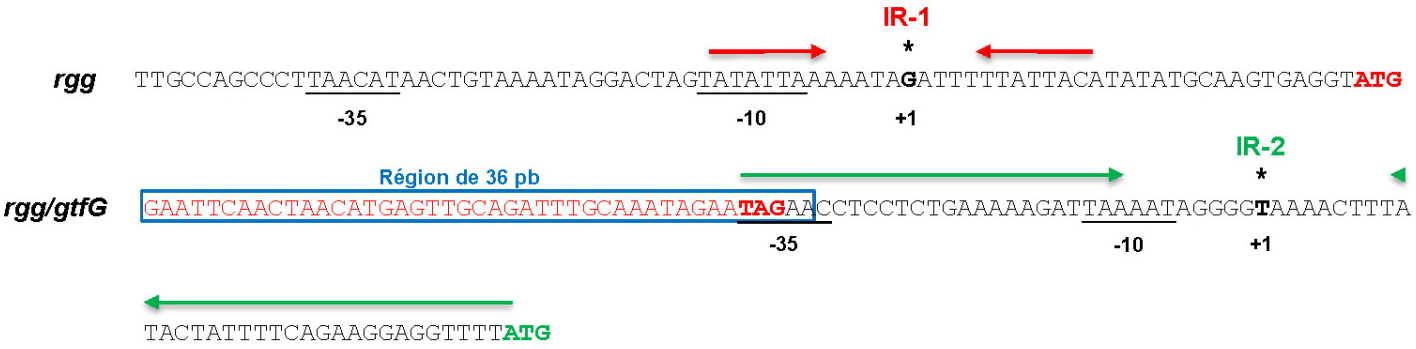


Figure 30 - Présentation des régions promotrices des gènes *rgg* et *gtfG* de *Streptococcus gordonii*.
Les régions répétées inversées identifiées (IR-1 et IR-2) sont indiquées par des flèches. L'utilisation de couleur rouge fait référence au gène *rgg* et la couleur verte, au gène *gtfG*. La totalité de région intergénique *rgg/gtfG* représente la sonde oligonucléotidique utilisée pour l'analyse SPR.

que le glucane. A la différence des autres streptocoques cités ci-dessus, l'espèce commensale *S. gordonii* ne possède qu'une seule enzyme GTF codée par le gène *gtfG* dont l'expression est contrôlée positivement par le produit du gène *rgg*, localisé en amont (Figure 29) (Sulavik *et al.*, 1992). Ce mécanisme de régulation, décrit comme complexe, n'est aujourd'hui toujours pas totalement élucidé.

Trois transcrits ont été associés au locus *rgg/gtfG* : un transcrit *rgg* (1,1 kb), un transcrit *gtfG* (5,5 kb) et un transcrit polycistronique *rgg/gtfG* (6,5 kb) (Sulavik and Clewell, 1996). Le produit du gène *rgg* n'affectant que le transcrit *gtfG*. L'étude des régions promotrices des gènes *rgg* et *gtfG* a révélé la présence de deux régions répétées inversées (Figure 30) (Sulavik *et al.*, 1992). La première région répétée inversée (IR-1), localisée entre la région -10 du promoteur et le site de fixation du ribosome (RBS, **R**ibosome **Bi**nding **S**ite) du gène *rgg*, présage l'action d'un répresseur transcriptionnel bloquant l'initiation de la transcription de ce dernier. La seconde région répétée inversée (IR-2) serait quant à elle à l'origine d'une structure tige-boucle localisée dans la région intergénique *rgg/gtfG*. Cette structure – incluant le codon STOP du gène *rgg* et la région promotrice du gène *gtfG* – influencerait de manière directe l'efficacité de l'expression du gène *gtfG*, en modulant l'accessibilité du RBS aux ribosomes lors d'un couplage traductionnel entre les deux gènes (Sulavik and Clewell, 1996; Vickerman and Minick, 2002). L'étude du mécanisme de régulation de la protéine Rgg sur l'activité GTF a ensuite souligné l'existence d'une région de 36 pb essentielle à une bonne activité GTF. Cette région, localisée en 3' du gène *rgg* (Figure 30), pourrait être impliquée soit directement comme site de fixation du Rgg, soit indirectement comme jouant un rôle dans la structure secondaire et dans l'interaction de la protéine Rgg avec le promoteur du gène *gtfG* (Vickerman and Minick, 2002). Aussi, la régulation de l'expression du gène *gtfG* est contrôlée à deux niveaux : à partir de son propre promoteur ou à partir du promoteur du gène *rgg*.

L'analyse de la séquence nucléotidique d'une souche exprimant 25% d'activité GTF en moins a mis en évidence une mutation près de l'extrémité 3' du gène *rgg* (Vickerman *et al.*, 1997). Par l'utilisation de fusions transcriptionnelles *gftG::lacZ* en système hétérologue, la même équipe a ensuite démontré que cette mutation influence de manière directe l'expression du gène *gtfG* (Vickerman *et al.*, 2003). Enfin, par résonnance plasmonique de surface (SPR, **s**urface **p**lasmon **r**esonnance), les auteurs ont ensuite mesuré, en comparaison avec une protéine Rgg « native », l'affinité de liaison de cette protéine tronquée ($RggC_1$) à l'ADN. De manière surprenante, le complexe $RggC_1$/ADN a présenté un niveau d'affinité deux fois plus élevé que celui du complexe Rgg/ADN (Vickerman *et al.*, 2003). Enfin, bien que l'interaction directe de la protéine Rgg au niveau de la région promotrice du

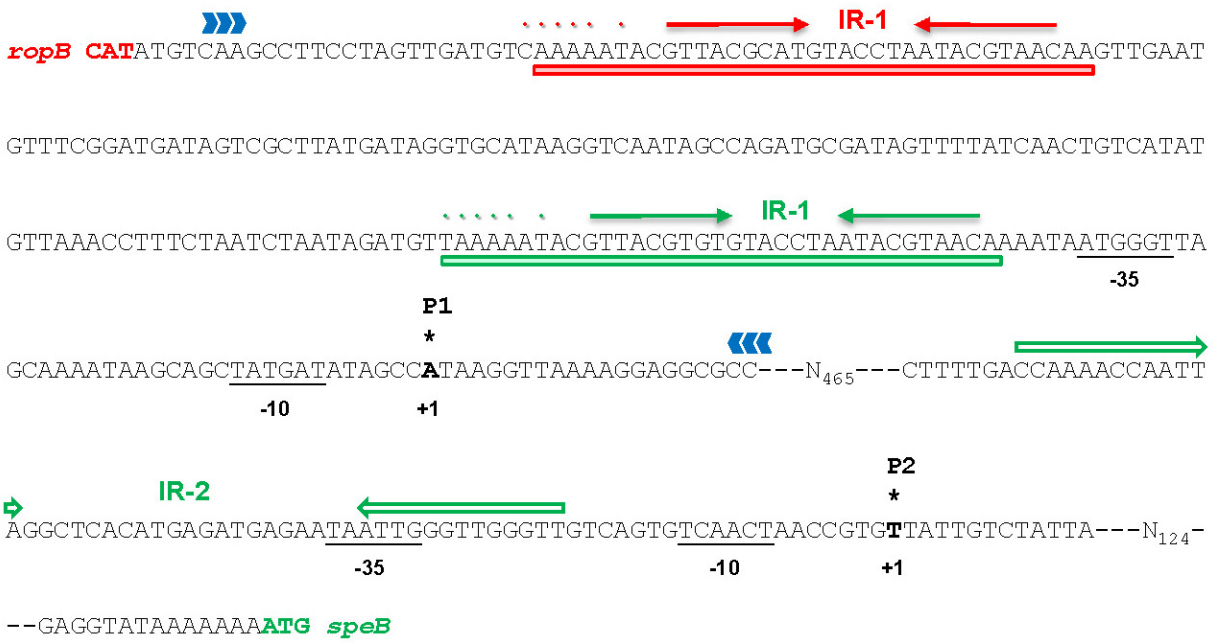


Figure 31 - Représentation schématique de la région intergénique *ropB*/*speB* chez *Streptococcus pyogenes* (adapté de Neely *et al.*, 2003)**.**
Les régions répétées inversées identifiées (IR-1 et IR-2) sont indiquées par les flèches et les terminateurs Rho-independants par des points. La sonde ADN utilisée lors des analyses de retard sur gel est indiquée par les flèches bleues. L'utilisation de la couleur rouge fait référence au gène *ropB*, le vert au gène *speB*.

gène *gtfG* ait été démontrée par cette expérience, l'utilisation d'une sonde de 107 pb (englobant à la fois la région de 36 pb et la séquence répétée inversée IR-2) n'a pas permis de définir précisément le motif de fixation du régulateur à l'ADN (Vickerman *et al.*, 2003).

Par ailleurs, l'étude détaillée de la région aval du locus *gtfG* a révélé la présence d'un deuxième gène codant un régulateur de la famille Rgg et appelé *rggD* (Figure 29). Néanmoins, l'activité GTF n'étant pas affectée dans un mutant *ΔrggD* et la quantité de transcrits des gènes adjacents *ylbN* et *dsg* ne semblant pas être modifiée dans ce même mutant, RggD a été proposé comme régulateur de gènes distaux chez *S. gordonii* (Vickerman *et al.*, 2001).

La séquence nucléotidique et des analyses de Southern blot ont permis d'identifier des déterminants *rgg*-like chez *Streptococcus oralis* (Fujiwara *et al.*, 2000) et *Streptococcus sanguinis* (Vickerman *et al.*, 1995), lesquels apparaissent – au même titre que *S. gordonii* – avoir une seule enzyme GTF. L'organisation génétique des loci *gtf* est par ailleurs conservée et l'inactivation du gène *rgg* chez *S. oralis* montre clairement une diminution de l'activité GTF (gène *gtfR*) chez cette bactérie (Fujiwara *et al.*, 2000). Ceci suggère que le produit du gène *rgg* régulerait également positivement la transcription du gène *gtfR* chez *S. oralis*. Néanmoins, l'analyse de la région intergénique *rgg/gtfR*, indique des différences dans la structure secondaire et suggère que les mécanismes de régulation des gènes *gtf* chez ces deux espèces (*S. oralis* et *S. gordonii*) seraient différents. Enfin, le gène *rgg* de *S. oralis* a été proposé comme marqueur pour l'identification rapide de cette espèce et sa différenciation vis-à-vis d'autres streptocoques du groupe viridans et principalement de *S. pneumoniae* (Park *et al.*, 2010).

2.2 Les protéines RopB et RovS : régulation de la virulence

2.2.1 Le facteur de virulence SpeB chez *Streptococcus pyogenes*

Un régulateur transcriptionnel appartenant à la famille des Rgg a été caractérisé chez *S. pyogenes*. Ce régulateur, appelé RopB (**R**egulator **o**f **p**roteinase), régule positivement l'expression du gène *speB* codant une exotoxine dont l'activité protéolytique contribue largement à la pathogénicité de la bactérie (Chaussee *et al.*, 1999; Lyon *et al.*, 1998). Les gènes *ropB* et *speB* sont adjacents et transcrits de manière divergente (Figure 31). Deux démarrages de transcription (appelés P1 et P2) ont été mis en évidence pour le gène *speB* et deux séquences répétées inversées ont été identifiées en amont de ces deux +1 de transcription. De plus, la séquence répétée inversée identifiée en amont du promoteur P1 a

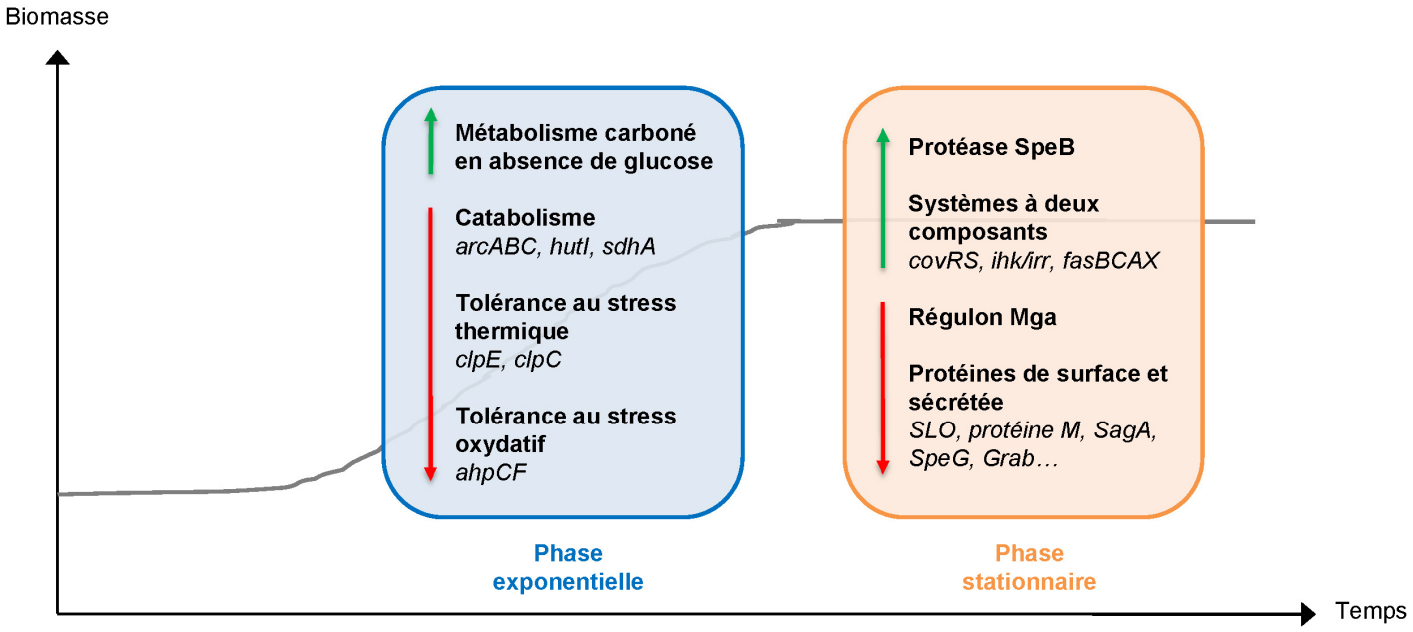


Figure 32 – Contrôle du régulon RopB au cours de la croissance de *Streptococcus pyogenes* (inspiré de McIver, 2009)**.**
Les flèches vertes et rouges indiquent respectivement que l'expression des gènes est activée ou réprimée par RopB.

également été retrouvée dans la région promotrice du gène *ropB* (Figure 31). L'interaction directe de RopB avec la région localisée en amont du promoteur P1 a ensuite été démontrée *in vitro* et l'étude de l'activité des promoteurs P_{speB} et P_{ropB} a révélé que cette région était essentielle à la transcription des deux gènes et pourrait être à la base d'une structure secondaire au niveau de l'ARN impliquée dans le contrôle de l'expression de la protéase SpeB (Figure 31) (Neely *et al.*, 2003).

RopB influence également l'expression de nombreux autres gènes, en phase exponentielle ou stationnaire de croissance (Figure 32). A titre d'exemple, ce régulateur est impliqué dans (i) le contrôle de l'activité catabolique de certaines enzymes impliquées dans le métabolisme de l'arginine (*arcABC*), de l'histidine (*hutI*) et de la sérine (*sdhA*), (ii) la résistance au stress oxydatif et thermique et, (iii) l'utilisation de sources carbonées alternatives telles que le fructose, mannose et sucrose (Chaussee *et al.*, 2004; Chaussee *et al.*, 2003; Dmitriev *et al.*, 2006; Pulliainen *et al.*, 2008). Il contrôle également l'expression de gènes codant des protéines secrétées ou localisées dans la membrane, telles que la protéine M, Mac, SagA et la streptolysine O (Kreikemeyer *et al.*, 2003). Enfin, RopB est également connu pour contrôler négativement le régulon Mga et positivement l'expression de plusieurs systèmes à deux composants tels que *covRS*, *fasBCAX* et *ihk/irr*, tous clairement impliqués dans la virulence et la pathogénicité de la bactérie (Chaussee *et al.*, 2001; Chaussee *et al.*, 2002). Aussi, ce régulateur transcriptionnel est décrit comme un régulateur global chez *S. pyogenes*. Cette fonction est probablement importante pour l'adaptation de ce pathogène à des fluctuations de disponibilités en substrats métaboliques en permettant la coordination entre facteurs de virulence et catabolisme. Néanmoins, le rôle direct ou indirect de RopB avec les cibles citées ci-dessus reste encore à démontrer. Très récemment, l'identification des sites de fixation de RopB a été entreprise par ChIP-chip (**Ch**romatin **I**mmuno**P**recipitation on **Chip**) dans le but de définir le régulon de ce régulateur (Anbalagan *et al.*, 2011). Il a été montré que RopB lie l'ADN au niveau de 65 régions localisées aléatoirement dans le chromosome et dans plus de la moitié des cas, les sites de liaison sont situés dans des régions codantes. L'analyse de ces régions indique que RopB ne semble pas se fixer en amont de régulateurs globaux mais serait plutôt spécifique de gènes impliqués dans le métabolisme. De manière intéressante, les auteurs ont également montré que RopB se lie de manière directe à l'ADN de prophages. Enfin, dans le cadre de l'étude de la régulation de SpeB, Loughman et Caparon ont mis en évidence une enzyme LacD.1 qui en séquestrant Rgg, empêcherait ce dernier d'aller contrôler l'expression des gènes cibles. LacD.1 codant une tagatose 1,6 biphosphate aldolase, impliqué dans le métabolisme du lactose, permettrait un couplage entre transcription et nutrition (Loughman and Caparon, 2006, 2007a).

L'étude de la variabilité des gènes *ropB*-like a permis l'identification de 19 allèles au sein des souches de *S. pyogenes* appartenant au sérotype M3. A l'exception du groupe majoritaire (*ropB1*, 58%), chaque allèle a été décrit spécifique à une lignée clonale. Seulement trois groupes (*ropB1*, *ropB2* et *ropB4*) seraient capables de produire de l'exotoxine SpeB active (Carroll and Musser, 2011). Aussi le polymorphisme observé sur les protéines RopB impacte directement l'expression du gène *speB* et par conséquent la virulence de la bactérie (Carroll and Musser, 2011; Hollands *et al.*, 2008; Ikebe *et al.*, 2010). En 2009, il a par ailleurs été démontré, *in vitro*, que l'absence de production d'exotoxine SpeB dans un variant, était lié à la présence d'une proline en position 103 inhibant l'affinité de liaison de la protéine Rgg à l'ADN dans la région promotrice du gène *speB* (Kappeler *et al.*, 2009). De plus, le polymorphisme observé influence l'ensemble du régulon RopB. Ceci a été démontré pour des souches appartenant au sérotype M1, M49 (Dmitriev *et al.*, 2008) et M3 (Carroll *et al.*, 2011). Le polymorphisme de RopB est principalement localisé dans la partie C-terminale des protéines, laquelle est prédite pour être impliquée dans la formation d'homodimères (Carroll and Musser, 2011). La prédiction structurale et la modélisation de RopB, apparentée à celle du régulateur PlcR de la famille des RNPP, indiquent par ailleurs que RopB fonctionnerait sous la forme d'un homodimère, présentant en N-terminal un domaine de liaison à l'ADN et en C-terminal un domaine intervenant dans la liaison des sous-unités entre elles (Carroll and Musser, 2011; Declerck *et al.*, 2007).

Plus récemment, un régulateur transcriptionnel Rgg a été caractérisé chez *S. suis*. En lien avec RopB, ce régulateur contribue à la virulence de la bactérie (mutant ∆*rgg* est moins pathogène *in vivo*). Il est également impliqué dans le métabolisme de l'arginine et nécessaire à l'utilisation de source carbonée alternative au glucose (Zheng *et al.*, 2011).

2.2.2 L'adhérence aux cellules épithéliales chez *Streptococcus agalactiae*

S. agalactiae est capable de coloniser différentes parties du corps humain grâce à un réseau de régulation contrôlant l'expression de facteurs de virulence. Au sein de ce réseau un régulateur transcriptionnel appartenant à la famille des Rgg a été caractérisé et désigné RovS pour **R**egulator **o**f **v**irulence in ***S****. agalactiae*.

Bien qu'influençant positivement l'expression de plusieurs gènes tels que *gbs0230*, *sodA*, *rogB* et l'opéron *cyl* ; RovS est principalement décrit comme régulant négativement l'expression du gène *fbsA* codant un récepteur au fibrinogène impliqué dans l'adhérence de la bactérie aux cellules épithéliales (Samen *et al.*, 2006). Ceci a été confirmé par des expériences de fixation au fibrinogène et a permis de démontrer un rôle prédominant de

Tableau 11 – Prévalence du gène *fbsA* et de ses régulateurs au sein des principaux complexes clonaux (CC) de *Streptococcus agalactiae* (adapté de Al Safadi *et al.*, 2011)**.**

	CC17	**CC19**	**CC23**	**CC1**	**CC10**
	N° (%)	**N° (%)**	**N° (%)**	**N° (%)**	**N° (%)**
	n = 38	**n = 21**	**n = 20**	**n = 29**	**n = 26**
fbsA	38 (100)	5 (24)	20 (100)	24 (83)	24 (92)
rovS	38 (100)	21 (100)	20 (100)	29 (100.0)	26 (100)
rogB	8 (21)	21 (100)	20 (100)	29 (100.0)	26 (100)
rgf	38 (100)	3 (14.3)	5 (25)	27 (93)	26 (100)

RovS dans la pathogénicité de la bactérie. Aussi, RovS contrecarre l'effet d'un autre régulateur RogB, activateur du gène *fbsA* et régulé positivement mais de manière indirecte par RovS (Gutekunst *et al.*, 2003; Samen *et al.*, 2006). L'étude du gène *rovS* et de son produit ont par ailleurs montré que la protéine RovS jouait son rôle d'activateur et de répresseur en se fixant spécifiquement au niveau d'une séquence palindromique (AWAAWVHTDAW-$N_{6/7}$-WTKWWAMDWAK) située dans la région promotrice de ses différents gènes cibles (Samen *et al.*, 2006). En 2004, Samen et ses collaborateurs ont également démontré, par l'utilisation de fusion rapportrice *fbsA::lacZ*, qu'une mutation dans le gène *oppB* du système de transport d'oligopeptides OppBCDF entrainait une diminution de 35% de l'expression du gène *fbsA* (Samen *et al.*, 2004). Les auteurs ont alors proposé que l'expression du gène *fbsA* pourrait être dépendante d'une régulation de type QS, *via* l'implication de molécules de signalisation réimportées par le transporteur d'oligopeptides Opp. Plus récemment, un système à deux composants, appelé *rgfBDAC*, a également été décrit comme jouant un rôle sur l'expression du gène *fbsA*. La quantité de transcrits de *rovS* n'étant pas altérée dans un mutant Δ*rgfAC*, RovS et RgfBDAC semblent réguler de manière indépendante l'expression du gène *fbsA* (Al Safadi *et al.*, 2011).

L'ensemble de ces études démontre la complexité d'un réseau de régulation bactérien où des activateurs (RogB, RftBDAC) et des répresseurs (RovS) sont finement impliqués dans le contrôle d'un facteur de pathogénicité. Néanmoins, la prévalence du gène *rovS* dans les divers complexes clonaux de *S. agalactiae* laisse présager un rôle majoritaire voir essentiel de ce régulateur au sein de ce réseau (Tableau 11).

2.3 Les protéines RggC et GadR : résistance au stress

2.3.1 La résistance au stress oxydatif chez *Streptococcus thermophilus*

Pour se défendre contre les dommages causés par le stress oxydatif, les streptocoques possèdent différentes enzymes telles que les superoxydes dismutases (Sod), les NADH oxydases, les peroxydases et les glutathions réductases. *S. thermophilus* est une bactérie anaérobie aérotolérante capable de se développer en présence d'oxygène et de survivre à de faibles concentrations en espèces réactives à l'oxygène (ROS, **r**eactive **o**xygen **s**pecies). Aussi, l'absence de peroxydase ou de catalase chez cette bactérie a suggéré l'existence d'un autre système de défense contre les effets délétères des ROS.

Afin de mieux appréhender le système de défense mis en place chez *S. thermophilus*, une collection de mutants obtenus par mutagénèse aléatoire (Thibessard *et al.*, 2002) a été

criblée vis-à-vis de leur sensibilité ou résistance au stress oxydatif (utilisant la ménadione comme agent oxydant). Ce criblage a permis de sélectionner 18 loci impliqués dans le comportement de *S. thermophilus* au stress oxydatif (Fernandez *et al.*, 2004; Thibessard *et al.*, 2004). Parmi les 8 loci présentant un phénotype de résistance, appelé *osr* (**O**xidative **s**tress **r**esistance), le gène *rggC*, codant un régulateur transcriptionnel de la famille des Rgg, a été caractérisé (Fernandez *et al.*, 2006) et son implication dans la résistance au stress oxydatif élargie à d'autres molécules oxydantes telles que le peroxyde d'hydrogène et le paraquat (Fernandez *et al.*, 2004). Aussi, RggC réprimerait l'expression des gènes impliqués dans le mécanisme de défense de la bactérie contre le stress oxydatif.
De manière plus précise, le gène *rggC* est composé de deux cadres ouverts de lecture (*rggC*$_1$ et *rggC*$_2$), tous deux ayant été décrits séparément comme impliqués dans la réponse au stress oxydatif (Fernandez *et al.*, 2006). Cependant, l'absence de RBS en amont de *rggC*$_2$ suggère que l'expression d'une protéine RggC pourrait être conditionnée par un mécanisme de « frameshift » programmé, impliquant un glissement du cadre de lecture généré par l'ARN polymérase au niveau d'un nonamère de thymine connu pour être instable (Fernandez *et al.*, 2006). Bien que cette étude décrive pour la première fois l'implication d'une protéine régulatrice Rgg dans la réponse au stress oxydatif, les cibles physiologiques contrôlées demeurent inconnues.

Plus récemment, l'implication de régulateurs transcriptionnels appartenant à la famille des Rgg a également été mise en évidence dans la réponse au stress oxydatif chez *S. pneumoniae* ainsi que dans la formation de biofilms et la virulence chez cette bactérie (Bortoni *et al.*, 2009). En effet, un mutant du gène *rgg* a été décrit comme plus sensible à l'oxygène et au paraquat et présente une capacité réduite à former des biofilms. De plus, les animaux infectés par ce mutant développent une septicémie plus tardivement que ceux infectés avec la souche sauvage.

2.3.2 Le système antiport glutamate/GABA chez *Lactococcus lactis*

L. lactis, microorganisme reconnu comme GRAS, est largement utilisé en industrie laitière pour ses capacités fermentaires conduisant à la production d'acide lactique et par conséquent à l'acidification de son environnement. Aussi, pour contrecarrer la baisse du pH et maintenir son pH intracellulaire, *L. lactis* met en œuvre différents système antiports couplant la décarboxylation d'un acide aminé (ou composé acide) à une consommation de protons présents dans le cytosol. Parmi ces systèmes, contribuant à l'homéostasie cellulaire, est décrite la décarboxylation du glutamate en GABA (**G**amma-**A**mino**b**utiric **a**cid, Acide gamma-aminobutyrique). Ce système antiport est codé sous la forme d'un opéron, *gadCB*,

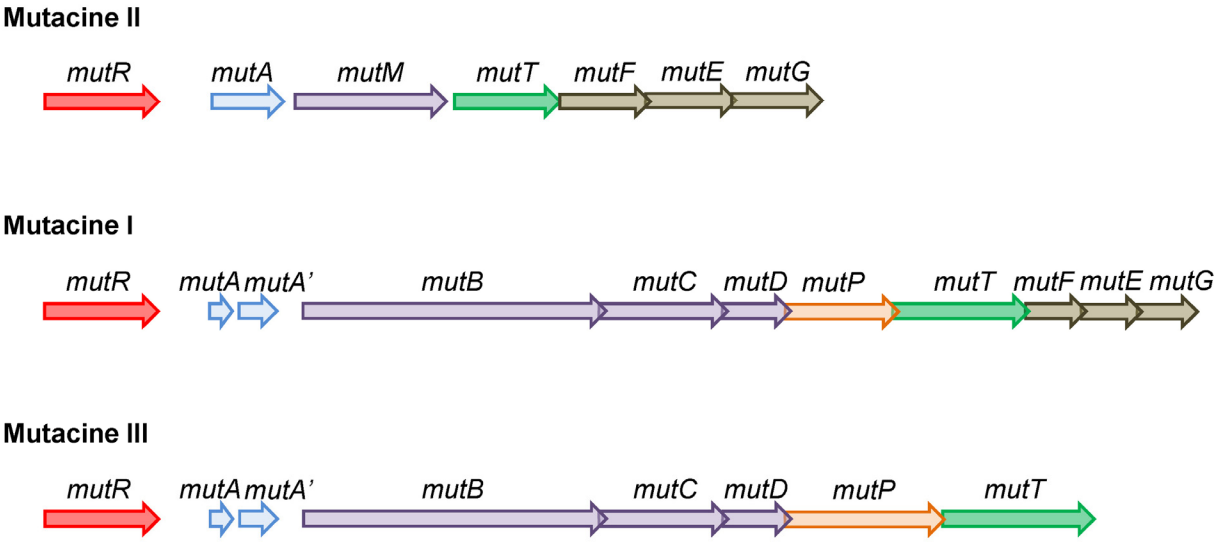


Figure 33 – Organisation génétique des loci codant les voies de biosynthèse de la mutacine II, I et III chez *Streptococcus mutans* (d'après Qi *et al.*, 1999ab ; 2000)**.**
Les couleurs utilisées indiquent les fonctions dans lesquelles les gènes sont impliqués : régulation, pre-pro-mutacine, modifications post-traductionnelles, processing, transport, immunité.

dont l'expression est dépendante de la quantité de glutamate et fonction du pH. Des analyses de Northern blot à partir d'une sonde localisée dans la partie 5' de *gadC*, ont néanmoins indiqué que le produit du gène *gadR*, localisé en amont de l'opéron *gadCB* et codant un régulateur de la famille des Rgg, était essentiel à l'expression des gènes *gadC* et *gadB*. De plus, l'expression de *gadR* a été démontrée, *via* l'utilisation de fusions transcriptionnelles, comme étant constitutive, c'est-à-dire non influencée par la quantité de glutamate présente dans le milieu et le pH de ce dernier. Enfin, l'identification d'une séquence répétée inversée de 21 pb dans la région promotrice de l'opéron *gadCB* a été suggérée comme site de liaison du régulateur à l'ADN (Sanders *et al.*, 1998).

2.4 Les protéines MutR et LasX : production de bactériocines

Les régulateurs Rgg ont été décrits comme intervenant dans la production de peptides antimicrobiens appelés bactériocines. Ces molécules, de par leur large spectre d'action, sont utilisées comme agents conservateurs dans les aliments inhibant ainsi le développement de bactéries indésirables.

2.4.1 La mutacine I, II et III chez *Streptococcus mutans*

Chez *S. mutans*, la production de bactériocines appelées « mutacines » est un facteur de virulence facilitant la dominance de cette espèce au sein de la plaque dentaire. Plusieurs mutacines, appartenant au groupe des lantibiotiques, ont été caractérisées chez cette bactérie ; on retrouve ainsi les mutacines I, II et III, synthétisées sous la forme de précurseurs subissant ensuite diverses maturations et/ou modifications post-traductionnelles. A titre d'exemple, la mutacine II est issue d'un précurseur de 53 acides aminés incluant une séquence signale contenant un motif double glycine reconnu lors du processus de maturation (Woodruff *et al.*, 1998).

L'étude des voies de biosynthèse de ces bactériocines a été entreprise par l'équipe de W. Caufield et F. Qi. De premiers travaux ont tout d'abord montré que la synthèse de la mutacine II était sous le contrôle d'un cluster de gènes composés de 7 cadres ouverts de lecture *mutRAMTFEG* (Figure 33). L'analyse de la séquence a par ailleurs révélé la présence de deux régions promotrices putatives en amont des gènes *mutR* et *mutA* et d'un terminateur de transcription entre ces deux gènes. Aucune région promotrice n'a été identifiée entre les autres gènes du locus dont les cadres ouverts de lecture se chevauchent (Chen *et al.*, 1999). Deux unités de transcription (opéron *mutR* et opéron *mutAMTFEG)* ont

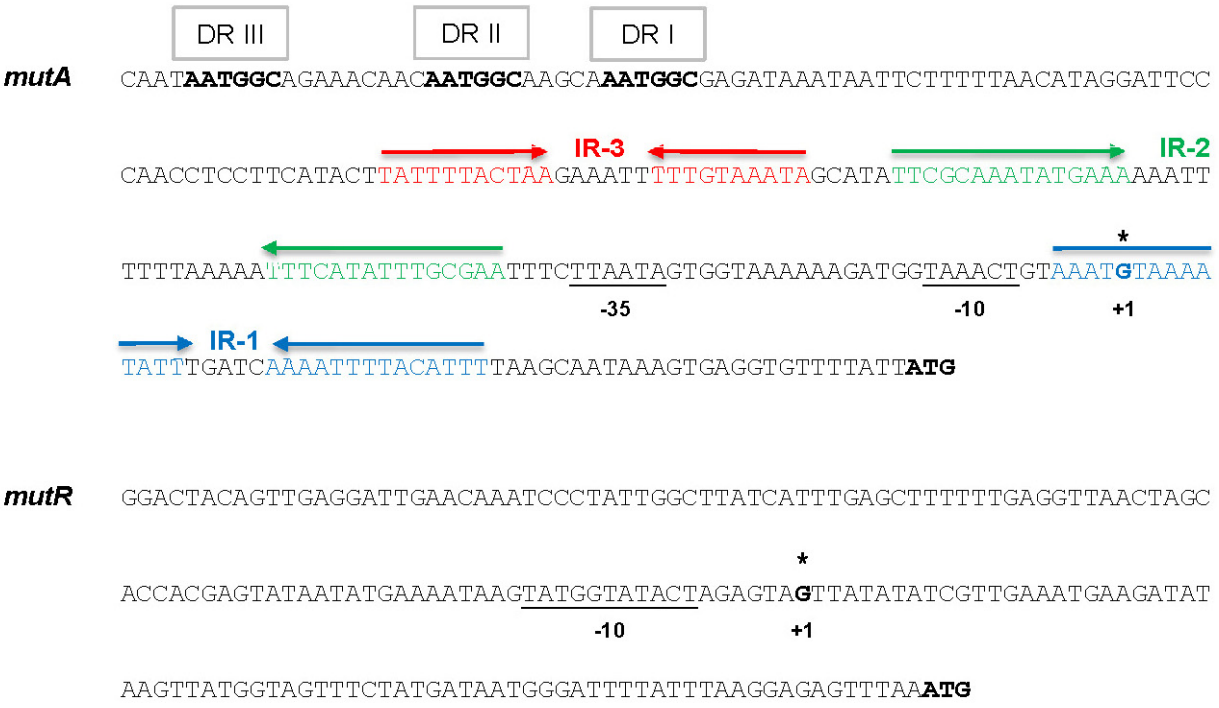


Figure 34 – Représentation schématique des régions promotrices des gènes *mutA* et *mutR* impliqués dans la voie de biosynthèse de la mutacine II chez *Streptococcus mutans* (d'après Qi *et al.*, 1999b)**.**
Les régions répétées inversées identifiées (IR-1, IR-2 et IR-3) sont indiquées par les flèches et les régions répétées directes (DR I, DR II et DR III) par des encadrés.

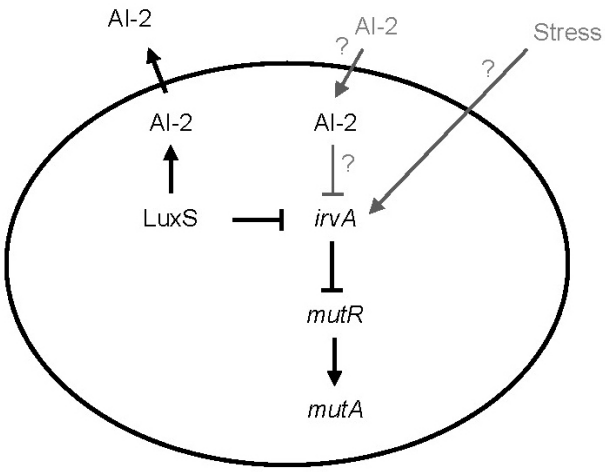


Figure 35 – Modèle proposé pour le contrôle de l'expression de la mutacine I par LuxS chez *Streptococcus mutans* (Merritt *et al.*, 2005)**.**
Les étapes du modèle validées expérimentalement apparaissent en noir.

ensuite été identifiées au sein du cluster (Qi *et al.*, 1999a). Le gène *mutR* étant un homologue du gène *rgg* chez *S. gordonii*, la protéine MutR a alors été proposé comme régulateur de l'opéron permettant la synthèse de la mutacine II. Cette hypothèse a ensuite été démontrée par extension d'amorces en analysant le taux de transcrits du gène *mutA* dans la souche sauvage et son mutant isogénique *⊿mutR* (Qi *et al.*, 1999a). Trois régions répétées inversées ont par ailleurs été identifiées dans la région promotrice du gène *mutA*. En revanche, aucune région répétée inversée ou répétée directe n'a été observée dans le cas de la région promotrice du gène *mutR*. Aussi, par la présence d'une région -10 étendue et l'absence de région -35, les auteurs ont suggéré une expression constitutive de ce dernier (Figure 34) (Qi *et al.*, 1999a).

D'autres études, portant cette fois-ci sur les voix de biosynthèse des mutacines I et III, ont montré la présence de clusters composés respectivement de 11 et 8 cadres ouverts de lectures (*mutRAA'BCDPTFEG* et *mutRAA'BCDPT*) (Figure 33). La comparaison des voix de biosynthèse de la mutacine I et III a révélé : (i) 99% de similitude au niveau de la séquence ADN comprenant la séquence du gène *mutR* et la région intergénique *mutR/mutA* (Qi *et al.*, 2000) et, (ii) 62% d'identité et 79% de similarité entre les protéines MutR des mutacines II et III (Qi *et al.*, 1999b). Ainsi, de par ces similarités de séquence et d'organisation génétique, MutR des mutacines I et III assurerait la même fonction que MutR de la mutacine II. Ceci a été validé *via* l'utilisation d'un système rapporteur pour la biosynthèse de la mutacine I (Kreth *et al.*, 2004). En effet, les auteurs ont montré que l'expression de la fusion *mutA::gusA* était abolie dans un mutant du gène *mutR* en comparaison à la souche sauvage. Enfin, MutR n'étant pas similaire aux autres régulateurs (de bactériocines) connus, il a été suggéré que les gènes de biosynthèse des mutacines aient été acquis par transfert horizontal à partir d'une autre espèce et que seulement le gène *rgg* pourrait avoir évolué au sein de l'espèce *S. mutans* (Qi *et al.*, 1999b).

Plus récemment, un mécanisme de régulation de type QS, décrit sur le locus de la mutacine I, a été mis en évidence suite à l'observation de l'induction du gène *mutA* à haute densité cellulaire. Ce système fait intervenir la protéine LuxS (impliquée dans la biosynthèse de l'autoinducteur 2) qui contrôle négativement l'expression d'un régulateur appelé IvrA (**I**nducible **r**epressor of **v**irulence), lui-même répresseur de MutR (Figure 35) (Merritt *et al.*, 2005). Ceci serait la deuxième démonstration d'une régulation de type QS contrôlant la synthèse d'une bactériocine chez *S. mutans* ; la première contrôlant la synthèse de la mutacine IV (appartenant au groupe des non-lantibiotiques) *via* le système à deux composants ComCDE chez cette bactérie (Kreth *et al.*, 2005).

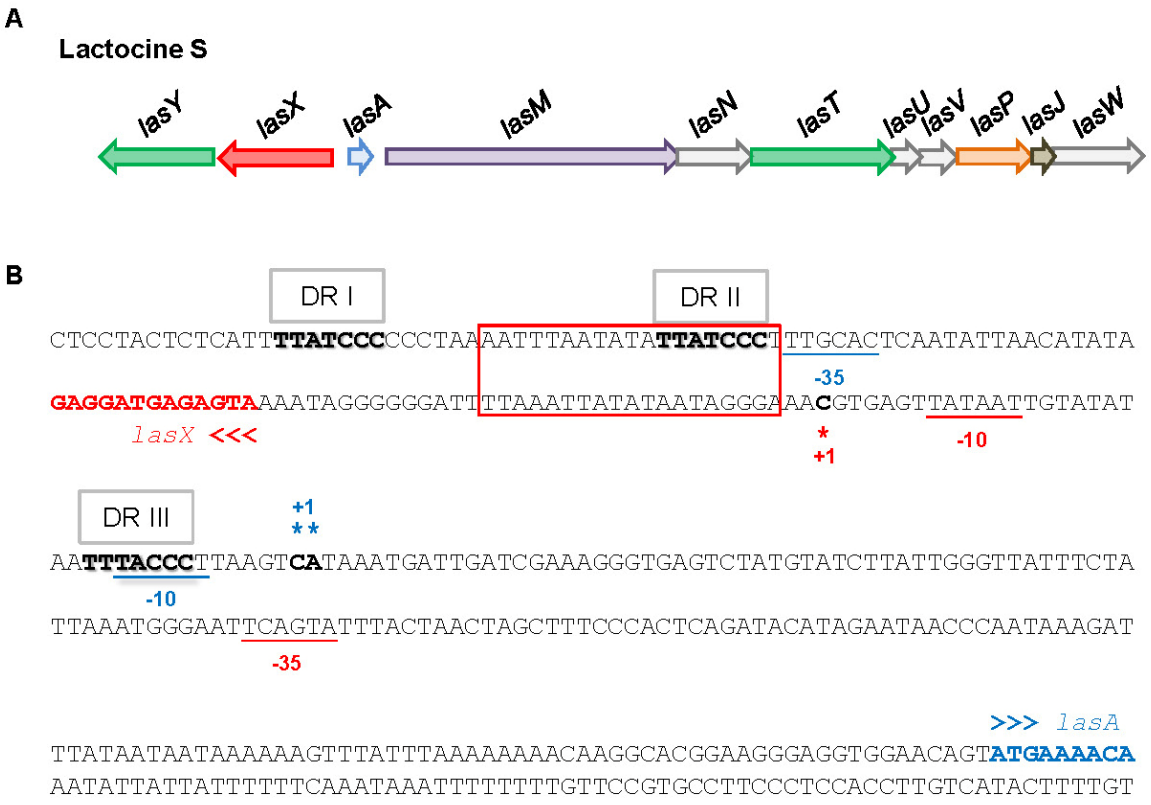


Figure 36 – Le locus de la lactocine S chez *Lactobacilles sakei* (d'après Skaugen *et al.*, 2002 et Rawlinson *et al.*, 2002)**.**

(A) Organisation génétique de la voie de biosynthèse de la lactocine S. Les couleurs utilisées indiquent les fonctions dans lesquelles les gènes sont impliqués : régulation, pre-pro-mutacine, modifications post-traductionnelles, processing, transport, immunité, inconnue.

(B) Représentation schématique des régions promotrices des gènes *lasX* et *lasA*. Les régions répétées directes (DR I, DR II et DR III) et la séquence de 19 pb, nécessaire à la fixation du régulateur à l'ADN sont respectivement identifiées par des encadrés gris et rouge. L'utilisation de la couleur rouge fait référence au gène *lasX*, le bleu au gène *lasA*.

2.4.2 La lactocine S chez *Lactobacillus sakei*

Chez *Lactobacillus sakei*, la production de la lactocine S, peptide antimicrobien appartenant au groupe des lantibiotiques, a également été décrite dépendante d'un régulateur de la famille des Rgg dont la voie de biosynthèse, opéron *lasA-W*, est présentée dans la Figure 36A. A la différence des voies de biosynthèse décrites pour les lantibiotiques produits par *S. mutans*, l'opéron *lasXY* – nécessaire à la production de la lactocine S et localisé en amont de l'opéron *lasA-W* – est transcrit de manière divergente (Figure 36A). Aussi, des analyses de Northern blot n'ayant pas permis de révéler des transcrits *lasA* dans un mutant du gène *lasX* en comparaison à la souche sauvage, le produit du gène *lasX* (*rgg*-like) a été décrit comme nécessaire à l'expression de l'opéron *lasA-W* (Skaugen *et al.*, 2002). L'identification des démarrages de transcription a par ailleurs mis en évidence un chevauchement des régions promotrices des gènes *lasX* et *lasA*, et l'analyse de la région intergénique a révélé la présence de 3 séquences répétées directes heptanucléotidiques (Figure 36B) (Skaugen *et al.*, 2002). Aussi, LasX pourrait être une protéine bifonctionnelle, activant la transcription de l'opéron *lasA-W* et réprimant la transcription de l'opéron *lasXY*. Cette hypothèse fut confirmée, en système hétérologue, par l'utilisation de fusions transcriptionnelles couplant les promoteurs P_{lasA-W} ou P_{lasXY} au gène *gusA*, codant une β-glucuronidase chez *Escherichia coli* (Rawlinson *et al.*, 2002). LasX fut alors le premier régulateur Rgg décrit comme réprimant l'expression de son propre gène. Enfin, des analyses de retards sur gel ont permis de déterminer une séquence de 19 pb considérée comme minimale pour la fixation du régulateur LasX à l'ADN. Cette séquence comprenant la deuxième séquence répétée directe souligne que cette dernière est requise pour l'activation de l'opéron *lasA-W* (Figure 36B) (Rawlinson *et al.*, 2005).

2.5 Conclusion

Les protéines Rgg constituent une famille de régulateurs transcriptionnels contrôlant des fonctions biologiques variées chez les bactéries à Gram positif à bas GC%. Bien que la majorité des protéines Rgg semble être associée à un seul gène cible ou opéron proximal, certaines études indiquent qu'elles seraient également impliquées dans de grands réseaux de régulation et joueraient un rôle global chez plusieurs organismes. Ceci a été décrit pour RopB de *S. pyogenes* (Chaussee *et al.*, 2002), RovS de *S. agalactiae* (Samen *et al.*, 2006), Rgg de *S. suis* (Zheng *et al.*, 2011).

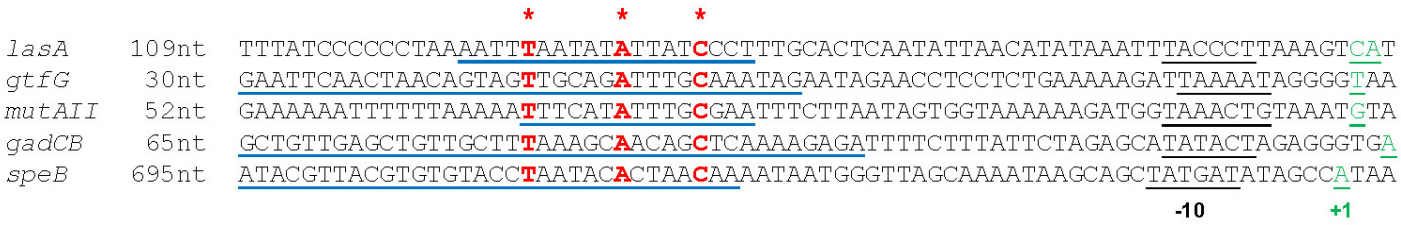


Figure 37 – Alignement des régions promotrices de plusieurs gènes cibles de régulateurs Rgg (adapté de Rawlinson *et al.*, 2005)**.**

Alignement des régions promotrices des gènes *lasA* de *Lactobacillus sakei*, *gtfG* de *Streptococcus gordonii*, *mutA* de *Streptococcus mutans*, de *gadCB* de *Lactococcus lactis* et de *speB* de *Streptococcus pyogenes*.

Les résidus conservés sont indiqués en rouge. En bleu sont soulignées les régions proposées ou décrites comme essentielles à l'activation des différents gènes cibles. La distance entre le +1 de transcription et l'ATG du gène est indiquée en amont de la séquence.

Les protéines Rgg présentent 32%[4] d'homologie et partagent trois résidus conservés dans leur séquence : une glycine et une arginine localisées dans leur partie N-terminale et un tryptophane dans leur partie C-terminale. Ces trois résidus ont par ailleurs été décrits comme essentiels à l'activité régulatrice de la protéine RopB chez *S. pyogenes* (Loughman and Caparon, 2007b). Aussi, bien que le rôle de la glycine soit encore inconnu, les résidus arginine et tryptophane seraient respectivement impliqués dans la stabilité du complexe ADN/protéine et dans l'interaction protéines/protéines.
De par la présence de leur domaine HTH, les protéines Rgg sont capables de se lier à l'ADN. Leur capacité à fixer les régions promotrices de leurs gènes cibles a clairement été démontrée *in vitro* pour RopB, RovS et LasX. Cependant, aucun site de fixation de ces régulateurs n'est encore précisément identifié. Toutefois, l'alignement des régions promotrices des gènes cibles de plusieurs régulateurs Rgg a souligné l'existence d'un motif dégénéré (ayant probablement évolué au sein des différentes espèces) recoupant les différentes régions proposées ou décrites comme essentielles à l'activation de ces différents gènes cibles (Figure 37). Enfin, des structures secondaires au niveau de l'ARN semblent également jouer un rôle non négligeable à l'activation de certains gènes cibles.

Malgré un vaste répertoire couvrant plusieurs genres bactériens, les études les plus détaillées se limitent essentiellement aux streptocoques (pathogènes). Aussi, l'étude de ces protéines dans d'autres genres bactériens, tels que Listeria, pourrait contribuer à élargir et compléter les connaissances actuelles sur ces protéines. En effet, malgré quelques informations concernant les signaux environnementaux ou métaboliques auxquels répondent ces protéines, aucune voie de transduction du signal n'est à ce jour clairement définie. De plus, le mécanisme de régulation de la transcription des gènes par les régulateurs Rgg « stand-alones » semble être un phénomène complexe et variable entre les espèces. Aussi, l'identification de ces éléments de régulation permettrait potentiellement de révéler l'implication de ces régulateurs dans de nouvelles fonctions physiologiques.

3. Les protéines Rgg associées à de petits peptides hydrophobes

Bien que les protéines Rgg soient principalement étudiées en tant que régulateurs « stand-alones », l'analyse des génomes de streptocoques a récemment fait ressortir deux sous-familles de protéines Rgg associées à de petits peptides hydrophobes. Ces deux familles connues sous le nom de systèmes « SHP/Rgg » et « XIP/ComR » sont présentées dans cette partie.

[4] Médiane calculée sur l'homologie de 13 paires de Rgg.

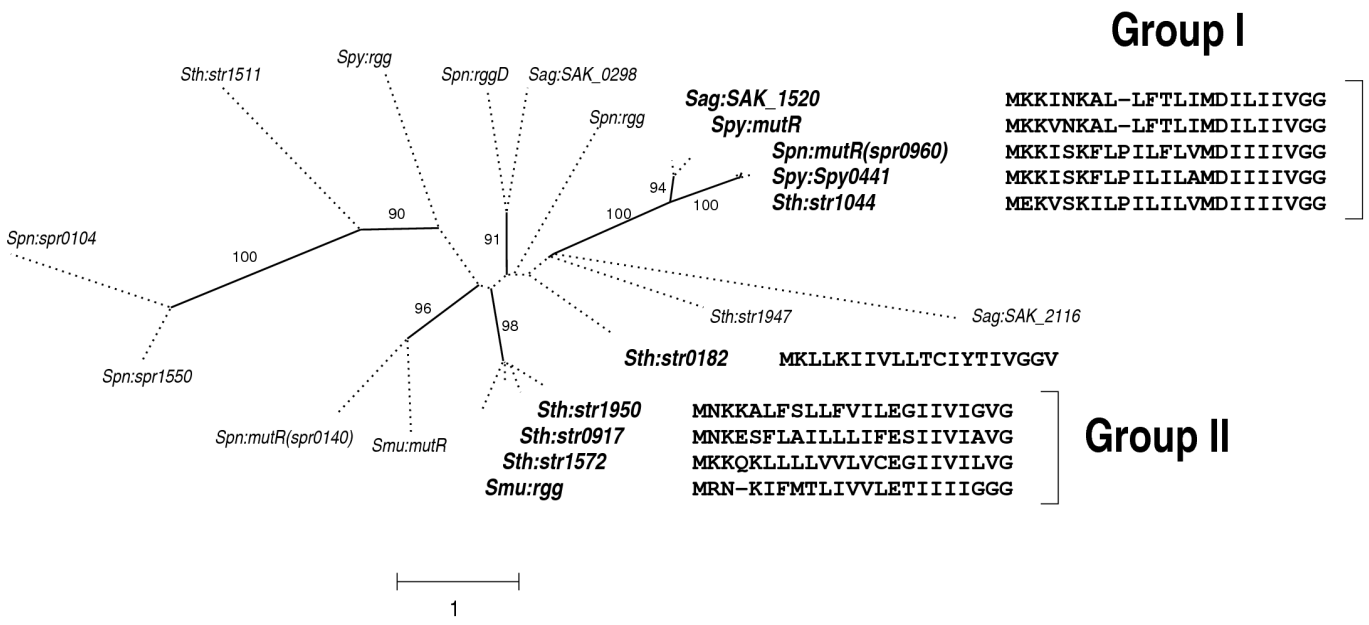


Figure 38 – Arbre phylogénétique des protéines Rgg chez les streptocoques (Ibrahim *et al.*, 2007b)**.**

Les travaux portant sur les systèmes SHP/Rgg ont été réalisés au sein de l'équipe ComBac et les résultats qui en ont découlé ont été à la base de mon travail de thèse. Par ailleurs, ces travaux représentaient l'ensemble des données disponibles dans la littérature au démarrage de ma thèse. Les travaux caractérisant la famille de systèmes XIP/ComR ont quand à eux été publiés durant la première moitié de ma thèse.

3.1 Détection de petits gènes dans l'environnement proche de gènes *rgg*

L'identification et l'annotation des petits gènes est un problème récurant en bio-informatique. Aussi, de par l'importance de leur rôle dans différentes fonctions, une étude portant principalement sur les génomes des streptocoques séquencés a été conduite au laboratoire afin d'identifier de nouveaux petits gènes pouvant potentiellement coder des peptides impliqués dans un mécanisme de communication de type QS (Ibrahim *et al.*, 2007b). Sur la base de critères intrinsèques (composition nucléotidique de la séquence et présence d'un RBS) et extrinsèques (conservation de la séquence étudiée et/ou de l'environnement génétique), une liste de petits gènes putatifs (spCDS, **s**hort **p**utative **C**oding **D**NA **S**equence) a été définie. En moyenne, 39 spCDS ont été détectés par génome de streptocoques et l'expression de sept d'entre eux, chez *S. thermophilus*, confirmée par RT-PCR (Ibrahim *et al.*, 2007b).

En parallèle à ces résultats, l'analyse des contextes génétiques de ces spCDS a fait ressortir une famille de spCDS localisés en amont de gènes codant des régulateurs transcriptionnels Rgg et transcrits de manière divergente. Cette famille de spCDS transversale au genre des streptocoques code des petits peptides hydrophobes, appelés SHP (**S**hort **H**ydrophobic **P**eptide), présentant une taille conservée allant de 20 à 23 acides aminés et caractérisés par une ou plusieurs charges positives en N-terminal (lysine), un domaine central hydrophobe présentant un glutamate (E) ou un aspartate (D) conservé et la présence d'une glycine en C-terminal. La construction d'un arbre phylogénétique des protéines Rgg chez les streptocoques a alors été entreprise afin de mieux définir les relations associant un gène *rgg* à une spCDS codant un SHP (Figure 38). Aussi, bien que tous les gènes *rgg* ne soient pas automatiquement associés à un gène *shp*, une coévolution des protéines Rgg avec ces petits peptides hydrophobes a pu être mise en évidence révélant ainsi l'existence de deux groupes de systèmes SHP/Rgg (Figure 38) (Ibrahim *et al.*, 2007b).

L'ensemble de ces résultats, couplé à la similarité des peptides SHP avec les anti-phéromones de *E. faecalis* (chapitre 2, § 2.2.4), ont suggéré un rôle éventuel des protéines Rgg associées à ces petits peptides dans des phénomènes de type QS.

3.2 Un transporteur d'oligopeptides impliqué dans la réimportation de peptides phéromones chez les streptocoques

3.2.1 Hypothèse d'un nouveau mécanisme de QS : les systèmes SHP/Rgg

Dans la littérature, un seul mécanisme de QS contrôlant la production d'une bactériocine *via* un système à deux composants est décrit chez *S. thermophilus* (Fontaine *et al.*, 2007). Aussi, en parallèle de l'approche bio-informatique et afin de mettre en évidence de nouveaux mécanismes de communication cellulaire de type QS chez *S. thermophilus*, la recherche de peptides phéromones dans les surnageants de cultures de cette bactérie a été entreprise au laboratoire. Les surnageants des souches LMD-9, CNRZ1066 et LMG18311 ont tout d'abord été analysés par RP-HPLC (**R**everse-**P**hase **H**igh **P**erformance **L**iquid **C**hromatography). Parmi les peptides identifiés, un seul spécifique de la souche LMD-9 et codé par le gène *ster_1357* a été caractérisé (Ibrahim *et al.*, 2007a). Ce peptide, nommé Pep1357C, est issu d'un clivage de son précurseur et présente une modification post-traductionnelle résultant d'une cyclisation entre une lysine et un tryptophane. L'analyse de l'environnement génétique sur le génome de *S. thermophilus* LMD-9 a révélé que le gène *ster_1357* était localisé en aval d'un gène *rgg* (*ster_1358*) associé à une spCDS appartenant à la famille des SHP décrite précédemment. Aussi, (i) les régulateurs Rgg étant principalement décrits comme contrôlant l'expression de gènes cibles adjacents et, (ii) les peptides SHP présentant des caractéristiques avec les anti-phéromones de *E. faecalis* dont l'importation dans les cellules bactériennes nécessite l'intervention d'un transporteur d'oligopeptides (Opp) ; les surnageants de *S. thermophilus* LMD-9 wt et des mutants du système de transport d'oligopeptides (Δ*ami*), du gène *rgg* (Δ*ster_1358*) et du gène *shp* (Δ*shp*) ont été analysés par RP-HPLC (Ibrahim *et al.*, 2007a). Le produit du gène *ster_1357* ne fut pas détecté dans les surnageants de ces différents mutants (Ibrahim *et al.*, 2007a). Ces résultats confirmés par RT-PCR ont alors permis de suggérer un nouveau mécanisme de communication cellulaire de type QS impliquant (i) le produit du gène *shp* jouant le rôle de phéromone, (ii) le transporteur d'oligopeptides Ami assurant le réimport de la phéromone à l'intérieur de la cellule et (iii) un régulateur transcriptionnel Rgg contrôlant l'expression de gène(s) cible(s), dans ce cas l'expression du gène *ster_1357*.

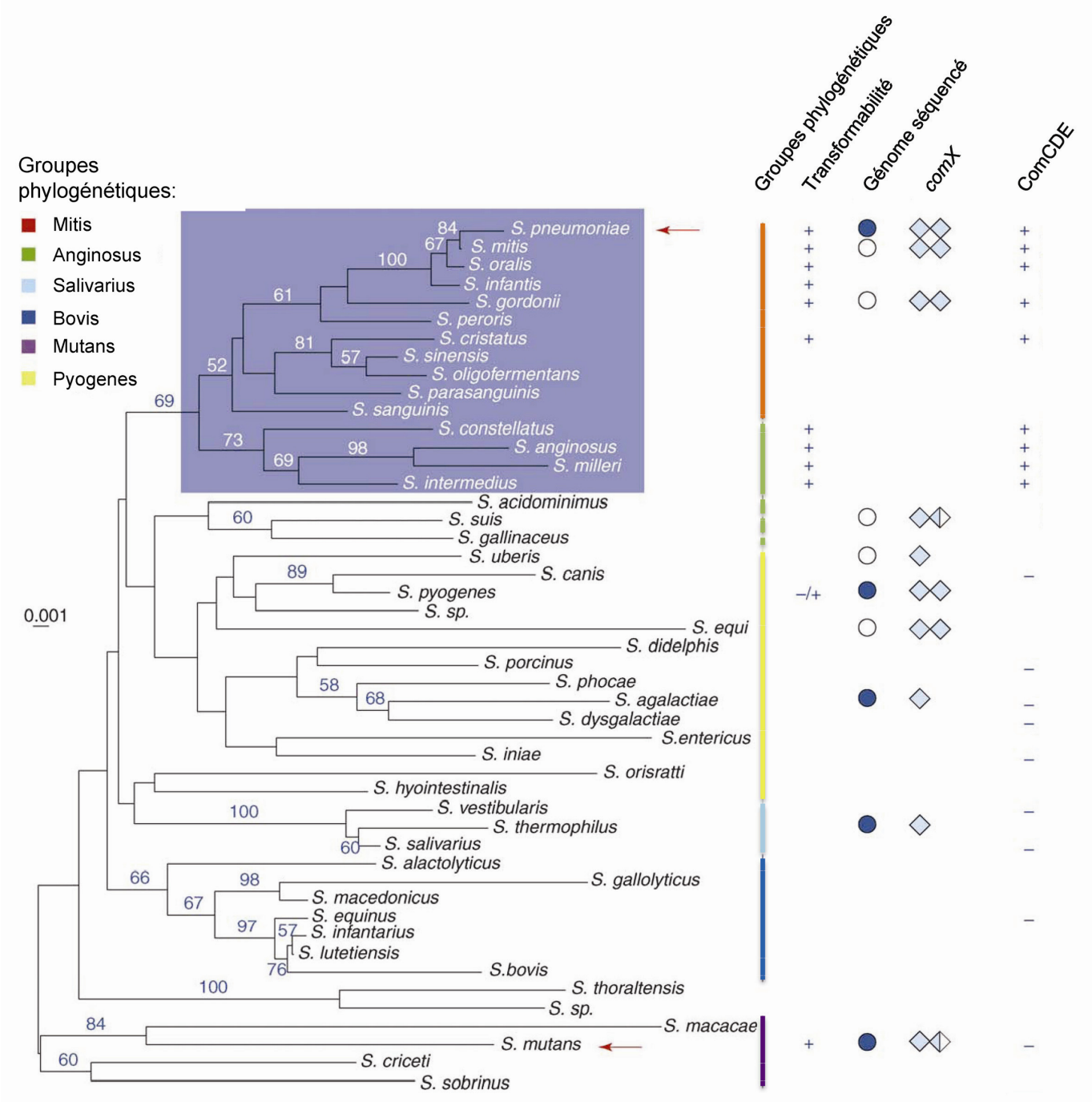


Figure 39 – Relations phylogénétiques basées sur l'ARN 16S et distribution du système ComCDE (et de ses paralogues) chez les streptocoques (adapté de Martin *et al.*, 2006)**.**
Les groupes phylogénétiques sont indiqués par l'attribution d'une couleur respective, le symbole « + » indique respectivement la transformabilité des espèces et la présence d'un système ComCDE dans le génome des streptocoques, les cercles indiquent la présence de génomes séquencés « ● » ou en cours de séquençage « ○ », le nombre de losanges indique le nombre de copies du gène *comX* identifié dans les génomes.

3.2.2 Le contrôle de la compétence au sein du groupe salivarius

Comme vu dans le chapitre 2 (§ 2.1.3) de cette introduction bibliographique, les streptocoques du groupe mitis et anginosus contrôlent l'expression d'un facteur sigma alternatif, ComX, nécessaire à l'expression des gènes tardifs de compétence, *via* le système à deux composants ComCDE. Ce système a par ailleurs été décrit comme absent des génomes de streptocoques séquencés issus des groupes salivarius, pyogenes et bovis, bien que le régulon ComX soit présent chez ces derniers (Figure 39) (Martin *et al.*, 2006).

En 2006, la surexpression de ComX chez *S. thermophilus* a permis de restaurer la compétence chez cette bactérie (Blomqvist *et al.*, 2006b). Ces résultats ont alors suggéré que la régulation de la compétence, sous-entendue du gène *comX*, avait été perdue chez cette bactérie ou plus probablement que celle-ci était dépendante d'un autre système de régulation. Aussi, qu'en était-il du/des éventuel(s) système(s) qui régule(nt) la compétence au sein des différents groupes de streptocoques (salivarius, pyogenes et bovis) décrits comme non naturellement compétents ?

Afin d'identifier par une approche globale des cibles des systèmes de QS qui sont contrôlés par des peptides réimportés, une approche protéomique (couplant gel 1D et LC-MS/MS) permettant de comparer le protéome d'une souche sauvage et mutée au niveau du système de transport d'oligopeptides (Δ*ami*) a été entreprise chez *S. thermophilus* LMD-9, à partir de cultures réalisées en **m**ilieu **c**himiquement **d**éfini (MCD) sans peptide (Gardan *et al.*, 2009). Parmi le pool de données, sept protéines (dont ComX) décrites comme essentielles pour le déclanchement de la compétence n'ont pas été détectées dans les protéomes issus du mutant du système de transport. Ces résultats validés par RT-PCR sur certains gènes de compétence sélectionnés (*comX*, *recA*, *dprA* et *comGA*) ont clairement démontré l'implication du transporteur d'oligopeptides dans le déclenchement de ce phénomène (Gardan *et al.*, 2009). Par ailleurs, les auteurs ont étudié le déclenchement de la compétence en MCD dépourvu de peptide. La souche sauvage de *S. thermophilus* s'est révélée être naturellement compétente dans ces conditions de croissance, en comparaison de son mutant isogénique Δ*ami*. Au même titre que *S. pneumoniae*, la compétence s'est également avérée être un phénomène transitoire chez *S. thermophilus*, se produisant en début de phase exponentielle de croissance (Gardan *et al.*, 2009). L'ensemble de ces résultats a alors suggéré que le déclenchement de la compétence chez *S. thermophilus* pouvait être assuré par un mécanisme de réimport de peptides *via* le système de transport d'oligopeptides Ami.

Peu de temps après et à partir des résultats d'une approche transcriptomique, le régulateur contrôlant l'expression de ComX chez *S. thermophilus* a été identifié. Ce

régulateur codé par le gène *ster_0316* a été décrit comme appartenant à la famille des Rgg (Fontaine *et al.*, 2010a). Aussi, sur la base des résultats décrits chez Ibrahim *et al.* (2007a,b), associant un régulateur Rgg à un petit peptide hydrophobe SHP, la présence d'une spCDS a été recherchée dans l'environnement proche du gène *ster_0316*. Une petite ORF localisée en aval, codant un petit peptide de 24 acides aminés, a alors été identifiée. Afin de démontrer le rôle de ces deux partenaires (régulateur Rgg et petit peptide) dans le déclenchement des gènes précoces de compétence, des fusions transcriptionnelles couplant le promoteur du gène *comX* à un système rapporteur (*luxAB*) ont été construites et leur activité mesurée dans différents mutants de *S. thermophilus* délétés du gène *rgg* (*ster_0316*), de la petite ORF ou du système de transport Ami (témoin négatif). L'absence d'activité du promoteur P_{comX} dans ces trois mutants a alors démontré l'essentialité du gène *rgg* et de la petite ORF (en plus du transporteur d'oligopeptides, précédemment décrit) pour le déclenchement de la compétence chez *S. thermophilus* (Fontaine *et al.*, 2010a).

L'identification de la forme mature du petit peptide de 24 acides aminés fut ensuite entreprise afin de savoir si celui-ci correspondait à la molécule de signalisation induisant le déclenchement de la compétence. En effet, si tel est le cas, ce peptide est supposé être maturé, sécrété et réimporté dans la cellule. Aussi, l'extrémité C-terminal du peptide fut synthétisée (7 à 10 aa). L'ajout des peptides synthétiques dans les surnageants de culture du mutant de la petite ORF contenant la fusion transcriptionnelle P_{comX}::*luxAB* dans son génome a alors permis de restaurer l'activité du promoteur P_{comX}, bien que celle-ci reste inférieure au niveau sauvage (Fontaine *et al.*, 2010a). En revanche, aucune complémentation n'a été observée avec le mutant du système de transport Ami. L'utilisation d'une banque de peptides synthétiques dérivant d'un précurseur a par ailleurs permis de définir une taille minimale nécessaire à l'activité du peptide et correspondant à un heptapeptide. De plus, la présence des deux résidus côté C-terminal du précurseur a été décrite comme essentielle à l'activité du peptide (Fontaine *et al.*, 2010a). Sur la base de ces résultats, les gènes codant le régulateur *rgg* et la petite ORF ont respectivement été appelés *comR* pour « **Com**petence **r**egulator » et *comS* pour « **Com**petence **s**ignal » (Fontaine *et al.*, 2010a). Enfin, l'analyse des régions promotrices des gènes précoces identifiés par transcriptomique a fait ressortir une région répétée inversée conservée appelée Ecombox (5'-TAGTGACATNTATGTCACTA-3'). Cette région a également été identifiée dans la région promotrice des gènes *comS* et *comX* (Fontaine *et al.*, 2010a). L'ensemble de ces résultats a alors permis de proposer un système de QS, similaire a celui exposé pour les systèmes SHP/Rgg, où la phéromone ComS est importée à l'intérieur de la cellule par le système Ami où elle interagirait avec le régulateur ComR (Rgg-like) pour la régulation des gènes cibles impliqués dans le contrôle de la compétence. La fonctionnalité de ce système semble par

ailleurs effective chez *S. salivarius* où un locus orthologue a pu être identifié (Fontaine *et al.*, 2010a). Aussi qu'en était-il des groupes pyogenes et bovis ?

3.2.3 Le contrôle de la compétence chez les streptocoques du groupe pyogenes et bovis : les systèmes XIP/ComR

A partir de l'analyse des régions promotrices des gènes *comX* (ou *sigX*), présents au sein des groupes pyogénique et bovis, Mashburn-Warren et ses collaborateurs ont pu mettre en évidence un motif conservé (5'-AACATGGACAN$_4$TGTCACTGTT-3') différent de celui décrit au sein du groupe salivarius (EcomBox). Aussi, afin d'identifier d'autres cibles putatives pouvant contribuer au déclenchement de la compétence (gènes précoces), la recherche de ce motif dans les génomes des streptocoques (tout groupe confondu) a été entreprise. Les résultats ont alors indiqué la présence de ce motif dans les génomes des streptocoques des groupes pyogénique et bovis ainsi que chez *S. mutans*, au niveau de la région promotrice d'un gène homologue au gène *comR* caractérisé chez *S. thermophilus*. Ces résultats ont alors suggéré que le système ComRS décrit chez *S. thermophilus* et impliquant un transporteur d'oligopeptides pouvait également être fonctionnel au sein d'autres groupes de streptocoques – ne présentant pas de système ComCDE dans leur génome – tels que pyogenes et bovis. De plus, ils ont également souligné une possible accumulation des deux systèmes impliqués dans le contrôle des gènes précoces de compétence – ComCDE et ComRS – chez *S. mutans* (Mashburn-Warren *et al.*, 2010).

Afin de conforter leurs hypothèses, les auteurs ont alors recherché la présence de spCDS dans l'environnement aval des gènes codant une protéine homologue à ComR. Ils ont ainsi identifié une nouvelle famille de petits peptides hydrophobes qu'ils ont caractérisé par la présence d'une charge globale positive et d'un motif double tryptophane en C-terminal (Mashburn-Warren *et al.*, 2010). Ensuite, en utilisant la même stratégie que celle développée par Fontaine *et al.* (fusions transcriptionnelles, complémentation et calcul du nombre de transformants dans différents mutants), les auteurs ont étudié la fonctionnalité de ces systèmes chez *S. mutans*, seul streptocoque pour lequel les conditions de croissance nécessaires au déclenchement de la compétence sont connues. Ils ont ainsi (i) démontré l'essentialité du produit du gène homologue à *comR*, du petit peptide codé par la petite ORF et du système de transport Opp pour l'expression du gène *comX*, (ii) démontré le rôle du petit peptide en tant que molécule de signalisation et (iii) proposé, à l'aide de peptides synthétiques, une forme mature possible de ce dernier (correspondant à un heptapeptide) pouvant être généré suite à un clivage côté C-terminal du précurseur et présentant le motif double tryptophane dans sa séquence (Mashburn-Warren *et al.*, 2010). Ils ont alors nommé

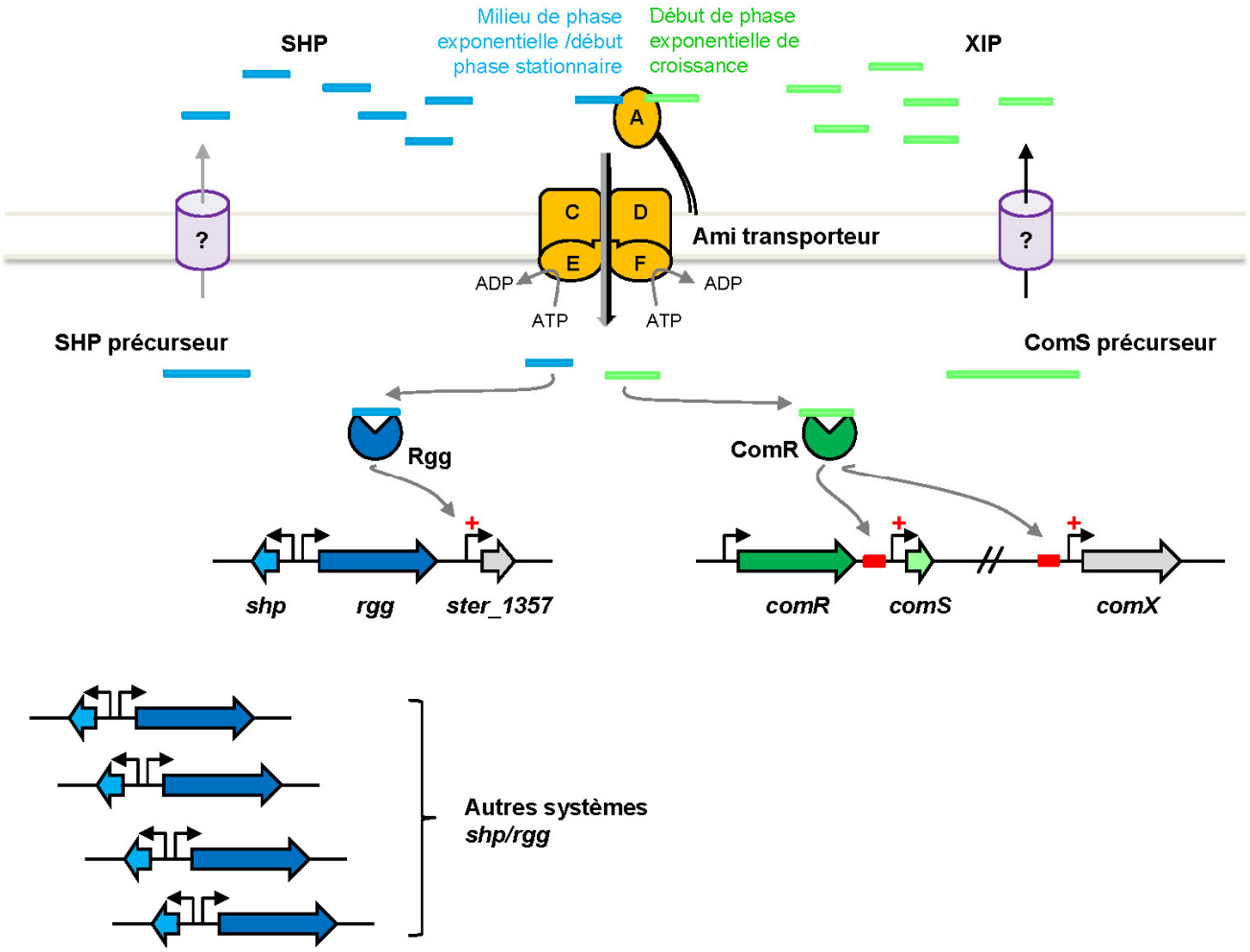


Figure 40 – Représentation schématique des systèmes SHP/Rgg et XIP/ComR chez *Streptococcus thermophilus* LMD-9.
Les étapes validées dans ce modèle sont indiquées par la présence de flèches noires. Les mécanismes sont détaillés dans le texte.

la forme mature du peptide codé par le gène *comS* « XIP » pour ***s****igX*-**i**nducing **p**eptide. Par ailleurs, les auteurs ont également décrit un rôle central de ce système pour le déclenchement de la compétence chez cette bactérie qui contient également le système ComCDE dans son génome. Enfin, ils ont démontré que ces deux systèmes fonctionneraient en tandem ; le système ComCDE agissant en amont de ComRS. Néanmoins, les éléments de régulation liant ces deux systèmes restent à ce jour inconnus (Mashburn-Warren *et al.*, 2010).

3.3 Conclusion

Sur la base de ces études et en regard des protéines Rgg « stand-alones », certaines protéines Rgg semblent être capables de répondre à des molécules de signalisation et ainsi participer à la mise en place de mécanismes de QS chez les streptocoques. L'ensemble des résultats présenté dans cette partie est résumé dans la Figure 40.

Les systèmes ComRS ou XIP/ComR ont clairement été décrits comme impliqués dans la compétence de certains groupes de streptocoques considérés jusqu'alors comme non transformables. Néanmoins, il n'en reste pas moins vrai que les conditions de transformation naturelle restent inconnues pour les groupes pyogenes et bovis. Par ailleurs, bien que les petits peptides hydrophobes ComS/XIP aient été décrits comme étant les phéromones nécessaires à l'activation de ces systèmes, l'utilisation de peptides synthétiques n'a pas permis de définir précisément la forme mature de ces phéromones. De plus, bien que ces phéromones semblent influencer l'état d'activation du régulateur ComR, leur interaction directe reste cependant à démontrer. Il en est de même concernant la régulation du gène *comX* par ComR *via* la fixation de ce dernier sur le motif ADN identifié. Enfin, l'identification de motifs de fixation différents de ceux décrits pour les protéines Rgg « stand-alones », présume d'une évolution de ces systèmes au sein de la famille des Rgg. La formation de structures secondaires au niveau de l'ARN semble cependant être un caractère conservé.

Objectifs du travail de thèse

Chez les bactéries à Gram positif, les études portant sur la communication cellulaire de type QS ont longtemps été focalisées sur les bactéries modèles (*e.g. B. subtilis*) ou pathogènes (*e.g. S. aureus*) dû à l'implication de ce mécanisme dans le contrôle de l'expression de gènes liés à la virulence. *A contrario*, peu d'études sur des bactéries d'intérêt industriel, telles que les bactéries lactiques, sont disponibles dans la littérature. Aussi, au laboratoire, une thématique s'est développée autour de l'étude des interactions bactériennes de type QS chez *S. thermophilus*.

Dans ce contexte de recherche, les travaux de thèse de P. Garault ont tout d'abord montré que le système de transport d'oligopeptides (Opp), appelé Ami chez *S. thermophilus*, ne semblait pas être impliqué uniquement dans la nutrition azotée. En effet, une différence de croissance entre la souche sauvage et un mutant du système de transport Δ*opp* a pu être observée en milieu MCD sans peptide (Garault *et al.*, 2002). Par spectrométrie de masse, il a ensuite été montré que le système de transport Opp intervient dans la réimportation de peptides allant de 3 à 23 résidus et pouvant présenter la particularité d'être hydrophobes (Garault *et al.*, 2002). Ces résultats ont alors suggéré un rôle de ce transporteur dans des phénomènes de signalisation de type QS chez *S. thermophilus*. Dans un deuxième temps, les travaux de thèse de M. Ibrahim, ont alors consisté à mettre en évidence et identifier des peptides « phéromones » chez *S. thermophilus*. Pour ceci, deux approches complémentaires ont été développées : une analyse *in silico*, permettant de détecter la présence de petits gènes dans les génomes des streptocoques et une analyse biochimique, consistant à analyser le pool de peptides présents dans les surnageants de culture de *S. thermophilus* en MCD. Alors que la première approche a permis – via la détection d'une nouvelle famille de petits gènes – d'identifier une nouvelle organisation génétique codant un petit peptide hydrophobe SHP (**s**hort **h**ydrophobic **p**eptide) et un régulateur transcriptionnel appartenant à la famille des Rgg ; la seconde a quant-à-elle mis en évidence la production d'un peptide cyclique secrété appelé Pep1357C sous la dépendance d'un système SHP/Rgg et du système de transport Ami.

L'ensemble de ces résultats a alors permis de suggérer un nouveau mécanisme de communication cellulaire de type QS impliquant (i) le produit du gène *shp* jouant le rôle de phéromone, (ii) le transporteur d'oligopeptides Ami assurant l'import de la phéromone à l'intérieur de la cellule et (iii) un régulateur transcriptionnel Rgg contrôlant l'expression d'un gène cible, *ster_1357*.

Mes objectifs de thèse consistaient alors :

1) A valider le modèle de QS proposé sur le locus *shp/ster_1358* de *S. thermophilus* LMD-9.

2) A étudier de manière globale les systèmes SHP/Rgg présents dans la quasi-totalité des streptocoques.

3) A identifier les cibles des systèmes SHP/Rgg afin de mieux appréhender les fonctions qu'ils contrôlent.

Résultats

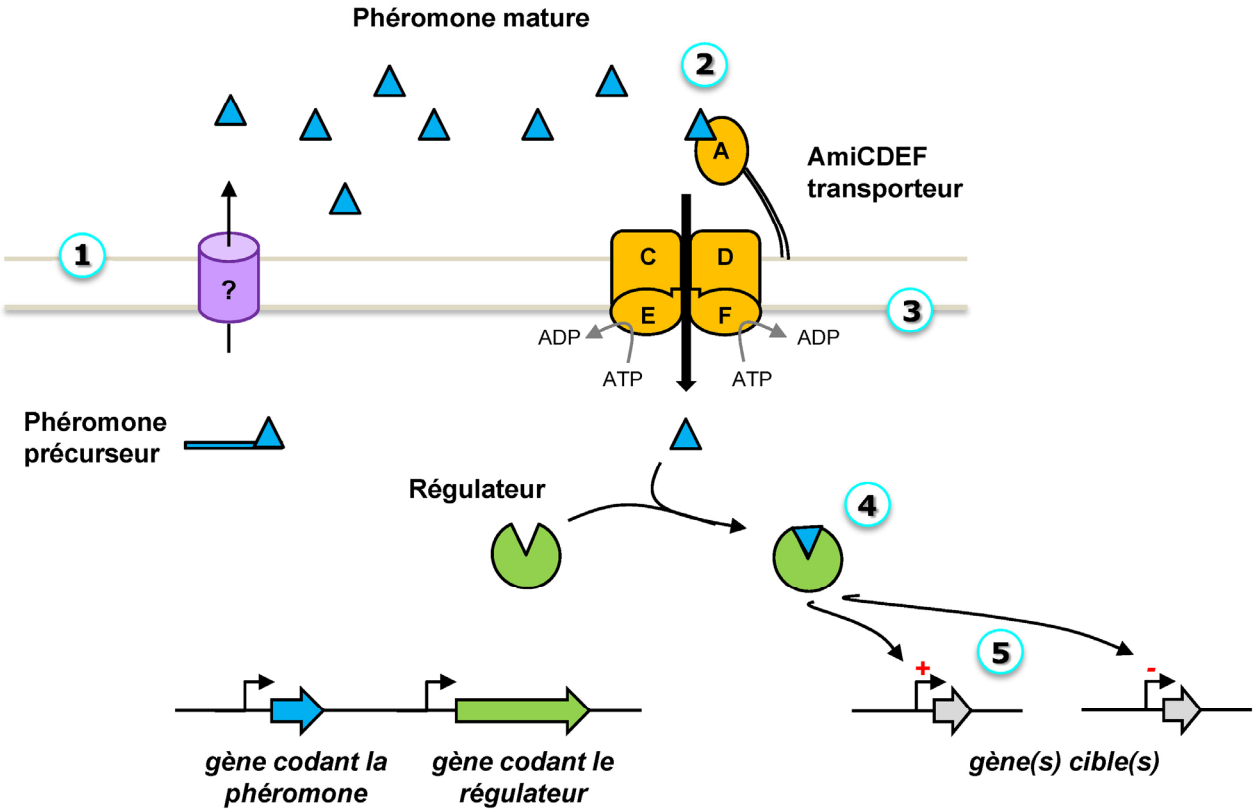


Figure 41 – Représentation schématique des différentes étapes intervenant dans un mécanisme de quorum-sensing.
(1) Sécrétion de la phéromone ; **(2)** Détection de la phéromone à une concentration seuil ; **(3)** Réimportation de la phéromone par le transporteur d'oligopeptides ; **(4)** Interaction de la phéromone avec le régulateur transcriptionnel ; **(5)** Fixation du régulateur transcriptionnel à l'ADN.

Chapitre 1 : Implication des protéines Rgg dans des phénomènes de quorum-sensing

Article I : « Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism is streptococci. »

1. Introduction et résultats préliminaires

En 1992, Sulavik et ses collaborateurs ont montré que l'expression du gène *gtfG*, codant une glucosyltransférase chez *S. gordonii*, était contrôlée positivement par le produit du gène adjacent et localisé en amont. En référence à la fonction du gène contrôlé, cet activateur a été appelé Rgg pour **R**egulator **g**ene of **g**lucosyltransferase (Sulavik *et al.*, 1992). Depuis, de nombreuses protéines similaires ont été identifiées et constituent aujourd'hui une famille de protéines régulatrices à part entière contrôlant l'expression de gène(s), proximaux et/ou distaux, impliqués dans un grand nombre de fonctions biologiques chez les bactéries à Gram positif (Bortoni *et al.*, 2009; Chaussee *et al.*, 2003; Dmitriev *et al.*, 2006; Fernandez *et al.*, 2006; Lyon *et al.*, 1998; Pulliainen *et al.*, 2008; Rawlinson *et al.*, 2002; Samen *et al.*, 2006; Sulavik *et al.*, 1992; Zheng *et al.*, 2011). A ce jour, les protéines Rgg sont principalement étudiées chez les streptocoques et malgré la connaissance de certains stimuli environnementaux nécessaires à l'expression de leurs cibles (pH, source de carbone par exemple), les mécanismes de transduction de ces signaux restent indéfinis.

En 2007, l'identification au laboratoire d'une nouvelle organisation génétique – spécifique des streptocoques – codant un petit peptide hydrophobe SHP et un régulateur transcriptionnel appartenant à la famille des Rgg a permis de suggérer un rôle de ces régulateurs dans des phénomènes de QS (Ibrahim *et al.*, 2007a; Ibrahim *et al.*, 2007b).
A mon arrivée au laboratoire, l'objectif était de démontrer l'implication de ces régulateurs dans un nouveau mécanisme de QS où le gène *shp* coderait un peptide phéromone qui serait maturé, sécrété puis réimporté dans la cellule (*via* un système de transport d'oligopeptides Ami) où il interagirait avec le régulateur Rgg permettant ainsi la régulation de gène(s) cible(s). Des travaux permettant de conclure sur la fonctionnalité des systèmes SHP/Rgg dans des phénomènes de QS chez les streptocoques ont alors été entrepris sur la base du locus *shp/rgg1358* de *S. thermophilus* LMD-9, locus contrôlant l'expression d'un gène cible *ster_1357* codant un peptide cyclique secrété appelé Pep1357C (Ibrahim *et al.*,

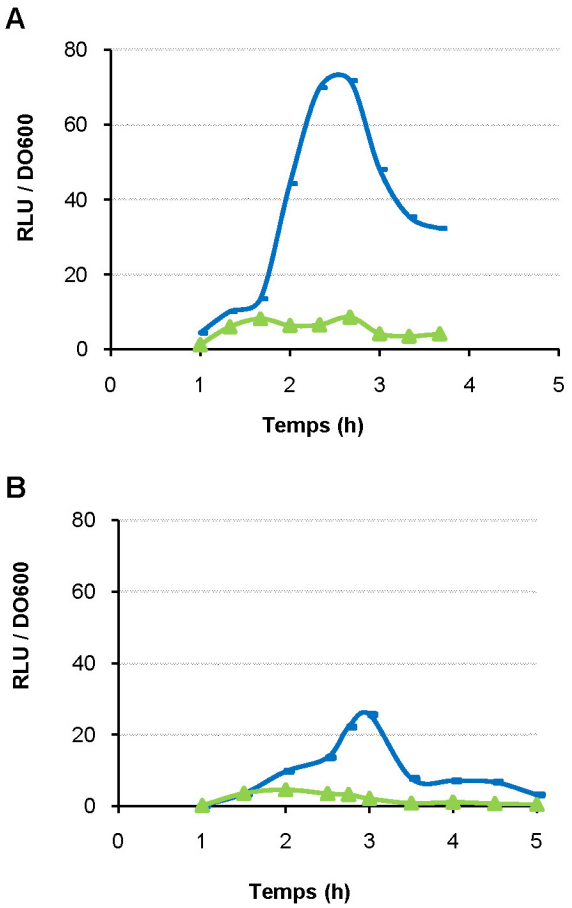


Figure 42 – Comparaison du niveau d'expression des promoteurs des gènes *shp1358* et *ster_1357*.
Bioluminescence (RLU/DO_{600}, Relative Light Unit) émise par les fusions *blp*::P*shp1358-luxAB* **(A)** et *blp*::P*ster_1357-luxAB* **(B)** dans un contexte sauvage (—) et mutant (Δ*shp*) (▲) de *Streptococcus thermophilus* LMD-9.

2007a). La stratégie utilisée a consisté en la caractérisation de l'ensemble des étapes décrivant un mécanisme de QS, à savoir : la sécrétion, la détection à une concentration seuil et la réimportation de la phéromone à l'intérieur de la cellule par le transporteur d'oligopeptides AmiCDEF, son interaction avec la protéine régulatrice Rgg et enfin la fixation de cette dernière à l'ADN (Figure 41).

Afin de valider les quatre premières étapes du mécanisme (Figure 41), une approche génétique a tout d'abord été développée. Cette approche met en œuvre l'utilisation de fusions transcriptionnelles couplant le promoteur d'un gène d'intérêt (cible du système) au système rapporteur *luxAB* d'une bactérie luminescente. Ces fusions sont intégrées par double recombinaison homologue au locus *blp* de différents mutants de *S. thermophilus* LMD-9 grâce à l'utilisation d'un plasmide à origine de réplication thermosensible (pGICB004). Elles permettent de suivre par bioluminescence l'activité d'un promoteur sous des conditions précises.
Aussi, l'étude de la régulation de l'expression du gène *shp* du locus *shp/rgg1358* de *S. thermophilus* LMD-9 ayant permis d'identifier ce dernier en tant que cible du système, deux gènes cibles étaient alors disponibles pour cette étude (*shp1358* et *ster_1357*). Le choix du promoteur d'intérêt à utiliser a donc été soulevé. L'activité des promoteurs des gènes *shp1358* et *ster_1357* a alors été mesurée dans un contexte sauvage et mutant (Δ*shp*) (Figure 42). Les résultats obtenus ont tout d'abord confirmé que l'activité des deux promoteurs était abolie dans un mutant Δ*shp*. De plus, l'activité de la fusion avec le promoteur $P_{shp1358}$ présentait un niveau de bioluminescence trois fois plus important que celui détecté avec le promoteur P_{ster_1357}. Nous avons donc choisi d'utiliser la fusion avec le promoteur $P_{shp1358}$ pour valider le système de régulation décrit sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9.
Dans un second temps, les techniques de SPR et de retard sur gel ont été utilisées afin d'étudier les deux dernières étapes du mécanisme.

En parallèle à ce travail expérimental, un arbre phylogénétique des protéines Rgg et Rgg-like chez les bactéries à Gram positif, a également été construit dans le but (i) de mettre à jour la liste des loci *shp/rgg* présents dans les génomes séquencés (en octobre 2010), (ii) de conforter leur spécificité au groupe des streptocoques et, (iii) de différencier deux systèmes de QS décrit chez les streptocoques (SHP/Rgg et XIP/ComR) impliquant des petits peptides hydrophobes associés à des protéines régulatrices Rgg ou Rgg-like.

L'ensemble des résultats a fait l'objet d'une publication et est présenté dans l'article 1.

2. Article 1

Title: Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci.

Running title: SHP/Rgg quorum-sensing mechanism in streptococci.

Fleuchot B.[1,2], Gitton C.[1,2], Guillot A.[1,2], Vidic J.[3], Nicolas P.[4], Besset C.[1,2], Fontaine L.[5], Hols P.[5], Leblond-Bourget N.[6], Monnet V.[1,2], Gardan R.[1,2*]

[1]INRA, UMR1319 MICALIS, F-78352 Jouy en Josas, France.
[2]AgroParistech, UMR MICALIS, F-78352 Jouy en Josas, France.
[3]INRA, UR892 Virologie Immunologie Moléculaire, F-78352 Jouy en Josas, France.
[4]INRA, UR1077 Mathématique, Informatique et Génome, F-78352 Jouy en Josas, France.
[5]Université catholique de Louvain, Unité de Génétique, B-1348 Louvain-la-Neuve, Belgium.
[6]INRA, UMR1128, IFR110, Nancy-Université, Laboratoire de Génétique et Microbiologie, BP239, 54506 Vandœuvre-lès-Nancy, France.

* Corresponding author:
E-mail: rozenn.gardan@jouy.inra.fr / Tel: 33-1-34-65-27-70 / Fax: 33-1-34-65-21-63.

Key words: streptococci, quorum-sensing, RNPP family, Rgg, ComR, hydrophobic peptide, oligopeptide transporter.

Supplementary data:
Annexe 1: Table S1 - Rgg-like proteins associated or not with SHP or XIP hydrophobic peptides in Gram positive bacteria.
Annexe 2: Fig. S1 – The triggering of the expression of the shp gene is a consequence of the presence of the $SHP1358_{(15-23)}$ pheromone in the culture medium and not of the transition from the exponential growth phase to the stationary phase.

Article publié dans Molecular Microbiology en 2011, volume 80, pages 1102-1119.

Molecular Microbiology (2011) 80(4), 1102–1119 ■
doi:10.1111/j.1365-2958.2011.07633.x
First published online 30 March 2011

Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci

B. Fleuchot,[1,2] C. Gitton,[1,2] A. Guillot,[1,2] J. Vidic,[3] P. Nicolas,[4] C. Besset,[1,2] L. Fontaine,[5] P. Hols,[5] N. Leblond-Bourget,[6] V. Monnet[1,2] and R. Gardan[1,2]*

[1]*INRA, UMR1319 MICALIS, F-78352 Jouy en Josas, France.*
[2]*AgroParistech, UMR MICALIS, F-78352 Jouy en Josas, France.*
[3]*INRA, UR892 Virologie Immunologie Moléculaire, F-78352 Jouy en Josas, France.*
[4]*INRA, UR1077 Mathématique, Informatique et Génome, F-78352 Jouy en Josas, France.*
[5]*Université catholique de Louvain, Unité de Biochimie et Génétique Moléculaire Bactérienne, B-1348 Louvain-la-Neuve, Belgium.*
[6]*INRA, UMR1128, IFR110, Nancy-Université, Laboratoire de Génétique et Microbiologie, BP239, 54506 Vandœuvre-lès-Nancy, France.*

Summary

We identified a genetic context encoding a transcriptional regulator of the Rgg family and a small hydrophobic peptide (SHP) in nearly all streptococci and suggested that it may be involved in a new quorum-sensing mechanism, with SHP playing the role of a pheromone. Here, we provide further support for this hypothesis by constructing a phylogenetic tree of the Rgg and Rgg-like proteins from Gram-positive bacteria and by studying the *shp/rgg1358* locus of *Streptococcus thermophilus* LMD-9. We identified the *shp1358* gene as a target of Rgg1358, and used it to confirm the existence of the steps of a quorum-sensing mechanism including secretion, maturation and reimportation of the pheromone into the cell. We used surface plasmon resonance to demonstrate interaction between the pheromone and the regulatory protein and performed electrophoretic mobility shift assays to assess binding of the transcriptional regulator to the promoter regions of its target genes. The active form of the pheromone was identified by mass spectrometry. Our findings demonstrate that the *shp/rgg1358* locus encodes two components of a novel quorum-sensing mechanism involving a transcriptional regulator of the Rgg family and a SHP pheromone that is detected and reimported into the cell by the Ami oligopeptide transporter.

Accepted 10 March, 2011. *For correspondence. E-mail rozenn.gardan@jouy.inra.fr; Tel. (+33) 1 34 65 27 70; Fax (+33) 1 34 65 21 63.

© 2011 Blackwell Publishing Ltd

Introduction

Quorum-sensing (QS) is a cell–cell communication mechanism in bacteria that controls gene expression via secreted signalling molecules, also called autoinducers or pheromones. Despite differences in the chemical nature of the signalling molecules between Gram-positive and Gram-negative bacteria (Bassler, 2002; Waters and Bassler, 2005; Antunes *et al*., 2010), QS is triggered by a similar circuit in the two groups. First, the signalling molecule is released into the extracellular environment by active or passive transport; it accumulates in the extracellular environment and once at a threshold concentration is detected by a sensor protein. This sensing leads cells to modulate gene expression in a co-ordinated manner in the bacterial population.

There are two general activation pathways for Gram-positive QS (Williams *et al*., 2007). The signal molecule, which is a peptide, can be sensed outside the cell by the histidine kinase of a two component system. Detection of the peptide leads to phosphorylation or dephosphorylation of the histidine kinase and then of a specific cytoplasmic transcriptional regulator. This phosphorylation state modifies the ability of the response regulator to bind DNA, modulating its ability to activate or repress the transcription of its target gene(s). This mechanism has been described in detail for many systems, including triggering of competence for natural transformation in *Streptococcus pneumoniae* and *Bacillus subtilis* (for reviews see Claverys and Håvarstein, 2002; Claverys *et al*., 2006) and control of the *agr* locus in *Staphylococcus aureus* inducing the regulation of accessory virulence genes (Novick and Geisinger, 2008). The second activation pathway involves sensing the signalling molecule inside the cell after its internalization by an oligopeptide permease transport system called Opp or Ami, member of the ubiquitous ATP-binding cassette superfamily (ABC transporters) (Linton and Higgins, 2007). Once internalized, the pheromone interacts with a transcriptional regulator or a Rap protein, both belonging to the RNPP family (for Rap,

NprR, PlcR and PrgX), thereby modifying their activity and consequently the expression of their target gene(s) (Declerck *et al*., 2007; Rocha-Estrada *et al*., 2010). Three groups of relevant peptides have been described in detail: (i) Rap-associated Phr peptides in *B. subtilis* involved in the control of sporulation, competence, and production of degradative enzymes and antibiotics (Pottathil and Lazazzera, 2003), (ii) peptides involved in the control of plasmid transfer in *Enterococcus faecalis* that interact with PrgX or PrgX-like proteins (Dunny, 2007) and (iii) PlcR-associated PapR peptides involved in triggering virulence of *cereus* group bacteria (Slamti and Lereclus, 2002; Bouillaut *et al*., 2008).

Rgg proteins are described as stand-alone transcriptional regulators in low-GC Gram-positive bacteria. They are composed of a N-terminal Helix–Turn–Helix DNA-binding domain belonging to the xenobiotic regulatory element family (PFAM01381) and a conserved C-terminal domain (Rgg-Cterm, TIGR01716). They are involved in several physiological functions such as: (i) expression of glucosyltransferases in *Streptococcus gordonii* (Sulavik *et al*., 1992; Vickerman and Minick, 2002) and in *Streptococcus oralis* (Fujiwara *et al*., 2000), (ii) regulation of the gene encoding the secreted cysteine proteinase virulence factor (SpeB) in *Streptococcus pyogenes* (Lyon *et al*., 1998; Chaussee *et al*., 1999; Neely *et al*., 2003; Dmitriev *et al*., 2006; Loughman and Caparon, 2007), (iii) the stress response in *Lactococcus lactis* (Sanders *et al*., 1998), *Streptococcus thermophilus* (Fernandez *et al*., 2006), *S. pyogenes* (Pulliainen *et al*., 2008) or *S. pneumoniae* (Bortoni *et al*., 2009), (iv) bacteriocin production in *Streptococcus mutans* (Qi *et al*., 1999) and *Lactobacillus sakei* (Rawlinson *et al*., 2002; Skaugen *et al*., 2002) and (v) regulation of genes involved in the pathogenicity of *Streptococcus agalactiae* (Chaussee *et al*., 2002; 2003; Samen *et al*., 2006). Although some Rgg proteins seem to be associated with a single target gene, they may also serve as global regulators in some organisms (Chaussee *et al*., 2002; Samen *et al*., 2006; Zheng *et al*., 2011). In general, the regulatory mechanisms of gene transcription by stand-alone Rgg proteins are complex. Although the environmental stimuli for the expression of some targets of Rgg regulators are known, the signal transduction processes have still to be discovered. Furthermore, most of the Rgg proteins annotated in genomes have not been studied.

We previously identified, by *in silico* analyses, a subcluster of Rgg regulators defined by the association of pairs of genes transcribed divergently. These pairs of genes, widespread only in the *Streptococcus* genus, combine a gene encoding a transcriptional regulator of the Rgg family with another gene encoding a small hydrophobic peptide called *shp*. One or two copies of such genetic contexts have been found in nearly all streptococci genomes, including *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*, and up to seven copies have been found in *S. thermophilus* (Ibrahim *et al*., 2007a). We have studied one of these *shp/rgg* loci, named *shp/rgg1358*, in *S. thermophilus* LMD-9 and identified one target gene of the encoded Rgg regulator. This target gene, called *ster_1357*, encodes a secreted cyclic peptide and its expression is positively controlled by Rgg1358 and indirectly by SHP1358 and the Ami oligopeptide transporter (Ibrahim *et al*., 2007b). These findings led us to investigate whether the Rgg protein and a small hydrophobic peptide (SHP) reimported pheromone are components of a novel QS regulatory system. Recently, another group of signalling peptides, associated with Rgg-like proteins named ComR, has been identified. These peptides are named ComS or XIP and are involved in triggering competence for transformation in *S. thermophilus*, *S. salivarius* and *S. mutans* and probably in other streptococci belonging to the pyogenic and bovis group (Gardan *et al*., 2009; Fontaine *et al*., 2010; Mashburn-Warren *et al*., 2010).

In this study, we started by constructing a phylogenetic tree of the Rgg and Rgg-like proteins found in Gram-positive bacteria to compare the respective distributions of SHP-associated Rgg with the XIP-associated ComR regulators. Then, we tested the function of the *shp*/*rgg* subcluster as a QS mechanism by studying the *shp/rgg1358* locus of *S. thermophilus* LMD-9. We identified another Rgg1358 target gene (*shp1358*) and used it to confirm the existence of the first steps of a QS mechanism: (i) secretion of the putative SHP pheromone, (ii) maturation of the pheromone, (iii) detection of the pheromone at a cell density threshold and (iv) reimportation of the pheromone into the cell by the Ami oligopeptide transporter. Surface plasmon resonance (SPR) analyses demonstrated that the pheromone interacts with the transcriptional regulator Rgg1358. Electrophoretic mobility shift assays (EMSA) revealed that the Rgg1358 binds to the promoter regions of target genes. Thus, this comprehensive survey of SHP-associated regulators combined with the study of the *shp/rgg1358* locus of *S. thermophilus* LMD-9 provides evidence of a novel QS mechanism: it involves a Rgg transcriptional regulator associated with a SHP pheromone detected and reimported into the cell by the Ami oligopeptide transporter.

Results

SHP-associated Rgg are phylogenetically distinct from the XIP-associated ComR regulators

Using genome sequences for 19 streptococci, we previously identified the SHP-associated Rgg as a subfamily of Rgg regulators (Ibrahim *et al*., 2007a). Since then,

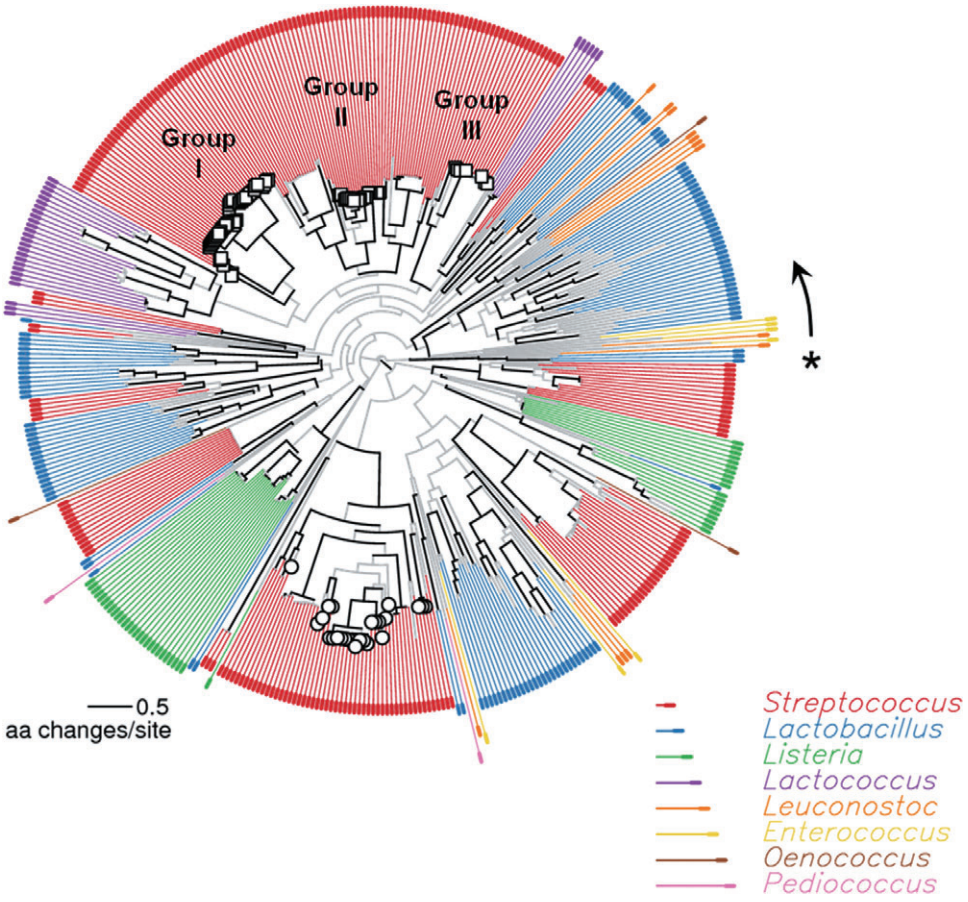


Fig. 1. Distribution of SHP/Rgg and XIP/ComR systems in the repertoire of Rgg and Rgg-like proteins. The evolutionary relationships between the protein sequences of Rgg and Rgg-like proteins are summarized in the phylogenetic tree shown at the centre of the figure. Each leaf corresponds to a Rgg or Rgg-like protein and the presence of SHP (□) and XIP (○) peptides at the corresponding locus is indicated. Internal branches supported by at least 80% of the bootstrap replicates are represented in black, other branches are shown in gray. Genera in which the loci are found are indicated using a different colour and a different radius length for each genus according to the correspondence code reported in the figure. The asterisk and the arrow indicate the first locus and the direction ordering all the loci in Table S1, where details are provided.

numerous genomes have been deposited in the databases. Moreover, another subfamily of Rgg-like regulators, ComR, also working with a peptide, ComS or XIP, imported by an oligopeptide transporter has been discovered (Fontaine *et al*., 2010; Mashburn-Warren *et al*., 2010). We therefore updated our list of SHP-associated Rgg and investigated their place in the repertoire of the Rgg and Rgg-like proteins. In particular we compared them with the XIP-associated ComR proteins.

Rgg proteins are only found in the order of *Lactobacillales* and the family *Listeriaceae* for which 90 complete genome sequences are available. We identified 484 sequences corresponding to Rgg proteins *sensu largo*, including all XIP-associated ComR proteins previously identified (Fontaine *et al*., 2010; Mashburn-Warren *et al*., 2010). Rgg regulators were found in all the genomes considered except the two *Lactobacillus salivarius* isolates. To map SHP and XIP peptides, the GenBank CDS annotation was complemented by predicting short CDSs with our software SHOW as described by Ibrahim *et al*. (2007a). Putative CDSs with length between 10 and 50 aa and adjacent to the 484 *rgg* genes were retrieved and screened manually to identify exhaustively all SHP and XIP peptides. The sequence of the 12 SHP (Ibrahim *et al*., 2007a) and 10 XIP (Mashburn-Warren *et al*., 2010) identified in complete genome sequences and already published were used to define the selection criteria as described in the *Experimental procedures* section. We recovered 61 SHP and 27 XIP. As a result of some redundancy between strains, only 22 and 12 unique amino acid sequences of SHP and XIP, respectively, were identified; they included 10 new SHP sequences and two new XIP.

The 484 sequences were further used to construct a phylogenetic tree in which SHP-associated Rgg and XIP-associated ComR were highlighted (Fig. 1). The complete list of Rgg and Rgg-like proteins and the sequences and

positions of the SHP and XIP peptides is provided in Table S1. This analysis confirmed that both SHP-associated Rgg and XIP-associated ComR are specific to the streptococci family. Although they share similarities, they form two different branches of the phylogenetic tree indicating that they correspond to two distinct subfamilies of Rgg regulators. Furthermore, the tree distinguished between the two groups of SHP-associated Rgg previously identified, i.e. groups I and II, which again correlated well with the presence of a conserved aspartate or glutamate respectively (Ibrahim *et al*., 2007a). We found no exception in group I and only one SHP (in six strains of *Streptococcus suis*) with a conserved aspartate associated with a Rgg belonging to group II. In this new phylogenetic tree, the two SHP previously described as unclassified because of the absence of conserved aspartate or glutamate residue clustered with group I Rgg proteins. Finally, detailed examination of the Rgg proteins in close vicinity to the SHP-associated Rgg revealed another group (group III) of seven peptides with amino acid sequences that fulfil the SHP definition but that are encoded by genes overlapping the end of the *rgg* genes in a convergent orientation. One is in *S. thermophilus* strain LMD-9 and the others are from six (of nine studied) strains of *S. pneumoniae*.

To confirm the function of the *shp*/*rgg* subcluster as a component of a QS mechanism, we investigated the *shp/rgg1358* locus of *S. thermophilus* LMD-9.

The expression of the shp1358 *gene is controlled by Rgg1358, SHP1358 and, AmiCDEF in* S. thermophilus *LMD-9*

The *shp/rgg1358* locus of *S. thermophilus* LMD-9 is composed of *shp1358* and *rgg1358*, two genes divergently transcribed with overlapping putative promoters. We studied the expression of the *shp1358* gene suspected to encode a pheromone (Ibrahim *et al*., 2007b). The production of most such pheromones is auto-regulated, so we constructed a $P_{shp1358}$-*luxAB* transcriptional fusion and used it to transform the LMD-9 strain and several mutants deleted for the *rgg1358*, *shp1358* or *ami* genes. The luciferase activity of the $P_{shp1358}$-*luxAB* fusion was 80 RLU/OD_{600} in the LMD-9 strain but only 6 and 3 RLU/OD_{600} in the *Δrgg1358* and *Δshp1358* mutants respectively; no activity was detected in the *amiCDE::spec* mutant (Fig. 2A). Therefore, the *shp1358* gene is a target of the Rgg1358 regulator, and the *shp1358* gene product is implicated in its own transcription. The Ami transporter is also involved in the positive control of the transcription of the *shp1358* gene, most probably by an indirect mechanism. The $P_{shp1358}$-*luxAB* fusion was then used to investigate all steps of the QS mechanism involving SHP1358 and Rgg1358.

The small hydrophobic peptide SHP1358 is the precursor of a secreted pheromone of nine amino acids

In our QS model, SHP1358 is proposed to be the precursor of the pheromone. We tested whether, like other pheromones in these systems, SHP1358 is secreted. Supernatants from cultures of strains producing or not producing the SHP1358 peptide (strain LMD-9 and its *Δshp1358* mutant, respectively) were mixed with pellets of cells of a strain containing a $P_{shp1358}$-*luxAB* fusion and deleted for the *shp1358* gene (TIL 1200, named the 'reporter strain' hereafter). The expression of the *shp1358* promoter was then tested (Fig. 2B). The expression of the *shp1358* promoter in the reporter strain was not activated by the supernatant from the *Δshp1358* mutant, whereas the supernatant from strain LMD-9 induced its activity (55 RLU/OD_{600}). This demonstrates functional complementation of the cells of the reporter strain by the supernatant of the LMD-9 strain possibly through secretion of the product of the *shp1358* gene.

To determine the amino acid sequence of the active form of the secreted pheromone, the supernatant from strain LMD-9 was analysed by mass spectrometry (LC-MS/MS) and compared with that from strain *Δshp1358* (data not shown). Only one mass, not detected in the supernatant of strain *Δshp1358*, was identical to the mass of a fragment of SHP1358. This mass corresponds to a highly hydrophobic nonapeptide with the amino acid sequence EGIIVIVVG. This form is the putative product of a C-terminal cleavage of the 23 amino acid peptide precursor. This sequence was validated by fragmentation followed by an analysis on the LTQ orbitrap (Fig. 3).

A synthetic peptide corresponding to this nine amino acid sequence was produced and added at different concentrations to cultures of the reporter strain. Functional complementation by the synthetic peptide was observed and was correlated with its concentration in the medium (Fig. 2C). These results demonstrate and identify one active secreted peptide, hereafter called $SHP1358_{(15–23)}$, derived from the SHP1358 precursor.

A critical biomass is required to activate the expression of the shp1358 *gene*

Quorum-sensing mechanisms involve signalling molecules reaching a threshold concentration to trigger the expression of their target gene(s) (Winzer *et al*., 2002; Podbielski and Kreikemeyer, 2004; Siehnel *et al*., 2010), and signalling molecule concentration correlates with cell density. To check whether this applies to the SHP/Rgg1358 system, the expression of the *shp1358* promoter was measured in cultures of strain LMD-9 containing the $P_{shp1358}$-*luxAB* fusion inoculated at different concentrations (Fig. 2D). In all cultures, irrespective of

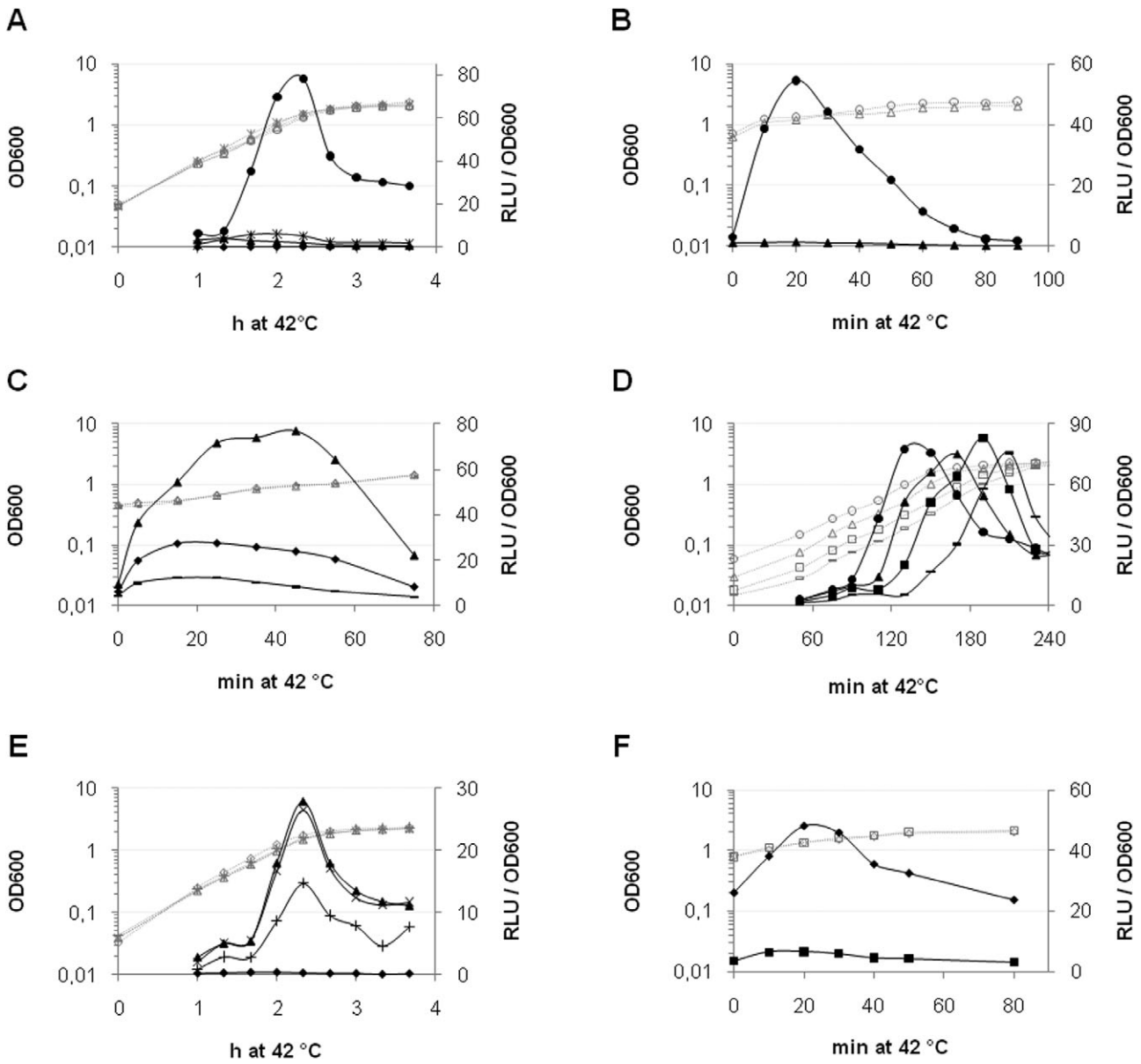


Fig. 2. Growth and luciferase activities of strains containing a $P_{shp1358}$-*luxAB* fusion in various genetic background and growth conditions. Growth curves (OD_{600}) are presented in gray and luciferase activities (RLU/OD_{600}) in black. Growth and luciferase activity of strains containing the *shp1358-luxAB* gene fusion: A, in the LMD-9 (●), *Δshp1358* (▲), *Δrgg1358* (✳) and *amiCDE::spec* (◆) genetic background; B, in the *Δshp1358* genetic background after adding the supernatant from cultures of strain LMD-9 at OD_{600} 0.6 (●) or the *Δshp1358* mutant (▲); C, in the *Δshp1358* genetic background after adding the synthetic $SHP1358_{(15–23)}$ pheromone at various concentrations: 25 ng ml^{-1} (-), 250 ng ml^{-1} (◆) and 2500 ng ml^{-1} (▲); D, in the LMD-9 background inoculated at various concentrations: OD_{600} 0.012 (-), OD_{600} 0.025 (■), OD_{600} 0.05 (▲) and OD_{600} 0.1 (●); E, in the *Δshp1358* (▲), *ΔamiA1 Δshp1358* (✖), *amiA3::erm Δshp1358* (✚) and *ΔamiA1 amiA3::erm Δshp1358* (◆) genetic background and in cocultures with strain LMD-9 serving as a SHP1358 supplier; F, in the *Δshp1358* genetic background after adding the supernatant from cultures of the *Δeep* pBV5030::P32-*shp1358* (◆) or the *Δopp* pBV5030::P32-*shp1358* (◆) mutant. Data shown are representative of three independent experiments.

initial cell density, a similar maximum of luciferase activity was observed (approximately 80 RLU/OD_{600}). However, the higher the initial cell density, the earlier the expression of the luciferase activity was triggered and its maximum reached.

To check that this effect was the result of the presence of the $SHP1358_{(15–23)}$ pheromone at a sufficient concentration, and not of the transition from exponential growth to stationary phase, we inoculated media with strain LMD-9 containing the $P_{shp1358}$-*luxAB* fusion at a low cell density and added or did not add synthetic $SHP1358_{(15–23)}$ at the beginning of the exponential phase. The addition of $SHP1358_{(15–23)}$ was sufficient to trigger the expression of the *shp1358* gene, independent of the growth phase of

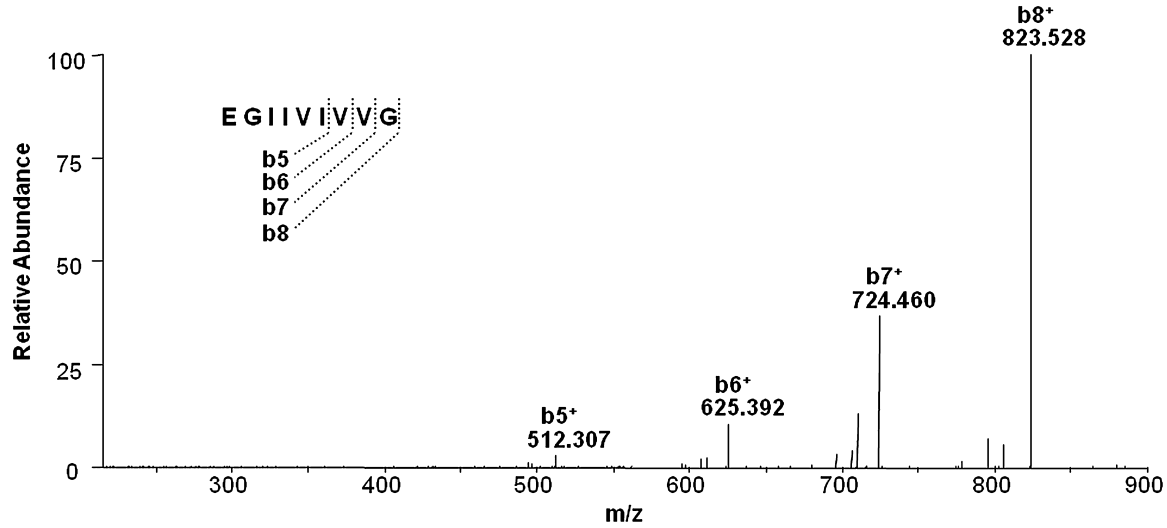


Fig. 3. The product of the *shp1358* gene is the precursor of a 9 amino acid secreted peptide. Fragmentation spectrum of the ion m/z 898.56, leading to the validation of the sequence of the putative mature form of the SHP1358 pheromone detected in the LMD-9 supernatant.

the culture (Fig. S1). These findings are consistent with a QS mechanism.

The lipoprotein AmiA3 detects the SHP1358$_{(15–23)}$ pheromone before its reimportation by the oligopeptide transporter, AmiCDEF

Oligopeptide permease transporters can be divided in two functional entities: a detection entity composed of oligopeptide binding proteins allowing the capture of the target substrate (AmiA1 and AmiA3 in the strain LMD-9) and a transport entity allowing the import of the substrate into the cell (AmiCDEF in the strain LMD-9). The expression of the *shp1358* gene is abolished in an *amiCDE::spec* mutant and also in a *ΔamiA1 amiA3::erm* mutant (data not shown). Therefore, it is plausible that the Ami transporter is involved in the reimportation of the SHP1358$_{(15–23)}$ pheromone into the cell. To confirm this, we introduced the P$_{shp1358}$-*luxAB* transcriptional fusion into various *Δami* mutants, also deleted for the *shp1358* gene: *ΔamiA1 amiA3::erm Δshp1358 blp*::P$_{shp1358}$-*luxAB* to confirm the role of the Ami transporter in the reimportation of the pheromone, *ΔamiA1 Δshp1358 blp*::P$_{shp1358}$-*luxAB* and *amiA3::erm Δshp1358 blp*::P$_{shp1358}$-*luxAB* to study, respectively, the activity of the AmiA3 and AmiA1 oligopeptide binding proteins. The reporter strain, TIL 1200, was also used as a positive control of luciferase activity. These four strains were then cocultivated with strain LMD-9, providing SHP1358$_{(15–23)}$ to the medium, and luciferase activity was followed (Fig. 2E). The luciferase activity of the positive coculture control increased to 28 RLU/OD$_{600}$ whereas there was no detectable activity in the LMD-9/*ΔamiA1 amiA3::erm Δshp1358 blp*::P$_{shp1358}$-*luxAB* coculture. Therefore, the SHP1358$_{(15–23)}$ pheromone provided by the LMD-9 strain was not sensed or imported by the cells lacking the oligopeptide binding proteins of the Ami transport system, but was by the *ami* wild type. Two different levels of luciferase activity were obtained for the cocultures of LMD-9 with single oligopeptide binding protein mutants. The luciferase activity of the LMD-9/*ΔamiA1 Δshp1358 blp*::P$_{shp1358}$-*luxAB* coculture was similar to that of the LMD-9/reporter strain coculture. The activity of the LMD-9/*amiA3::erm Δshp1358 blp*::P$_{shp1358}$-*luxAB* coculture was half that of the positive control coculture. Therefore, detection of the SHP1358$_{(15–23)}$ pheromone mainly requires the AmiA3 oligopeptide binding protein.

The membrane peptidase Eep is involved in the SHP1358 maturation process

The Eep protease plays a key role in the maturation of various pheromones and anti-pheromones of *E. faecalis* (An *et al.*, 1999) and *S. gordonii* (Vickerman *et al.*, 2010). The presence of an *eep*-like gene in the *S. thermophilus* LMD-9 genome and some similarities between the anti-pheromone of *E. faecalis* and the SHP1358 pheromone (sequences of approximately 23 amino acids, presence of lysine in the N-terminal part and high hydrophobicity) led us to investigate the involvement of this protease in the SHP1358 maturation process. Eep proteases also contribute to the cleavage of the peptide signal of some lipoproteins (An and Clewell, 2002; Denham *et al.*, 2008). As the oligopeptide binding proteins of the Ami transporter are lipoproteins, we hypothesized that they may be matured by the Eep protease and therefore tested their functionality in a *Δeep* mutant. The toxicity of the aminopterin peptide for the *Δeep* strain was compared with that for the wild-type strain and the *ΔamiCDE* strain. The tox-

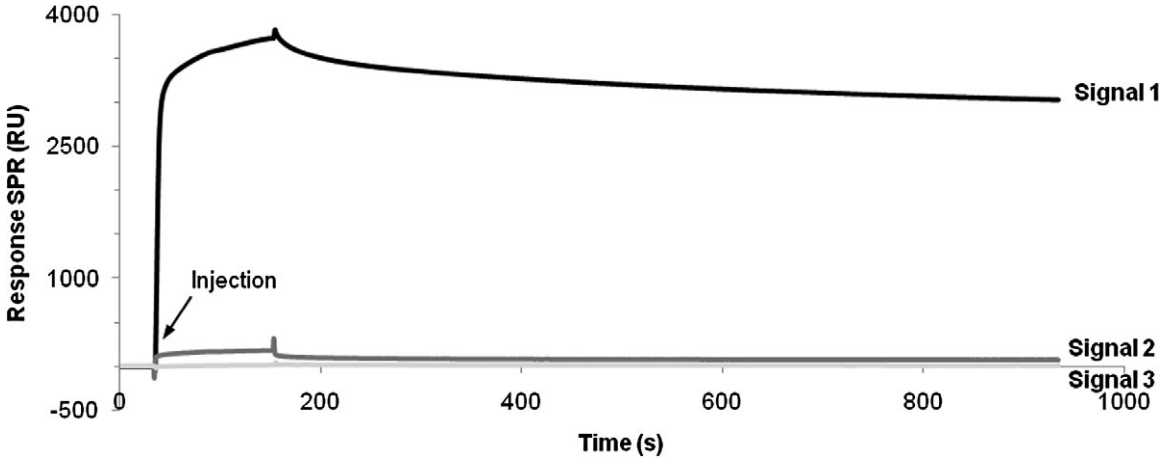


Fig. 4. Interaction profile between the His-tagged Rgg1358 protein and the SHP1358$_{(15–23)}$ pheromone. Sensorgram of immobilized SHP1358$_{(15–23)}$ pheromone and the regulatory protein Rgg1358 (signal 1) obtained after injection of 5 μM of the purified His-tagged Rgg1358 protein over 2 min, at a flow of 25 μl min^{-1} in 20 mM HBS. Two controls were performed in the same conditions to test the interaction of the purified His-tagged Rgg1358 protein with immobilized ComS peptide of *S. thermophilus* (signal 2), and the interaction of bovine serum albumin with the immobilized SHP1358$_{(15–23)}$ pheromone (signal 3). The sensorgrams observed were corrected by subtracting the results with the reference surface (ethanolamine).

icity of the aminopterin peptide for the *Δeep* strain was intermediate between that for strain LMD-9 (sensitive) and the *ami* mutant (resistant) (data not shown), suggesting that the functionality of the Ami transporter is probably partially impaired in the *Δeep* mutant. Consequently, a *ΔamiCDE* mutant, rather than the wild-type strain, was used as a positive control for the following experiment. A replicative plasmid containing the *shp1358* gene under the control of a constitutive promoter was introduced into the *Δeep* and *ΔamiCDE* strains, leading to strains TIL 1206 and TIL 1213 respectively. The effects of culture supernatants from both these strains were then tested on reporter strain (TIL 1200) cell pellets (Fig. 2F). The luciferase activity of the reporter strain was tenfold lower with supernatant from TIL 1206 than with supernatant from TIL 1213, indicating that the extracellular concentration of the pheromone is lower in TIL 1206 cultures than in TIL 1213 cultures. The endopeptidase Eep is therefore involved in the SHP1358 maturation process.

The SHP1358$_{(15–23)}$ pheromone interacts with the Rgg1358 regulatory protein

We used SPR analyses to investigate whether the SHP1358$_{(15–23)}$ pheromone binds the regulatory protein Rgg1358. The interaction between chip-immobilized SHP1358$_{(15–23)}$ (898 Da) and purified His-tagged Rgg1358 protein (at 5 μM) gave a high SPR signal response corresponding to approximately 3700 RU (Fig. 4, signal 1), under the flow condition tested. Moreover, very little dissociation was observed, suggesting a high affinity between these two partners. In contrast, when purified Rgg1358 protein was injected over another immobilized small hydrophobic peptide (IAILPYFAGCL, ComS of *S. thermophilus* LMD-9) in similar conditions, no signal response was obtained (Fig. 4, signal 2). The specificity of the binding was confirmed by replacing the regulatory protein Rgg1358 by BSA (Fig. 4, signal 3). These SPR analyses demonstrate that SHP1358$_{(15–23)}$ and Rgg1358 specifically interact with each other.

Rgg1358 binds the promoter regions of shp1358 *and* ster_1357 *genes*

Electrophoretic mobility shift assays were carried out to test for interactions between the Rgg1358 regulatory protein and DNA fragments of promoter regions of the two known target genes, *shp1358* and *ster_1357*. To design relevant DNA probes, we mapped the transcriptional start site of both genes and identified putative –10 promoter sequences (Fig. 5A). A conserved motif – GCAWATATGGGAATA – of 15 nt was found 25 nt upstream from the –10 motif of the *shp1358* and *ster_1357* genes. This motif differs from that described for PlcR, PrgX and ComR regulators (Fig. 5B). DNA probes centred on the conserved motif were amplified by PCR. Nine pg of each labelled probe was mixed with purified recombinant His-tagged Rgg1358 protein and analysed by EMSA. The recombinant Rgg1358 protein bound specifically to the *shp1358* and *ster_1357* probes and the intensity of the shift increased with increasing concentration of the Rgg1358 (data not shown). In parallel, the binding to these two probes of the Rgg1358 protein mixed with the SHP1358$_{(15–23)}$ pheromone was also tested. In this case,

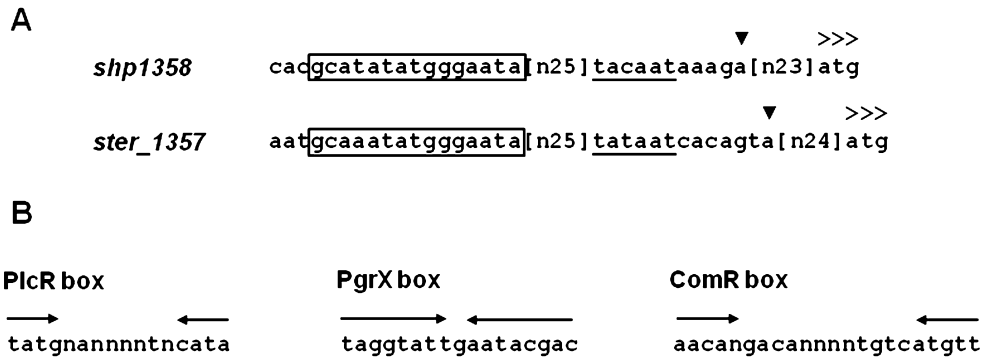


Fig. 5. Comparison of the DNA-binding site of the RNPP and RNPP-like proteins.
A, Nucleotide sequences of promoter regions of the genes *shp1358* and *ster_1357* of *S. thermophilus* strain LMD-9. Transcriptional start sites are indicated with ▼ symbol and the start codon with >>>. The putative binding site for the Rgg1358 protein is boxed and putative −10 sequences are underlined. B, DNA-binding sites for PlcR (Lereclus *et al.*, 1996; Agaisse *et al.*, 1999), PrgX (Bae *et al.*, 2002) and ComR (Mashburn-Warren *et al.*, 2010) transcriptional regulators.

to prevent peptide-dependent binding being missed because of a possible excess of Rgg1358 protein, an intermediate Rgg1358 protein concentration (at 70 nM) was used; this concentration did not allow a complete shift of the *shp1358* and *ster_1357* probes (Fig. 6A, lanes 1 and 2). The experiments were performed as described above in the presence of 27.8 nM, 278 nM and 2.78 μM, of the synthetic $SHP1358_{(15–23)}$ pheromone (corresponding to 25, 250 and 2500 ng ml^{-1}, respectively). DNA shifts were observed for the *shp1358* and *ster_1357* probes (Fig. 6A, lanes 3 to 5) but the presence of the pheromone did not appear to influence the binding of the Rgg1358 protein to DNA. For all experiments, the *ldh* probe was used as negative control (Fig. 6A, lanes 6 and 7). These experiments show that Rgg1358 binds sequences found in the *shp1358* and *ster_1357* gene promoters and that the presence of the pheromone does not seem to be required for binding. To confirm this novel result, a $P_{rgg1358}$-*luxAB* fusion was constructed and used to transform strain LMD-9 and mutants deleted for the *rgg1358* or *shp1358* genes. The maximal expression of the *rgg1358* promoter was the same in the wild-type strain background and the *Δshp1358* mutant (70 RLU/OD_{600}) but was twofold higher in the *Δrgg1358* mutant (130 RLU/OD_{600}) (Fig. 6B). This indicates that the Rgg1358 protein is a repressor for the transcription of its own gene with or without the $SHP1358_{(15–23)}$ pheromone. This repressor effect is most probably direct. Indeed, the DNA-binding motif recognized by the Rgg1358 protein in the promoter region of the *shp1358* gene is located 16 nt upstream from the ATG of the *rgg1358* gene and consequently in its promoter region. These findings are in good agreement with the EMSA results and suggest that the Rgg1358 protein does not require the presence of pheromone to bind DNA.

Discussion

Here, we demonstrate the existence of, and describe, a QS mechanism involving a Rgg transcriptional regulator encoded by the *rgg1358* gene and a small hydrophobic peptide called SHP1358 (summarized in Fig. 7). We show that the triggering of the mechanism is biomass-dependent and relies on the presence, in the extracellular medium, of the mature form of the *shp1358* gene product, $SHP1358_{(15–23)}$, which is released by a C-terminal cleavage of the SHP1358 precursor. $SHP1358_{(15–23)}$ is most probably the pheromone of this QS mechanism. However, we can not completely exclude the possibility that it is an intermediate precursor to a more active species, or that there are other minor derivatives involved. The maturation of the SHP1358 precursor involves the transmembrane endopeptidase Eep. The capture of the pheromone from the extracellular medium appears to be performed mainly by one of the two oligopeptide binding proteins of the AmiCDEF oligopeptide transporter, AmiA3. Following the reimportation of $SHP1358_{(15–23)}$, it interacts with the regulatory protein Rgg1358. This interaction probably changes the Rgg1358 protein to an activated state, resulting in, at least, upregulation of the expression of the genes *shp1358* and *ster_1357*. The effect of the interaction remains unclear, but we suggest that it may involve a change of either conformation or oligomerization. At a low cell density, the Rgg1358 regulator may already be bound to its recognition sequences upstream from the promoters of the genes *shp1358* and *ster_1357* but in an inactive form.

Further study of two steps of this mechanism would be particularly interesting. First, the secretion mechanism responsible for the export of SHP pheromones is unknown. As the SHP sequences are similar to signal sequences, it is difficult to predict whether or not they use

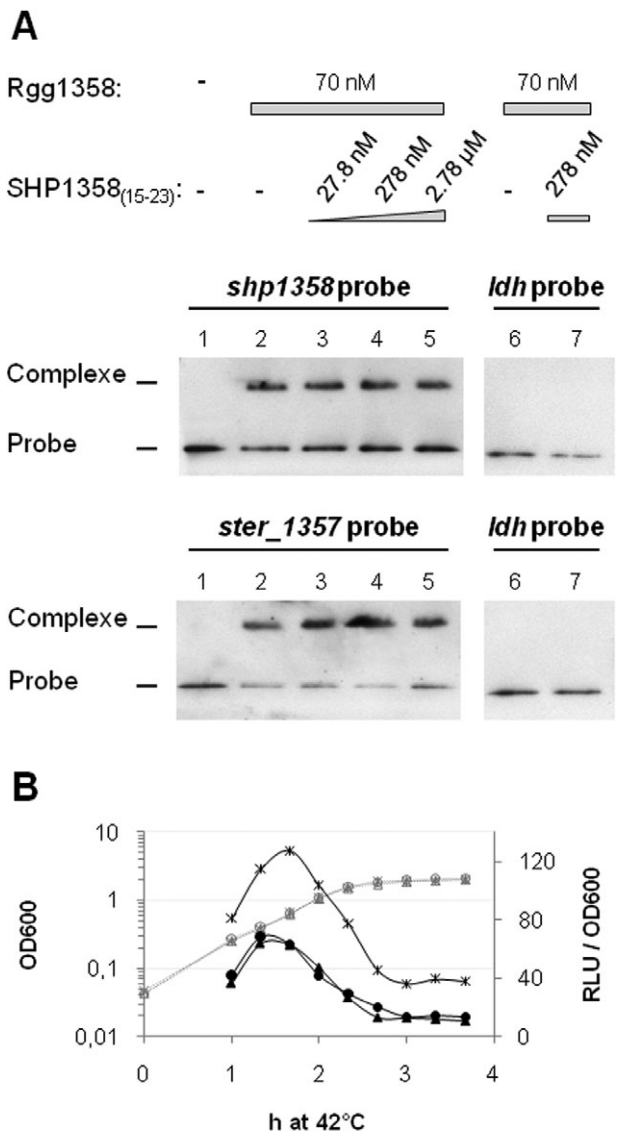


Fig. 6. Analysis of the Rgg1358 binding to DNA.
A. DNA binding of the His-tagged Rgg1358 protein to promoter regions of the two target genes of the SHP/Rgg1358 system, *ster_1357* and *shp1358*, with or without the addition of the synthetic SHP1358$_{(15–23)}$ pheromone. The *ldh* probe is an *ldh* promoter fragment used as a negative control.
B. Growth and luciferase activity of strains containing the *rgg1358-luxAB* gene fusion in the LMD-9 (●), Δ*shp1358* (▲) and Δ*rgg1358* (✳) genetic backgrounds. Growth curves (OD$_{600}$) are presented in gray and luciferase activities (RLU/OD$_{600}$) in black. Data shown are representative of three independent experiments.

the Sec machinery, need specific transporter(s) or insert into the cytoplasmic membrane to be exported. Second, the details of the roles of the two oligopeptide binding proteins are unclear. The oligopeptide binding protein AmiA3 is also the major actor for the capture and the sensing of the signalling peptide ComS (Gardan *et al.*, 2009; Fontaine *et al.*, 2010). The *amiA1* gene is the first gene of the *ami* operon whereas the *amiA3* gene is transcribed independently, and is separated from the *ami* operon by an insertion sequence. It would be of interest to determine if AmiA1 is more dedicated to nutritional function and if AmiA3 is largely responsible for the transport of pheromones into the cell as a result of a higher affinity for these hydrophobic peptides; alternatively, the presence of larger amounts of AmiA3 may be responsible for its dominant role.

This QS mechanism, involving a signalling peptide that is detected in the intracellular medium and imported by an oligopeptide transporter, has similarities with those described for the peptide-associated transcriptional regulators of the RNPP family. This family includes the PapR/PlcR and the NprX/NprR activators and the cCF10-iCF10/PrgX repressor (Declerck *et al.*, 2007). Only the QS mechanisms involving PlcR and PrgX have been studied in detail: the role of the peptide in the control of the activation state of their cognate regulator has been revealed by structural studies (Shi *et al.*, 2005; Declerck *et al.*, 2007). PSI-BLAST analyses indicate amino acid sequence similarities between various Rgg proteins (MutR of *S. mutans* and RggD of *S. gordonii*) and the transcriptional regulators PlcR and PrgX, throughout the protein (Declerck *et al.*, 2007). Like PlcR and PrgX, Rgg proteins have a Helix–Turn–Helix domain of the xenobiotic regulatory element family, which allows their binding to DNA (Neely *et al.*, 2003; Vickerman *et al.*, 2003; Rawlinson *et al.*, 2005; Samen *et al.*, 2006). We demonstrate such binding in this study. Rgg-associated peptides, like the anti-pheromone peptides from *E. faecalis*, resemble signal sequences and, like the pheromone and anti-pheromones, SHP1358 is matured by the endopeptidase Eep (An *et al.*, 1999; An and Clewell, 2002). Rgg-associated peptides are not similar to the PapR peptide, which is longer and in which the presence of a pro-peptide leader upstream from the mature form of the peptide indicates that it is probably exported and matured by the Sec machinery. A common feature of the RNPP regulators is the presence of tetratricopeptide repeat (TPRs) domains in their C-terminal part that are involved in protein/protein or protein/peptide interactions (Goebl and Yanagida, 1991). TPRpred software (Karpenahalli *et al.*, 2007) failed to identify such domains in the Rgg proteins. However, the TPR domains of the PrgX repressor were only discovered after structural analysis (Shi *et al.*, 2005). Consequently, structural analyses, involving for example crystallography experiments, are required to determine whether or not the Rgg regulatory proteins contain TPR domains. We therefore propose that the SHP-associated Rgg are new members of the RNPP family. Indeed, SHP/Rgg systems have also some specific features. By searching the complete genome sequences of streptococci, we retrieved 68 SHP representing 28 different amino acid sequences. Among the various peptides asso-

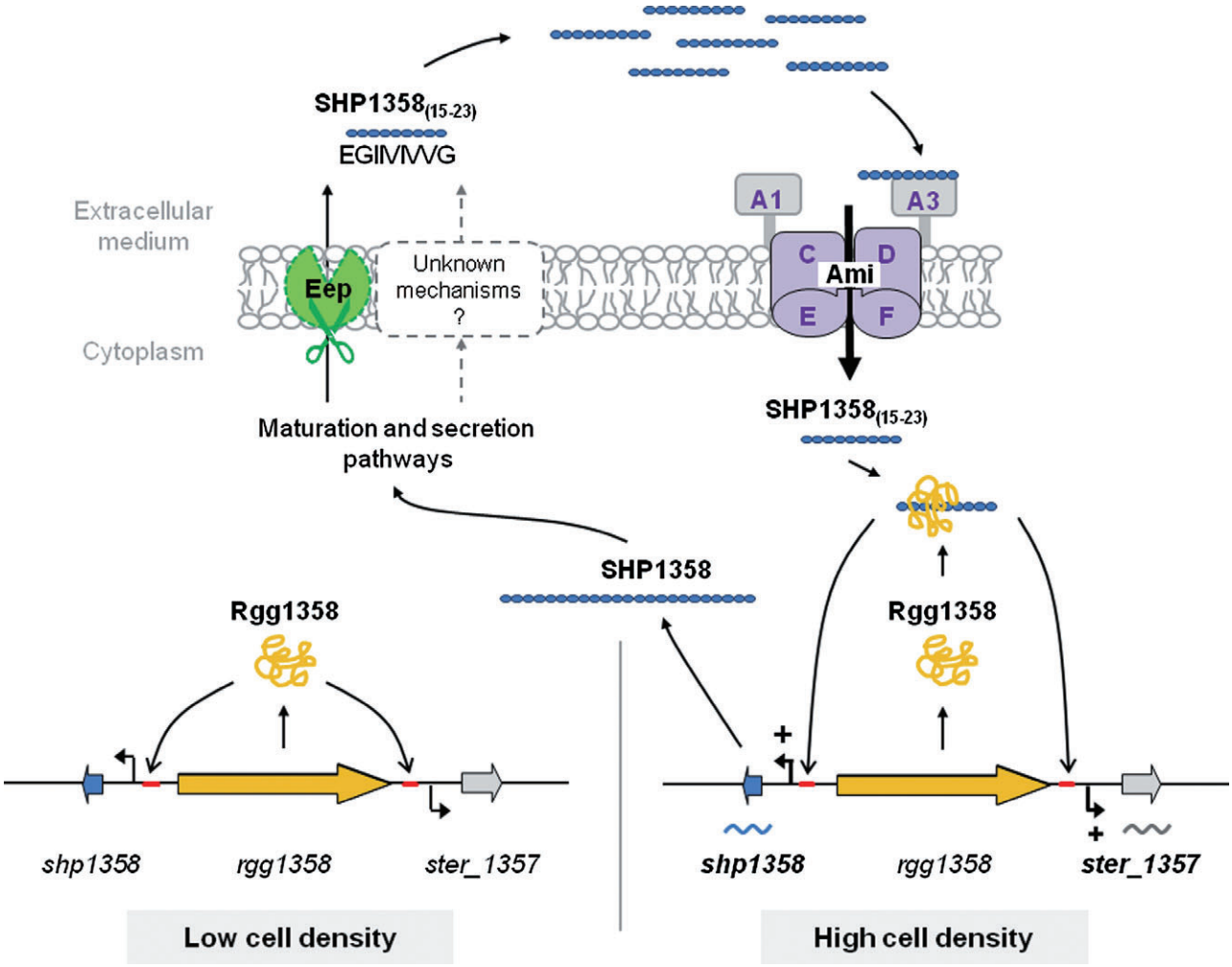


Fig. 7. Schematic representation of the SHP/Rgg1358 quorum-sensing mechanism of *S. thermophilus* LMD-9. The quorum-sensing signal is encoded by the *shp1358* gene. The pheromone, a small hydrophobic peptide called SHP1358$_{(15–23)}$ (EGIIVIVVG), is produced by C-terminal cleavage of a 23 amino acid peptide precursor, a process that involves the endopeptidase Eep localized in the membrane. At high cell density, the secreted SHP1358$_{(15–23)}$ is sensed by the lipoprotein AmiA3 and reimported into the cell by the AmiCDEF transporter. Intracellular SHP1358$_{(15–23)}$ then interacts with the regulatory protein Rgg1358. The effect of the interaction is unknown but binding of the SHP1358$_{(15–23)}$/Rgg1358 complex to the promoter region of the two target genes, *shp1358* and *ster_1357*, leads to the activation of their transcription. At low cell density, we hypothesize that the SHP1358$_{(15–23)}$ is present in a negligible amount and the Rgg1358 protein is bound alone to the promoter region of *shp1358* and *ster_1357* in an inactivated state that does not allow their transcription.

ciated with regulators of the RNPP family, SHPs constitute the largest group. They have one conserved aspartate or glutamate residue in their C-terminal part, possibly important for the maturation step and nearly always a conserved glycine residue at the end of their sequence. The DNA recognition sequence also seems to be specific with no similarities to the inverted repeat sequences recognized by PlcR or PrgX (Fig. 5B). Similar sequences have been found upstream from 80 % of *shp* genes belonging in *S. thermophilus* and other streptococci (e.g. B. Fleuchot, unpublished) and has the particularity of not being palindromic. Finally, our *in vitro* and genetic experiments demonstrate that SHP1358$_{(15–23)}$ interacts with the transcriptional regulator Rgg1358 but that it is probably not necessary for the binding of the Rgg1358 protein to DNA. In *B. cereus*, PapR pheromone binding to the PlcR activator leads to its multimerization, which is essential for its binding to the PlcR box and therefore for the regulation of the PlcR regulon (Slamti and Lereclus, 2002; Declerck *et al.*, 2007). In *E. faecalis*, the situation is more complex. Indeed, although PrgX is able to bind DNA without the absence of its associated pheromones (Bae *et al.*, 2002), their interaction is essential to regulate the expression of the target genes. The binding of the cCF10 sex pheromone to the PrgX repressor modifies its tetrameric conformation to become two dimers with less affinity for DNA, leading to the opening of a DNA loop and allowing access of the RNA polymerase to the target *prgQ* promoter. In

contrast, the binding of iCF10 neutralizes the cCF10 pheromone activity and favours the return to the tetrameric form of PrgX (Shi *et al*., 2005; Dunny, 2007). Further experiments are needed to determine if the $SHP1358_{(15–23)}$ pheromone influences the oligomerization state of the Rgg1358 protein, and if this is the cause of the activation of the Rgg1358 regulator, independently of its binding capacity. However, these results indicate that this QS mechanism involving Rgg regulators differs from those described for PlcR and PrgX.

The regulation of competence in *S. thermophilus* LMD-9 is controlled by a Rgg-like protein (ComR) associated with a hydrophobic pheromone (ComS) reimported into the cell by the Ami oligopeptide transporter (Gardan *et al*., 2009; Fontaine *et al*., 2010). Mashburn-Warren *et al*. (2010) showed that *comS/comR* orthologues are also present in the genomes of *S. mutans* and in pyogenic and bovis streptococci. Moreover, they found results *in S. mutans* similar to those in *S. thermophilus*: (i) a deletion of *comR*, *comS* or of subunits of the oligopeptide transporter abolish transformability, (ii) ComR controls the transcription of the gene encoding the alternative sigma factor, SigX, itself involved in the control of the transcription of the genes encoding the DNA uptake and recombination machinery and (iii) extracellular addition of a synthetic peptide derived from the ComS precursor induces development of competence. The *comS* orthologues encode peptides, named XIP, which are 17 to 32 amino acids long and contain a conserved double-tryptophan in their C-terminal part (except in the case of *S. thermophilus*). The ComR regulators have similarity with both Rgg and PlcR proteins, and their predicted structures have similarities with the crystal structure of PlcR. These various observations led the authors to propose that XIP-associated ComR regulators are new members of the RNPP family. Although we agree with this, we think that SHP/Rgg and XIP/ComR pairs are distinct branches of the RNPP family for the following reasons: first, our phylogenetic analysis of the Rgg and Rgg-like proteins clearly shows that SHP-associated Rgg and XIP-associated ComR cluster in different parts of the tree and that the two families of SHP and XIP peptides differ in terms of their amino acid sequences and their conserved amino acids; second, their predicted DNA recognition sequences are not similar; and finally, the *shp* and *rgg* genes are in all cases transcribed in opposite directions whereas the *comS* genes all map downstream from the *comR* genes and in the same orientation. This last point has an incidence on the SHP/Rgg mechanism. Where the *shp* gene is located upstream from the *rgg* gene and divergently transcribed, Rgg can simultaneously be the repressor of its own gene and the activator of the *shp* gene. This potentially prevents the system from becoming over-active or overexpressed. For the new SHP/Rgg group described in this study (group III, in which the *shp* gene is downstream from the *rgg* gene and convergently transcribed), the 3′ ends of the two genes always overlap. This genetic configuration may also have consequences for the regulation of the system.

To facilitate discrimination between these small ComS/XIP hydrophobic peptides and the SHP we describe herein, we propose to restrict the use of the acronym SHP to small hydrophobic peptides with a conserved aspartate or glutamate residue about nine amino acids before the C-ter end of the sequence, encoded by a gene that is not transcribed in the same orientation as the *rgg* gene.

Another important issue is the range of physiological functions controlled by these systems. The discovery of 68 SHP/Rgg systems suggests a QS mechanism that is widespread in the streptococci family. Unlike XIP/ComR systems that all seem to control the transcription of at least the *sigX* gene and consequently the triggering of the competence state in different species of streptococci, the targets controlled by the SHP/Rgg systems are not obvious. For the SHP/Rgg1358 QS mechanism described here, one target, *ster_1357*, encoding a cyclic secreted peptide called Pep1357C was identified. This gene, mapping downstream from the *rgg1358* gene, is followed by two genes encoding a protein belonging to the S-adenosylmethionine radical enzyme family and an efflux transporter; it has been suggested that the proteins encoded by these two genes are responsible for the maturation and the secretion of Pep1357C outside the cell (Ibrahim *et al*., 2007b). A nearly identical locus including the *shp/rgg* genes has been found in a recently sequenced *Streptococcus mitis* strain (SK564). We found a small number of homologous genes i.e. encoding a peptide, a S-adenosylmethionine radical enzyme and an exporter downstream from SHP-associated Rgg, in various streptococci species. Four similar genetic structures in *B. subtilis* and more recently in *Bacillus thuringiensis* have been studied. None of them are close to genes encoding Rgg proteins (Zheng *et al*., 2000; Brede *et al*., 2004; Butcher *et al*., 2007; Rea *et al*., 2010). All these loci encode modified peptides with antimicrobial activities. We therefore tested Pep1357C for bactericidal and bacteriostatic activity against several strains and species but no bacterial growth inhibition or killing effects were detected in the conditions tested.

Finally, it would also be interesting to explore other pathways linked to this QS mechanism. For example, *shp/rgg* loci are present in multicopy in streptococci, so it would be interesting to test for cross-activation within cells, or between strains or species. Any such interspecies interaction would be highly relevant to the study of various natural ecosystems, particularly for example the

human oral microbiome where several species of streptococci live together (Dewhirst *et al*., 2010). Another level of complexity is brought by the presence of other RNPP systems in the same strain. Elucidating how the bacteria manage this plethoric repertoire of peptides and how it could be manipulated would be potentially valuable.

Experimental procedures

Phylogenetic analysis of the Rgg family

A preliminary PSI-BLAST search, against the non-redundant NCBI database, detected protein sequences with statistically significant similarity to Rgg regulators only in the order of *Lactobacillales* and the family *Listeriaceae*. We thus conducted a comparative analysis of the Rgg and Rgg-like repertoire in the genomes belonging to these two taxonomic groups. The analysis included a total of 90 complete genome sequences corresponding to all the sequenced isolates available in Genome Reviews release 125 (05-OCT-2010). This list includes representatives of eight genera: *Streptococcus* (48 genome sequences), *Lactobacillus* (22), *Listeria* (9), *Lactococcus* (5), *Leuconostoc* (3), *Enterococcus* (1), *Oenococcus* (1) and *Pediococcus* (1). Three prey representative of the three types of SHP/Rgg systems described in Ibrahim *et al*. (2007a) (*stu.0182* in *S. thermophilus* LMG18311, *smu.1509* in *S. mutans* UA159 and *gbs1555* in *S. agalactiae* NEM316) were used to retrieve an exhaustive list of Rgg and Rgg-like proteins. Hits accounting for a least 80% of the length of one prey were collected after three iterations with PSI-BLAST (*E*-value $\leq$ 0.001). The sequences were aligned with muscle (default parameters) and a maximum-likelihood phylogenetic tree was constructed with phyml (JTT model of protein evolution, gamma distributed rate of evolution, 4 discrete categories of sites). The statistical support for each branch was quantified by analysing 200 bootstrap replicates. The tree was drawn in R using the library 'ape' (Paradis *et al*., 2004). Rgg-associated SHP or XIP peptides were recovered as follows. First, all hydrophobic peptides with basic residues (lysine and more rarely arginine) at the N-terminus were selected. Then, peptides with a glutamate or an aspartate at their C-terminus and encoded by a gene upstream from the *rgg* gene and divergently transcribed were considered to be SHP. Peptides with a double tryptophan motif and encoded by a gene downstream from the *rgg* gene and in the same orientation were considered to be XIP.

Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 1. *S. thermophilus* strains were grown at 28°C or 42°C in M17 medium (Difco) supplemented with 10 g l^{-1} lactose (M17lac) or in a chemically defined medium (CDM) without shaking, under atmospheric air and with a ratio of air space to liquid of approximately 90% (Letort and Juillard, 2001). *Escherichia coli* strains were grown at 30°C or 37°C in Luria–Bertani (LB) broth with shaking (Sambrook & Russell, 2001). Agar (1.5%) was added to the media as appropriate. When required, antibiotics were added to the media at the following final concentrations: erythromycin, 200 μg ml^{-1} for *E. coli* or 5 μg ml^{-1} for *S. thermophilus*, kanamycin, 30 μg ml^{-1} for *E. coli* or 1 mg ml^{-1} for *S. thermophilus* and spectinomycin, 100 μg ml^{-1} for *S. thermophilus*. The optical density at 600 nm of the cultures was measured with a Uvikon 931 spectrophotometer (Kontron).

DNA manipulation and sequencing

Standard methods were used for DNA purification, restriction digestion, PCR, ligation and sequencing. *E. coli* TG1 or TG1 *repA*$^+$ strains were used as hosts for cloning experiments. The oligonucleotides used for PCR are listed in Table 2. *S. thermophilus* was transformed using natural competent cells or electrocompetent cells (Ibrahim *et al*., 2007b). The 5′ tag-RACE method described by Fouquier d'Hérouel *et al*. (2011) was used to determine the transcriptional start site of the genes *ster_1357* and *shp1358* with oligonucleotides shp1358-tag and ster_1357-tag.

Plasmid and strain constructions

The plasmids used are listed in Table 3. Plasmid pGICB004, which allows the integration of transcriptional fusions to the *luxAB* reporter genes at the *blp* locus in *S. thermophilus*, was constructed in two steps. First, a 0.95 kb fragment including the *blpR* promoter sequence, the *blpR* ORF and the first 96 bp of *blpH* was amplified by PCR with primers UPINSLUX1 and UPINSLUX2, digested with PshAI and SpeI, and ligated upstream from *luxAB* into similarly digested pJIM4900. Next, a 1.05 kb fragment located downstream from *blpX* was amplified by PCR with primers DNINSLUX1 and DNINSLUX2, digested with SalI and PvuII, and inserted downstream from *luxAB* between the SalI and PvuII sites of the plasmid obtained in the first step. Plasmid pGICB004::$P_{shp1358}$ and pGICB004::$P_{rgg1358}$ were constructed as follows. The *shp1358* and *rgg1358* promoter were amplified by PCR with oligonucleotides Pshp1358-EcoRI/Pshp1358-SpeI and Prgg1358-EcoRI/Prgg1358-SpeI, respectively, digested with restriction enzymes SpeI/EcoRI and ligated between the same restriction sites of pGICB004. Integration and excision of these plasmids in strains LMD-9, TIL 778 (*Δshp1358*), TIL 775 (*Δrgg1358*) and TIL 1198 (*ΔamiA1*) (Garault *et al*., 2000 for experimental procedure) led to the construction of strains TIL 1165 (*blp*::$P_{shp1358}$-*luxAB*) and TIL 1164 (*blp*::$P_{rgg1358}$-*luxAB*), TIL 1200 (*Δshp1358 blp*::$P_{shp1358}$-*luxAB*) and TIL 1208 (*Δshp1358 blp*::$P_{rgg1358}$-*luxAB*), TIL 1169 (*Δrgg1358 blp*::$P_{shp1358}$-*luxAB*) and TIL 1168 (*Δrgg1358 blp*::$P_{rgg1358}$-*luxAB*) and TIL 1204 (*ΔamiA1 blp*::$P_{shp1358}$-*luxAB*). Competent cells of strain TIL 1204 were transformed with pG$^+$host9::updown.*shp*1358. Integration and excision of the plasmid gave strain TIL 1303 (*ΔamiA1 Δshp1358 blp*::$P_{shp1358}$-*luxAB*). The SHP1358 peptide was overproduced by inserting the *shp1358* gene into pBV5030 downstream from the P32 promoter. The *shp1358* gene was amplified by PCR with oligonucleotides shp1358-NcoI and shp1358-PstI. The resulting fragment was digested with restriction enzymes NcoI/PstI and inserted into pBV5030::P_{32}-*ster_1357* digested with the same enzymes. The resulting plasmid pBV5030::P_{32}-*shp1358* allowed the overproduction of a modified SHP1358 peptide containing a

Table 1. Bacterial strains used in this study.

Bacterial strain and genotypes		Resistance[a]	Description[b]	Source or reference
Streptococcus thermophilus				
LMD-9	Wild type			Makarova *et al.* (2006)
TIL 773	Δ*eep*		pG$^+$host9::updown.*eep* → LMD-9	This study
TIL 775	Δ*rgg1358*			Ibrahim *et al.* (2007b)
TIL 778	Δ*shp1358*			Ibrahim *et al.* (2007b)
TIL 883	Δ*amiCDE*			Ibrahim *et al.* (2007b)
TIL 1164	*blp*::$P_{rgg1358}$-*luxAB*		pGICB004::$P_{rgg1358}$ → LMD-9	This study
TIL 1165	*blp*::$P_{shp1358}$-*luxAB*		pGICB004::$P_{shp1358}$ → LMD-9	This study
TIL 1168	Δ*rgg1358 blp*::$P_{rgg1358}$-*luxAB*		pGICB004::$P_{rgg1358}$ → TIL 775	This study
TIL 1169	Δ*rgg1358 blp*::$P_{shp1358}$-*luxAB*		pGICB004::$P_{shp1358}$ → TIL 775	This study
TIL 1197	*amiA3::erm*	Erm		Gardan *et al.* (2009)
TIL 1198	Δ*amiA1*			Gardan *et al.* (2009)
TIL 1199	Δ*amiA1 amiA3::erm*	Erm		Gardan *et al.* (2009)
TIL 1200	Δ*shp1358 blp*::$P_{shp1358}$-*luxAB*		pGICB004::$P_{shp1358}$ → TIL 778	This study
TIL 1202	*amiCDE::spec*	Spec	PCR fragment *amiCDE::spec* → LMD-9	This study
TIL 1203	*amiCDE::spec blp*::$P_{shp1358}$-*luxAB*	Spec	TIL 1202 DNA → TIL 1165	This study
TIL 1204	Δ*amiA1 blp*::$P_{shp1358}$-*luxAB*		pGICB004::$P_{shp1358}$ → TIL 1198	This study
TIL 1206	Δ*eep* pBV5030::P32-*shp1358*	Erm	pBV5030::P32- *shp1358* → TIL 773	This study
TIL 1208	Δ*shp1358 blp*::$P_{rgg1358}$-*luxAB*		pGICB004::$P_{rgg1358}$ → TIL 778	This study
TIL 1213	Δ*amiCDE* pBV5030::P32- *shp1358*	Erm	pBV5030::P32- *shp1358* → TIL 883	This study
TIL 1303	Δ*amiA1* Δ*shp1358 blp*::$P_{shp1358}$-*luxAB*		pG$^+$host9::updown.*shp1358* → TIL 1204	This study
TIL 1304	*amiA3::erm* Δ*shp1358 blp*::$P_{shp1358}$-*luxAB*	Erm	TIL 1197 DNA → TIL 1200	This study
TIL 1305	Δ*amiA1 amiA3::erm* Δ*shp1358 blp*::$P_{shp1358}$-*luxAB*	Erm	TIL 1199 DNA → TIL 1200	This study
Escherichia coli				
TG1			*supE hsdΔ5 thi* Δ(*lac-proAB*) F'[*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15]	Gasson, 1983
TG1 *repA*$^+$		Km	TG1 derivative with *repA* gene integrated into the chromosome	Pierre Renault[c]
Rosetta		Cm	F$^-$ *ompT hsdS*$_B$(r_{B^-} m_{B^-}) *gal dcm* pRARE	Novagen
TIL 1318	Rosetta pET28a::Rgg1358	Km	pET28a::Rgg1358 → Rosetta	This study

a. Erm, Spec, Km, Cm indicate resistance to erythromycin, spectinomycin, kanamycin and chloramphenicol respectively.
b. Arrows indicate construction by natural transformation or electroportation with chromosomal DNA or plasmid; 'updown' followed by the name of a gene indicates that the fragments upstream and downstream from the gene were inserted into pG$^+$host9 to construct in frame deletions of the gene by double cross-over event (Garault *et al.*, 2000).
c. INRA, Jouy-en-Josas, France.

glycine residue between the methionine at position 1 and lysine at position 2 of the wild-type sequence. Strain TIL 1202 (*amiCDE::spec*) was constructed using the overlapping PCR method. Briefly, the spectinomycin cassette (spec) was amplified by PCR with oligonucleotides Spec-R and Spec-F2 and the pAT28 as the template (Trieu-Cuot *et al.*, 1990); it was fused by PCR to the beginning of the *amiC* gene (upstream fragment) and the end of the *amiE* gene and the *amiF* gene (downstream fragment). Oligonucleotides used for the amplification of the upstream fragment were amiCDE_up-F and amiCDE_up-R and for the downstream fragment amiCDE_down-F and amiCDE_down-R. The resulting 3 kb fragment was used to transform LMD-9 cells. Strains TIL 1304 (*amiA3::erm Δshp1358 blp*::$P_{shp1358}$-*luxAB*) and TIL 1305 (*ΔamiA1 amiA3::erm Δshp1358 blp*::$P_{shp1358}$-*luxAB*) were constructed as follows. Cells of strain TIL 1200 were transformed with chromosomal DNA from strains TIL 1197 (*amiA3::erm*) and TIL 1199 (*ΔamiA1 amiA3::erm*) respectively. A similar procedure was used to construct strain TIL 1203 (*amiCDE::spec blp*::$P_{shp1358}$-*luxAB*): TIL 1165 was transformed with chromosomal DNA from strain TIL 1202. Strain TIL 773 (*Δeep*) was constructed by deleting an internal fragment of the gene by a double cross-over event using pG$^+$host9. Briefly, oligonucleotides eep-XhoI with eep-EcoRIA and eep-EcoRIB with eep-SpeI were used to amplify upstream and downstream fragments of the *eep* gene. These two fragments were double digested with restriction enzymes XhoI with EcoRI and EcoRI with SpeI, respectively, and ligated between the XhoI and SpeI restriction sites of pG$^+$host9. The resulting plasmid was used to transform electrocompetent cells of strain LMD-9. Integration and excision of the plasmid led to the in-frame deletion of the *eep* gene. Strains TIL 1206 (*Δeep* pBV5030::P_{32}-*shp1358*) and TIL 1213 (*ΔamiCDE* pBV5030::P_{32}-*shp1358*) were constructed by electrotransformation of strains TIL 773 and TIL 883, respectively, with pBV5030::P_{32}-*shp1358*. Plasmid pET28a::Rgg1358 was constructed to allow the expression and the purification of the N-terminus His-tagged Rgg1358 protein in *E. coli* as follows. The *rgg1358* gene was amplified by PCR with oligonucleotides Rgg1358-NdeI and Rgg1358-XhoI. The resulting 882 bp fragment was digested with restriction enzymes NdeI/XhoI and ligated into in the expression vector pET28a digested with the same restriction enzymes. The resulting plasmid was used to transform *E. coli* strain Rosetta to obtain TIL 1318 (*E. coli* Rosetta pET28a::Rgg1358). All constructions were verified by PCR and validated by sequencing.

Table 2. Primers used in this study.

Primer	Sequence (5′ → 3′)[a,b]
Mutants and plasmids constructions	
UPINSLUX1	AAGACCACAGTCCAGACGAATCTTCTGTCCCA
UPINSLUX2	CGACTAGTGTATCCAGTCTCTAAAATTCAG
DNINSLUX1	AAGTCGACTTCCACTGTCTTGGCTATATC
DNINSLUX2	AACAGCTGCGTGTCAAAGCTACTGTAGTC
Pshp1358-EcoRI	GAAGAATTCTGCTTTTTCATGTGATAAT
Pshp1358-SpeI	AACTACTAGTAGACGTTGATATACCAGC
Prgg1358-EcoRI	GAAGAATTCAATTTTTTCATACTACACCTCC
Prgg1358-SpeI	ACCTACTAGTCGACGATAACGATAATGCC
shp1358-NcoI	CCACCATGGGTAAAAAGCAAATTTTACTAAC
shp1358-PstI	CCTGCTGCAGAATACAGAACTAAATCC
Spec-R	GAAAAATTCTATAGAAACTTCTCTCAATTAGGCT
Spec-F2	TACAGATTAATAATTATTCTTTATTATACAGATC
amiCDE_up-F	CGGCTTTGACCTACACG
amiCDE_up-R	GAAAAATTCTATAGAAACTTCTCTCAATTAGGCTAAGGCGATGGTTGGTAGTG
amiCDE_down-F	TACAGATTAATAATTATTCTTTATTATACAGATCTGAGGAAAAAGCACCAC
amiCDE_ down-R	AAGTGACCTGGGCGTAC
eep-XhoI	GAAGAATTCTTAAAAAAGTCCTAGTAAGTGC
eep-EcoRIA	CTCCTCGAGTACCTTTATCGTGATCC
eep-EcoRIB	GAAGAATTCTTAGAAACAAAGAAATGAAAGG
eep-speI	AACTACTAGTTTGAGTCACAACCTTAG
Rgg1358-NdeI	CCACCACATATGAAAAAATTTGGAG
Rgg1358-XhoI	CCGCTCGAGCTACTAACTACTAGCTAAAA
5′ tag-RACE	
shp1358-tag	ACGATAACGATAATGCCTTCAAATACGA
ster_1357-tag	ATCACTCCCAGCCGTCACCTTTTGCCATC
EMSA	
EMSA-shpF	CTTTATTGTATGATATGATG
EMSA-shpR	CTCTCGATTCTCGAAAC
EMSA-ster_1357F	CTTATGAAGGACATTTAG
EMSA-ster_1357R	TGATACTGTGATTATATTTC
EMSA-ldhF	ACGCTTTCACTTAATAATTC
EMSA-ldhR	TGGTCTAAACATCTCCTTA

a. Restriction enzyme recognition sequences are underlined.
b. Complementary sequences of the spectinomycin cassette are underlined with dashes.

Table 3. Plasmids used in this study.

Plasmid	Description[a]	Source or reference
pG$^+$host9::updown.*shp1358*	Erm, Ts plasmid for *shp1358* gene replacement by double cross-over integration.	Ibrahim *et al.* (2007b)
pJIM4900	Erm, Ts derivative of pG$^+$host9 containing the *luxAB* genes of *Photorhabdus luminescens* and a transcriptional terminator.	E. Guédon[b] (e.g. unpublished)
pGICB004	Erm, pJIM4900 with a 0.95 kb insert containing two fragments upstream and downstream respectively from the *luxAB* gene and allowing the integration of the plasmid by double cross-over event at the *blp* locus of *S. thermophilus*.	This study
pGICB004::$P_{shp1358}$	Erm, Derivative of pGICB004 used to introduce a $P_{shp1358}$-*luxAB* transcriptional fusion at the *blp* locus.	This study
pGICB004::$P_{rgg1358}$	Erm, Derivative of pGICB004 used to introduce a $P_{rgg1358}$-*luxAB* transcriptional fusion at the *blp* locus.	This study
pBV5030	Erm, Replicative plasmid in *E. coli* and Gram positive bacteria.	Bojovic *et al.* (1991)
pBV5030::P_{32}-*ster_1357*	Erm, Derivative of pBV5030 allowing the overexpression of the *ster_1357* gene of *S. thermophilus* LMD-9.	Gardan (e.g. unpublished)
pBV5030::P_{32}-*shp1358*	Erm, Derivative of pBV5030::P_{32}-*ster_1357* allowing the overexpression of the *shp1358* gene of *S. thermophilus* LMD-9.	This study
pET28a	Km, vector for production of His tagged proteins	Novagen
pET28a::Rgg1358	A derivative of pET28a used to overproduce Rgg1358	This study

a. Erm and Km indicate resistance to erythromycin and kanamycin respectively; Ts indicates that the plasmid encodes a thermosensitive RepA protein.
b. INRA, Jouy-en-Josas, France.

Luciferase assays

Cells were grown overnight at 42°C in CDM. These cultures were then diluted in 50 ml of CDM to a final OD_{600} of 0.05 and incubated at 42°C, except for the experiments reported in Fig. 2D where cultures were diluted to a final OD_{600} of 0.012, 0.025, 0.05 or 0.1. Aliquots of 1 ml of the culture were sampled at regular intervals until the culture reached stationary phase and analysed as follows: OD_{600} was measured, then 10 µl of a 0.1% nonyl-aldehyde solution was added and the luminescence was measured with a Luminoskan TL (Labsystems). Results are reported in Relative Luminescent Units divided by the OD_{600} (RLU/OD_{600}). For cocultures, each culture was diluted to an OD_{600} of 0.025 and results are expressed as luminescence divided by the total OD_{600}. For experiments involving mixing supernatants of one strain with cell pellets of another strain containing a *luxAB* reporter fusion, both supernatants and pellets were harvested by centrifugation (5000 *g* for 10 min at RT) of equal volumes of cultures at OD_{600} 0.6. Finally, the synthetic peptide EGIIVIVVG, stored in lyophilized form and freshly prepared in DMSO, was added to cultures with an OD_{600} of 0.5.

LC-MS/MS

Supernatants from cultures of strains LMD-9 and TIL 778 were recovered by centrifugation. Aliquots of 10 µl of ultrafiltered supernatant (0.22 µM, Millipore) were loaded on a Pepmap C18 column (length 150 mm, 75 µm ID, 100 Å; Dionex, Voisin-le-Bretonneux) and analysed on-line by mass spectrometry on a LTQ-Orbitrap Discovery apparatus (Thermo Fischer, San Jose). First, only ion current of m/z corresponding to all possible C-terminal amino acid sequences produced by the cleavage of the SHP1358 sequence were sought in the supernatant of strain LMD-9. Next, we checked that these masses, detected in the LMD-9 supernatant, were not present in the supernatant of the *Δshp1358* mutant and that they were eluted from the Pepmap C18 column with a retention time compatible with the hydrophobicity of the putative amino acid sequence. Then, the ion with mass that fulfilled these conditions was fragmented and analysed on the Orbitrap mass analyser, to identify accurately the sequence of the corresponding peptide. Finally, the absence of the mature form of this peptide from the *Δshp1358* supernatant was confirmed after fragmentation on LTQ by monitoring the current of the b7 ions daughter (724.46 m/z) in the Orbitrap mass analyser.

Production and purification of Rgg1358

A preculture of TIL 1318 was diluted in 500 ml of LB with kanamycin to an OD_{600} of 0.05 and incubated at 37°C with shaking. At OD_{600} 0.5, recombinant protein production was induced by adding IPTG at a final concentration of 1 mM and the incubation continued for 4 h at 30°C. Cells were harvested by centrifugation (5000 *g* for 10 min at 4°C), rinsed and resuspended in disrupter buffer (20 mM Tris HCl pH 8.0, 250 mM NaCl, 0.1% Tween 20, 10% v/v glycerol and 10 mM Imidazole) to obtain a OD_{600} of 100 ml^{-1}. Cells were disrupted at 2.0 kbars with a Basic Z Cell Disruption System (Constant Systems). The soluble fraction was collected by centrifugation (5000 *g* for 20 min at 4°C) and production of the recombinant His-tagged Rgg1358 protein checked by SDS-PAGE. The product was then purified with the QIAexpressionist kit (Qiagen), according to the manufacturer's instructions with the following modification: elution was carried out with an imidazole concentration gradient from 100 mM to 250 mM.

SPR analysis

Real-time binding kinetics experiments were conducted on a BIAcore 3000 apparatus (GE Healthcare Europe). The synthetic peptide EGIIVIVVG was immobilized on a CM5 sensor chip using amine-coupling chemistry. For this, the surface of the chip was activated for 7 min with a mixture of 0.05 M NHS and 0.2 M EDC (Amine coupling kit, Biacore, GE Healthcare). The synthetic pheromone (37 µM, in sodium acetate buffer, pH 5.0) was then covalently linked to the surface giving up to 2000 resonance units (RU). Ethanolamine (1 M, pH 8.5) was injected for 7 min to block the remaining activated groups. The purified Rgg1358 protein (5 µM) was injected over 2 min, at flow of 25 µl min^{-1} in 20 mM HBS (pH 7.4) to allow binding to the immobilized $SHP1358_{(15–23)}$ pheromone; dissociation was recorded for 12 min after the end of injections. The surface was regenerated by injecting glycine buffer (pH 2.0) for 2 min. The sensorgram observed in a cell immobilizing pheromone was corrected by subtracting the response observed in a cell immobilizing only ethanolamine (reference surface). Two sets of control experiments were performed to check the specificity of binding: (i) BSA was injected over the immobilized $SHP1358_{(15–23)}$, and (ii) Rgg1358 was injected over a chip surface with the control peptide, IAILPYFAGCL, covalently immobilized (purchased from eurogentec). All measurements were performed at 25°C. Sensorgrams were analysed using BIAevaluation Software.

EMSA

DNA probes of approximately 110 bp of the *shp1358* and *ster_1357* promoter regions were amplified by PCR with oligonucleotides EMSA-shpF with EMSA-shpR and EMSA-ster_1357F with EMSA-ster_1357R respectively. The *ldh* promoter region was amplified with oligonucleotides EMSA-ldhF and EMSA-ldhR as a control. DNA probes were 3′ end labelled and gel shift reactions carried out by using the DIG GEL Shift Kit, 2nd Generation (Roche), according to the manufacturer's instructions. The DNA binding reactions, coupling probes to the Rgg1358 protein with or without the synthetic peptide EGIIVIVVG (purchased from eurogentec), involved incubation for 30 min at 42°C. Samples were then loaded on a 4–16% native polyacrylamide gels (nativePAGE 4–16% Bis-Tris Gel, Invitrogen) and subjected to 70 V for approximately 2 h in 1 × TBE buffer. The labelled probe/protein complexes were transferred to a positively charged nylon membrane (GE Healthcare Amersham Hybond – N^+) by electro-blotting for 30 min in a Mini Trans-Blot Cell (Biorad) in 0.5 × TBE buffer. DNA complexes were detected by chemiluminescence on X-ray film (Amersham Hyperfilm ECL) according to the manufacturer's instructions.

Acknowledgements

This study was supported by the Institut National de la Recherche Agronomique (INRA) and the Ministère de l'Education Nationale de la Recherche et de la Technologie (MENRT). The 'Plateforme d'Analyse Protéomique Paris Sud-Ouest' (PAPPSO, Alain Guillot) received the financial support from the Ile de France regional council and from CEMAGREF. Work in the group of P. Hols was financially supported by FNRS. L. Fontaine is postdoctoral researcher at FNRS. P. Hols is research associate at FNRS. We thank L. Topisirovic for plasmid pBV5030. We thank F. Rul, V. Juillard and P. Serror for their critical reading of the manuscript, E. Chambellon and M. Cote for their technical assistance, and E. Bruneau for her technical advice. Finally, we are grateful to F. Repoila for helpful advice concerning his 5′ tag-RACE method.

References

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A.B., and Lereclus, D. (1999) PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol Microbiol* **32:** 1043–1053.

An, F.Y., and Clewell, D.B. (2002) Identification of the cAD1 sex pheromone precursor in *Enterococcus faecalis*. *J Bacteriol* **184:** 1880–1887.

An, F.Y., Sulavik, M.C., and Clewell, D.B. (1999) Identification and characterization of a determinant (*eep*) on the *Enterococcus faecalis* chromosome that is involved in production of the peptide sex pheromone cAD1. *J Bacteriol* **181:** 5915–5921.

Antunes, L.C., Ferreira, R.B., Buckner, M.M., and Finlay, B.B. (2010) Quorum sensing in bacterial virulence. *Microbiology* **156:** 2271–2282.

Bae, T., Kozlowicz, B., and Dunny, G.M. (2002) Two targets in pCF10 DNA for PrgX binding: their role in production of Qa and prgX mRNA and in regulation of pheromone-inducible conjugation. *J Mol Biol* **315:** 995–1007.

Bassler, B.L. (2002) Small talk: Cell-to-cell communication in bacteria. *Cell* **109:** 421–424.

Bojovic, B., Djordjevic, G., and Topisirovic, L. (1991) Improved vector for promoter screening in lactococci. *Appl Environ Microbiol* **57:** 385–388.

Bortoni, M.E., Terra, V.S., Hinds, J., Andrew, P.W., and Yesilkaya, H. (2009) The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology* **155:** 4123–4134.

Bouillaut, L., Perchat, S., Arold, S., Zorrilla, S., Slamti, L., Henry, C., *et al*. (2008) Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. *Nucleic Acids Res* **36:** 3791–3801.

Brede, D.A., Faye, T., Johnsborg, O., Ødegård, I., Nes, I.F., and Holo, H. (2004) Molecular and genetic characterization of propionicin F, a bacteriocin from *Propionibacterium freudenreichii*. *Appl Environ Microbiol* **70:** 7303–7310.

Butcher, B.G., Lin, Y.P., and Helmann, J.D. (2007) The *yydFGHIJ* operon of *Bacillus subtilis* encodes a peptide that induces the LiaRS two-component system. *J Bacteriol* **189:** 8616–8625.

Chaussee, M.S., Ajdic, D., and Ferretti, J.J. (1999) The *rgg* gene of *Streptococcus pyogenes* NZ131 positively influences extracellular SPE B production. *Infect Immun* **67:** 1715–1722.

Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Smoot, L.M., Graham, M.R., Watson, R.O., and Musser, J.M. (2002) Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*. *Infect Immun* **70:** 762–770.

Chaussee, M.S., Somerville, G.A., Reitzer, L., and Musser, J.M. (2003) Rgg coordinates virulence factor synthesis and metabolism in *Streptococcus pyogenes*. *J Bacteriol* **185:** 6016–6024.

Claverys, J.P., and Håvarstein, L.S. (2002) Extracellular-peptide control of competence for genetic transformation in *Streptococcus pneumoniae*. *Front Biosci* **7:** 1798–1814.

Claverys, J.P., Prudhomme, M., and Martin, B. (2006) Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* **60:** 451–475.

Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., *et al*. (2007) Structure of PlcR: insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proc Natl Acad Sci USA* **104:** 18490–18495.

Denham, E.L., Ward, P.N., and Leigh, J.A. (2008) Lipoprotein signal peptides are processed by Lsp and Eep of *Streptococcus uberis*. *J Bacteriol* **190:** 4641–4647.

Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C., Yu, W.H., *et al*. (2010) The human oral microbiome. *J Bacteriol* **192:** 5002–5017.

Dmitriev, A.V., McDowell, E.J., Kappeler, K.V., Chaussee, M.A., Rieck, L.D., and Chaussee, M.S. (2006) The Rgg regulator of *Streptococcus pyogenes* influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon. *J Bacteriol* **188:** 7230–7241.

Dunny, G.M. (2007) The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc Lond B Biol Sci* **362:** 1185–1193.

Fernandez, A., Borges, F., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2006) The *rggC* locus, with a frameshift mutation, is involved in oxidative stress response by *Streptococcus thermophilus*. *Arch Microbiol* **186:** 161–169.

Fontaine, L., Boutry, C., Henry de Frahan, M., Delplace, B., Fremaux, C., Horvath, P., *et al*. (2010) A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol* **192:** 1444–1454.

Fouquier d'Hérouel, A., Wessner, F., Halpern, D., Ly-Vu, J., Kennedy, S.P., Serror, P., *et al*. (2011) A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen Enterococcus faecalis. *Nucleic Acids Res* (in press).

Fujiwara, T., Hoshino, T., Ooshima, T., Sobue, S., and Hamada, S. (2000) Purification, characterization, and molecular analysis of the gene encoding glucosyltransferase from *Streptococcus oralis*. *Infect Immun* **68:** 2475–2483.

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.

Gardan, R., Besset, C., Guillot, A., Gitton, C., and Monnet, V. (2009) The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. *J Bacteriol* **191:** 4647–4655.

Gasson, M.J. (1983) Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. *J Bacteriol* **154:** 1–9.

Goebl, M., and Yanagida, M. (1991) The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* **16:** 173–177.

Ibrahim, M., Nicolas, P., Bessières, P., Bolotin, A., Monnet, V., and Gardan, R. (2007a) A genome-wide survey of short coding sequences in streptococci. *Microbiology* **153:** 3631–3644.

Ibrahim, M., Guillot, A., Wessner, F., Algaron, F., Besset, C., Courtin, P., *et al*. (2007b) Control of the transcription of a short gene encoding a cyclic peptide in *Streptococcus thermophilus*: a new quorum-sensing system? *J Bacteriol* **189:** 8844–8854.

Karpenahalli, M.R., Lupas, A.N., and Söding, J. (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8:** 2.

Lereclus, D., Agaisse, H., Gominet, M., Salamitou, S., and Sanchis, V. (1996) Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. *J Bacteriol* **178:** 2749–2756.

Letort, C., and Juillard, V. (2001) Development of a minimal chemically-defined medium for the exponential growth of *Streptococcus thermophilus*. *J Appl Microbiol* **91:** 1023–1029.

Linton, K.J., and Higgins, C.F. (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch* **453:** 555–567.

Loughman, J.A., and Caparon, M.G. (2007) Contribution of invariant residues to the function of Rgg family transcription regulators. *J Bacteriol* **189:** 650–655.

Lyon, W.R., Gibson, C.M., and Caparon, M.G. (1998) A role for trigger factor and an Rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of *Streptococcus pyogenes*. *EMBO J* **17:** 6263–6275.

Makarova, K., Slevarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., *et al*. (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* **103:** 15611–15616.

Mashburn-Warren, L., Morrison, D.A., and Federle, M.J. (2010) A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol Microbiol* **78:** 589–606.

Neely, M.N., Lyon, W.R., Runft, D.L., and Caparon, M. (2003) Role of RopB in growth phase expression of the SpeB cysteine protease of *Streptococcus pyogenes*. *J Bacteriol* **185:** 5166–5174.

Novick, R.P., and Geisinger, E. (2008) Quorum sensing in staphylococci. *Annu Rev Genet* **42:** 541–564.

Paradis, E., Claude, J., and Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20:** 289–290.

Podbielski, A., and Kreikemeyer, B. (2004) Cell density – dependent regulation: basic principles and effects on the virulence of Gram-positive cocci. *Int J Infect Dis* **8:** 81–95.

Pottathil, M., and Lazazzera, B.A. (2003) The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Front Biosci* **8:** 32–45.

Pulliainen, A.T., Hytönen, J., Haataja, S., and Finne, J. (2008) Deficiency of the Rgg regulator promotes H_2O_2 resistance, AhpCF-mediated H_2O_2 decomposition, and virulence in *Streptococcus pyogenes*. *J Bacteriol* **190:** 3225–3235.

Qi, F., Chen, P., and Caufield, P.W. (1999) Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in *Streptococcus mutans*. *Appl Environ Microbiol* **65:** 652–658.

Rawlinson, E.L., Nes, I.F., and Skaugen, M. (2002) LasX, a transcriptional regulator of the lactocin S biosynthetic genes in *Lactobacillus sakei* L45, acts both as an activator and a repressor. *Biochimie* **84:** 559–567.

Rawlinson, E.L., Nes, I.F., and Skaugen, M. (2005) Identification of the DNA-binding site of the Rgg-like regulator LasX within the lactocin S promoter region. *Microbiology* **151:** 813–823.

Rea, M.C., Sit, C.S., Clayton, E., O'Connor, P.M., Whittal, R.M., Zheng, J., *et al*. (2010) Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc Natl Acad Sci USA* **107:** 9352–9357.

Rocha-Estrada, J., Aceves-Diez, A.E., Guarneros, G., and de la Torre, M. (2010) The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Appl Microbiol Biotechnol* **87:** 913–923.

Sambrook, J., and Russell, D.W. (2001) Molecular cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Samen, U.M., Eikmanns, B.J., and Reinscheid, D.J. (2006) The transcriptional regulator RovS controls the attachment of *Streptococcus agalactiae* to human epithelial cells and the expression of virulence genes. *Infect Immun* **74:** 5625–5635.

Sanders, J.W., Leenhouts, K., Burghoorn, J., Brands, J.R., Venema, G., and Kok, J. (1998) A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol Microbiol* **27:** 299–310.

Shi, K., Brown, C.K., Gu, Z.Y., Kozlowicz, B.K., Dunny, G.M., Ohlendorf, D.H., and Earhart, C.A. (2005) Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in *Enterococcus faecalis*. *Proc Natl Acad Sci USA* **102:** 18596–18601.

Siehnel, R., Traxler, B., An, D.D., Parsek, M.R., Schaefer, A.L., and Singh, P.K. (2010) A unique regulator controls the activation threshold of quorum-regulated genes in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **107:** 7916–7921.

Skaugen, M., Andersen, E.L., Christie, V.H., and Nes, I.F. (2002) Identification, characterization, and expression of a second, bicistronic, operon involved in the production of lactocin S in *Lactobacillus sakei* L45. *Appl Environ Microbiol* **68:** 720–727.

Slamti, L., and Lereclus, D. (2002) A cell-cell signaling

peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *EMBO J* **21:** 4550–4559.
Sulavik, M.C., Tardif, G., and Clewell, D.B. (1992) Identification of a gene, *rgg*, which regulates expression of glucosyltransferase and influences the Spp phenotype of *Streptococcus gordonii* Challis. *J Bacteriol* **174:** 3577–3586.
Trieu-Cuot, P., Carlier, C., Poyart-Salmeron, C., and Courvalin, P. (1990) A pair of mobilizable shuttle vectors conferring resistance to spectinomycin for molecular cloning in *Escherichia coli* and in gram-positive bacteria. *Nucleic Acids Res* **18:** 4296.
Vickerman, M.M., and Minick, P.E. (2002) Genetic analysis of the *rgg-gtfG* junctional region and its role in *Streptococcus gordonii* glucosyltransferase activity. *Infect Immun* **70:** 1703–1714.
Vickerman, M.M., Wang, M., and Baker, L.J. (2003) An amino acid change near the carboxyl terminus of the *Streptococcus gordonii* regulatory protein Rgg affects its abilities to bind DNA and influence expression of the glucosyltransferase gene *gtfG*. *Microbiology* **149:** 399–406.
Vickerman, M.M., Flannagan, S.E., Jesionowski, A.M., Brossard, K.A., Clewell, D.B., and Sedgley, C.M. (2010) A genetic determinant in *Streptococcus gordonii* Challis encodes a peptide with activity similar to that of enterococcal sex pheromone cAM373, which facilitates intergeneric DNA transfer. *J Bacteriol* **192:** 2535–2545.
Waters, C.M., and Bassler, B.L. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **21:** 319–346.
Williams, P., Winzer, K., Chan, W.C., and Cámara, M. (2007) Look who's talking: communication and quorum sensing in the bacterial world. *Philos Trans R Soc Lond B Biol Sci* **362:** 1119–1134.
Winzer, K., Hardie, K.R., and Williams, P. (2002) Bacterial cell-to-cell communication: sorry, can't talk now – gone to lunch! *Curr Opin Microbiol* **5:** 216–222.
Zheng, G., Hehn, R., and Zuber, P. (2000) Mutational analysis of the *sbo-alb* locus of *Bacillus subtilis*: identification of genes required for subtilosin production and immunity. *J Bacteriol* **182:** 3266–3273.
Zheng, F., Ji, H., Cao, M., Wang, C., Feng, Y., Li, M., *et al.* (2011) Contribution of the Rgg transcription regulator to metabolism and virulence of *Streptococcus suis* serotype 2. *Infect Immun* **79:** 1319–1328.

Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

3. Conclusion

Le travail présenté dans l'article 1 a permis de décrire, sur la base du locus *shp/rgg1358* de *S. thermophilus* LMD-9, un nouveau mécanisme de QS impliquant deux gènes transcrits de manière divergente et codant un petit peptide hydrophobe jouant le rôle de phéromone (SHP) et un régulateur transcriptionnel appartenant à la famille des Rgg.

Pour valider ce mécanisme de QS, une approche génétique a tout d'abord été mise en œuvre afin de mesurer l'activité du promoteur $P_{shp1358}$ dans différentes conditions de croissance (cultures, co-cultures, etc.). L'ensemble des résultats obtenus a permis de valider les trois premières étapes de ce nouveau mécanisme de régulation par QS à savoir : la sécrétion, la détection et la réimportation de la phéromone SHP1358. La forme mature de la phéromone a également été identifiée par LC-MS/MS et confirmée par complémentation fonctionnelle du mutant Δ*shp*. De plus, un rôle indirect de la metalloprotéase Eep dans la maturation et/ou la sécrétion de la phéromone a également été démontré.
La spécificité de l'interaction de la phéromone SHP1358 avec la protéine régulatrice Rgg1358 a été démontrée par SPR. Enfin, la technique de retard sur gel a permis d'étudier la fixation de la protéine régulatrice Rgg1358 au niveau des régions promotrices des gènes cibles du système (*shp1358* et *ster_1357*), présentant un motif ADN conservé. Les +1 de transcription des deux gènes cibles du système étudié ont par ailleurs été déterminés en utilisant la technique de 5'tag-RACE. L'ensemble des résultats valide l'existence d'un nouveau système de communication de type QS chez les streptocoques. Néanmoins, bien que la présence des deux partenaires (phéromone et protéine régulatrice) soit nécessaire à l'activation des gènes cibles connus du système, les résultats obtenus indiquent que la présence de la phéromone SHP1358 n'influence pas la fixation de la protéine régulatrice Rgg1358 à l'ADN dans les conditions testées. Aussi, l'hypothèse d'un changement de conformation et/ou la modification de l'état d'activation de la protéine régulatrice Rgg1358 en présence de la phéromone a été soulevée dans cet article.

Enfin, la construction d'un arbre phylogénétique des protéines Rgg et Rgg-like a permis de distinguer les systèmes SHP/Rgg et XIP/ComR. Aussi, ces résultats vont dans le sens de deux familles de systèmes de QS différentes. Par ailleurs, l'analyse des peptides issus des données *in silico* a révélé l'existence d'un troisième groupe de système SHP/Rgg.

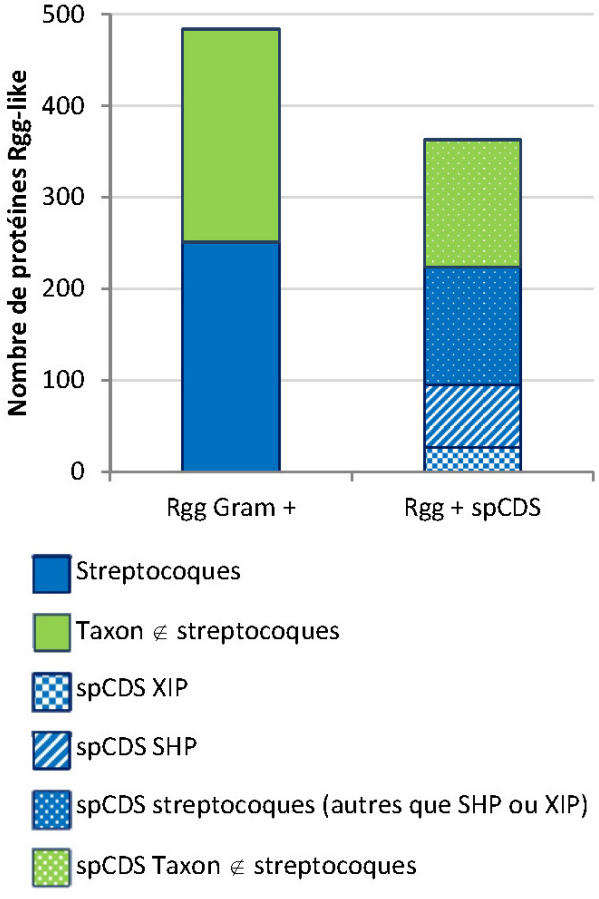


Figure 43 – Nombre de protéines Rgg et Rgg-like, associées ou non à des spCDS, chez les bactéries à Gram positif à bas GC% appartenant au taxon Rgg.

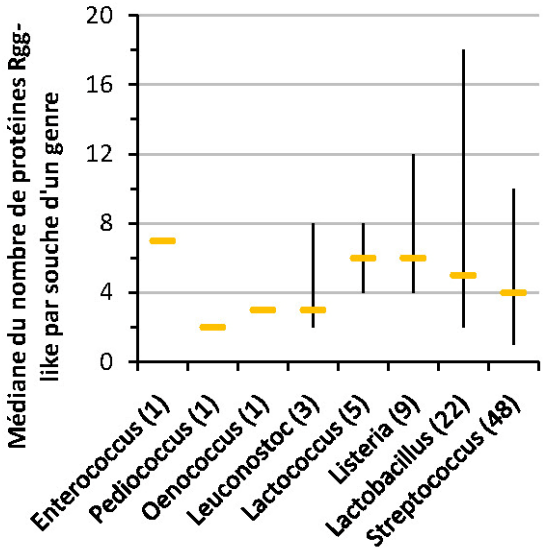


Figure 44 – Nombre de protéines Rgg et Rgg-like, associées ou non à des spCDS, au sein des différents genres bactériens appartenant au taxon Rgg.

Entre parenthèses est précisé le nombre de souches séquencées disponibles au moment de l'analyse dans le genre considéré. La barre verticale indique respectivement le nombre minimal et maximal de protéines Rgg-like identifiées au sein des souches d'un genre.

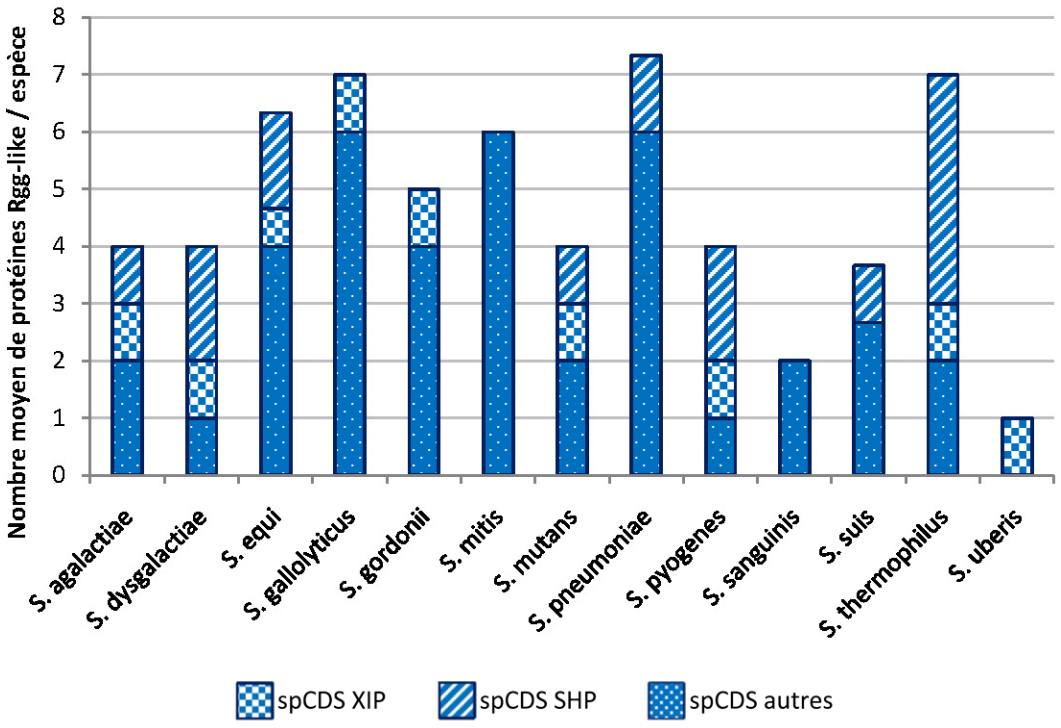


Figure 45 – Nombre de protéines Rgg et Rgg-like, associées à des spCDS, au sein du genre *Streptococcus*.

4. Résultats non publiés

4.1 Introduction

Afin de répondre aux objectifs liés à la construction de l'arbre phylogénétique, et comme décrit dans l'article 1, un taxon Rgg a tout d'abord été défini. Au sein de ce taxon, composé de l'ordre des *Lactobacillales* et de la famille des *Listeriaceae*, 484 protéines Rgg-like ont été identifiées (Figure 43). La répartition de ces dernières au sein des différents genres bactériens composant le taxon s'est révélée être relativement homogène (Figure 44) et représentative de la quantité de génomes séquencés. Ceci est représenté pour le genre *Streptococcus* sur la Figure 43 ; en effet, 52% des protéines Rgg et Rgg-like proviennent des 48 génomes de streptocoques, représentant 53% du nombre total des génomes disponibles pour l'analyse. Néanmoins, une hétérogénéité inter-souche peut être soulignée dans certains cas et notamment au sein du genre *Lactobacillus* regroupant des souches pouvant contenir de 2 à 18 protéines Rgg-like dans leur génome (Figure 44). A partir du taxon identifié, une recherche des petites phases ouvertes de lecture putatives (spCDS) dans l'environnement proche des protéines Rgg ou Rgg-like a été mise en place et 363 protéines se sont avérées être associées à une ou plusieurs spCDS (Figure 43). Aussi, l'association d'une protéine Rgg-like avec une ou plusieurs spCDS s'avère être plus fréquente au sein des streptocoques (89%) que dans le reste du taxon (60%). Parmi ces spCDS et sur la base des critères définis dans le Tableau 12, 68 ont été considérées comme codant un peptide SHP et 27 comme codant un peptide XIP. La Figure 45 représente le nombre de peptides SHP et XIP répertoriés pour chaque espèce de streptocoques. Les résultats soulignent l'accumulation des systèmes SHP/Rgg au sein de l'espèce *S. thermophilus* et l'absence des systèmes XIP/ComR au sein des streptocoques du groupe mitis (*S. pneumoniae*, *S. mitis*, *S. sanguinis*) possédant le système à deux composants ComDE pour le déclenchement de la compétence. Enfin, sur l'ensemble du taxon Rgg, 74% des peptides codés par des spCDS localisées dans l'environnement proche de protéines Rgg-like n'ont pas été décrits dans l'article 1 (Figure 43). Une analyse complémentaire centrée sur ces peptides est présentée dans cette partie.

4.2 Matériel et méthode

A partir du répertoire des protéines Rgg et Rgg-like obtenues lors de l'analyse *in silico* décrite dans l'article 1, une recherche de nouvelles familles de peptides associés aux protéines régulatrices Rgg et Rgg-like a été réalisée. Dans cette analyse, seulement les peptides n'ayant pas été décrits comme SHP ou XIP ont été pris en compte. Afin de mettre

Tableau 12 – Caractéristiques des peptides « phéromones » SHP et XIP.

	SHP (Ibrahim *et al.*, 2007b)	**XIP** (Mashburn-Warren *et al.*, 2010)
Séquence des peptides « phéromones » publiés	MKKINKALLFTLIMDILIIVGG MKKVNKALLFTLIMDILIIVGG MKKISKFLPILILAMDIIIIVGGEVT MKKISKFLPILILAMDIIIIVGG MKKISKFLPILFLVMDIIIIVGG MEKVSKILPILILVMDIIIIVGG MKKQKLLLLVVLVCEGIIVILVG MNKESFLAILLLIFESIIVIAVG MKKQILLTLLLVVFEGIIVIVVG MRNKIFMTLIVVLETIIIIGGG MKLLKIIVLLTCIYTIVGGV MKLLKIIVLLTCIYIIVGGV	MKTLKIFVLFSLLIAILPYFAGCL MTLVIKLVGTLLTMGWWGL MFKVFFTVMTGVFWWG MFSILTSILMGLDWWSL MLKKYKYYFIFAALLSFKVVQELSAVDWWRL MLKKVKPFLLLAAVVAFKVARVMHEFDWWNLG MFKKYQYYLFLAALFLLHSAQLLSDIDWWRVG MFKRYHYYFILTAMLAFKAAQMISQVDWWRL MLNIFSIVITGWWGL MFKKIHFYVTTFSFLAVALITFLSEKDWWHIG
Organisation génétique		
Localisation du gène codant le peptide	. En amont du gène *rgg*	. En aval du gène *rgg*
Brin portant le gène codant le peptide	. Région codante sur le brin opposé	. Région codante sur le même brin
Orientation des gènes	. Transcrits de manière divergente	. Transcrits de manière convergente
Séquence peptidique		
Taille en aa	. Taille homogène 20 à 26	. Taille hétérogène 15 à 32
Hydrophobicité (Gravy)	. Compris entre 1.355 à 2.205 (moy 1.70)	. XIP (31/32 aa) moy = 0.4932, XIP (15 à 24aa) moy = 1.4538
Charge globale	. Une charge nette positive	. Une charge nette positive
Partie N-ter	. Présence de 2 résidus basiques*	. Présence de 2 résidus basiques*
Partie centrale	. Asp (D) ou Glu (E) au milieu d'un stretch hydrophobe composé de Ile (I), Leu (L), Val (V) et Met (M)	.
Partie C-ter	.	. Motif double Trp (WW) conservé
	. Présence de Gly (G)	. Présence de Gly (G)

* en général des lysines (K) ; en noir sont indiqués les critères retenus pour la sélection ; « moy » signifie moyenne.

en évidence d'autres familles de peptides potentiellement secrétés ou « phéromones », une analyse basée sur des critères définis a été mise en place. En effet, la comparaison des peptides SHP et XIP (Tableau 12) a souligné que ces deux familles de peptides – malgré leurs différences – partagent certaines similarités probablement en lien avec leur mécanisme de maturation, d'export, de sécrétion ou de réimportation. Aussi, les critères communs et spécifiques de ces deux familles de peptides (présence de lysine en N-ter, hydrophobicité dans la partie centrale, présence d'acides aminés conservés – aromatiques, aspartate, glutamate – en C-ter) ont servi de critères de sélection pour la détection et l'élaboration de nouvelles familles.

4.3 Résultats principaux

- **Identification de deux nouvelles familles au sein du groupe mitis**

La recherche de nouvelles familles sur la base de critères connus et définis dans le Tableau 12 a fait ressortir deux familles putatives de peptides détectés au sein de deux espèces de streptocoques appartenant au groupe mitis, *S. pneumoniae* et *S. mitis*.

Les peptides appartenant à la première famille appelée « mitis 1 » sont retrouvés dans 8 des 12 souches séquencées de *S. pneumoniae* ainsi que dans l'unique souche séquencée de *S. mitis* disponible au moment de l'analyse. Les caractéristiques générales de ces peptides, nommés $peptides_{SHOW_mitis1}$, sont présentées dans le Tableau 13. Ces derniers présentent plusieurs critères communs à ceux décrits pour les familles SHP et XIP. Leur principale particularité, au même titre que les peptides SHP, est la présence d'un aspartate ou d'un glutamate conservé dans la partie C-terminale. De manière intéressante, la nature de l'acide aminé conservé semble être corrélée à l'espèce de streptocoque possédant le peptide. Aussi, tous les $peptides_{SHOW_mitis1}$ issus de l'espèce *pneumoniae* contiennent un glutamate et le $peptide_{SHOW_mitis1}$ issu de l'espèce *mitis*, un aspartate. Enfin, une mauvaise identification des démarrages de traduction de ces peptides peut également être soulevée. En effet, au regard des séquences peptidiques détectées, un deuxième démarrage de traduction peut être envisagé au niveau des acides aminés MKK (Tableau 13). Si tel est le cas, les $peptides_{SHOW_mitis1}$ partageraient de très fortes similarités de séquence avec les peptides SHP connus tout en préservant une organisation génétique différente de celles déjà décrites pour les groupes I, II et III (cf. article 1). En effet, les spCDS codant ces peptides sont localisées en amont et transcrites dans le même sens que le gène *rgg*-like. Enfin, la stricte conservation des séquences peptidiques appartenant à l'espèce *pneumoniae* indique une absence d'évolution des spCDS codant ces derniers pouvant être due à des contraintes fonctionnelles précises ou à une acquisition récente de ces petits gènes au sein de l'espèce.

Tableau 13 – Familles de peptides$_{SHOW}$ identifiées dans le groupe mitis et le genre *Listeria*.

		Gène *rgg*-like[a]	Peptide détecté par le logiciel SHOW			
			Localisation[b]		Distance[c]	Séquence
Famille mitis 1[d]	*Streptococcus pneumoniae* D39	*SPD_1952*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae JJA	*SPJ_2147**	1	amont		MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae Taiwan19F-14	*SPT_2134**	1	amont		MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae G54	*SPG_2060*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae CGSP14	*SPCG_2091*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae TCH8431/19A	*HMPREF0837_10124*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae 70585	*SP70585_2248**	1	amont		MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus pneumoniae Hungary19A-6	*SPH_2314*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**IRNQG**E**VYGAD**W**KSL
	Streptococcus mitis B6	*smi_0116*	1	amont	C	MINILYFLIILTIWQVFDEFSEKYDK**MKK**ILNQG**D**AYGAD**W**KSL
Famille mitis 2	*Streptococcus pneumoniae* R6	*spr0104**	-1	amont	(183 pb)	**MNKK**IFCILVCIILLISLAIIFPW**G**W**W**PI
	Streptococcus pneumoniae D39	*SPD_0112*	-1	amont	183 pb	**MNKK**IFCILVCIILLISLAIIFPW**G**W**W**PI
	Streptococcus pneumoniae JJA	*SPJ_0139*	-1	amont	183 pb	**MNKK**IFCILVCIISLISLAIIFPW**G**W**W**PI
	Streptococcus pneumoniae Taiwan19F-14	*SPT_0150*	-1	amont	183 pb	**MNKK**IFCILVCIISLISLAIIFPW**G**W**S**I
	Streptococcus pneumoniae CGSP14	*SPCG_0110**	-1	amont	(183 pb)	**MNKK**IFCILVCIISLISLAIIFPW**G**W**W**PI
	Streptococcus pneumoniae TCH8431/19A	*HMPREF0837_10406*	-1	amont	183 pb	**MNKK**IFCILVCIISLISLAIIFPW**G**W**S**I
	Streptococcus pneumoniae 70585	*SP70585_0179*	-1	amont	183 pb	**MNKK**IFCILVCIISLISLAIIFPW**G**W**S**I
	Streptococcus mitis B6	*smi_1977*	-1	amont	184 pb	**MNKK**VFRFLVVLAFLFSIATVFPW**G**W**P**I
Famille Listeria	*Listeria innocua* Clip11262	*lin2541*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLFKSIN**W**LF
	Listeria seeligeri serovar 1/2b str. SLCC3954	*lse_2346*	-1	amont	76 pb	(**LKK**EGKN)MLTFLTSIN**W**IF
	Listeria monocytogenes EGD-e	*lmo2447*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLLKSIN**W**LF
	Listeria monocytogenes serotype 4b str. F2365	*LMOf2365_2420*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLLKSIN**W**LF
	Listeria monocytogenes serotype 4b str. CLIP 80459	*Lm4b_02416*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLLKSIN**W**LF
	Listeria monocytogenes HCC23	*LMHCC_0153*	-1	amont	76 pb	(**MKK**EGE**D**)MLTFLTSVN**W**LF
	Listeria monocytogenes 08-5923	*LM5923_2592*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLLKSIN**W**LF
	Listeria monocytogenes 08-5578	*LM5578_2643*	-1	amont	76 pb	(**MKK**EGE**D**)MLTLLKSIN**W**LF

[a] Les annotations sont issues de la base de données GenBank ; [b] Les termes « -1 » et « 1 » indiquent respectivement une localisation amont ou avale de la spCDS codant le peptide détecté par rapport au gène *rgg* de référence ; [c] Distance nucléotidique entre les ATG de la spCDS et le gène *rgg* de référence, donnée précisée uniquement quand les gènes sont divergents ; [d] La majorité des peptides de cette famille sont annotés dans GenBank. L'astérisque indique une annotation du démarrage de transcription du *rgg* différente de celle identifiée pour la majorité des gènes *rgg* du groupe. La lettre « C » indique que les parties 3' de la spCDS et 5' du gène *rgg* se chevauchent.

Les peptides appartenant à la seconde famille appelée « mitis 2 » sont quant à eux retrouvés dans 7 des 12 souches séquencées de *S. pneumoniae* ainsi que dans l'unique souche séquencée de *S. mitis* disponible au moment de l'analyse. Les caractéristiques générales de ces peptides, nommés $peptides_{SHOW_mitis2}$, sont présentées dans le Tableau 13. Cette famille de peptides présente différentes caractéristiques telles que la présence de résidus basiques en N-terminal, d'un stretch hydrophobe au niveau de leur partie centrale et d'acides aminés aromatiques (une phénylalanine et deux tryptophanes) en C-terminal.

- **Identification d'une nouvelle famille de peptides spécifique du genre *Listeria*.**

Sur la base des similarités de séquences des peptides codés par les spCDS, une famille de peptides a pu être distinguée à l'échelle du genre bactérien. Ces peptides, identifiés au sein du genre *Listeria*, sont représentés dans 8 des 9 souches séquencées disponibles au moment de l'analyse et présents à la fois dans les génomes des souches pathogènes et non pathogènes. Les caractéristiques générales de ces peptides, nommés $peptides_{SHOW_listeria}$, sont présentées dans le Tableau 13. De manière inattendue, l'analyse de ces loci a révélé une possible erreur d'annotation des démarrages de traduction de ces peptides. En effet, une séquence peptidique conservée en N-terminal, correspondant à MKKEGED, attribuerait à ces peptides de nouveaux critères ; telle qu'une sécrétion potentielle du à la présence de lysines et la présence d'un aspartate conservé (Tableau 13).

- **Identification de peptides « souche spécifique »**

Dans cette étude, ont été considérés comme peptides « souche specifique » tous les peptides présentant des caractéristiques intéressantes (Tableau 12) mais détectés dans moins de la moitié des souches séquencées d'une espèce considérée ou dans toute espèce ne présentant qu'une souche séquencée au moment de l'analyse. Les caractéristiques générales de ces peptides sont présentées dans le Tableau 14.

Les peptides « souche specifique » ont été identifiés au sein de 4 espèces appartenant au groupe des streptocoques (*pneumoniae*, *mutans*, *sanguinis* et *equi*). Tous possèdent des lysines dans la partie N-terminale de leur séquence peptidique. Le peptide retrouvé chez *S. pneumoniae* à la particularité d'être riche en acides aminés aromatiques et plus particulièrement en tryptophane. Le peptide identifié chez *S. sanguinis* présente également cette particularité mais de manière plus localisée dans sa partie C-terminale. Les spCDS codant ces peptides isolés sont localisées en amont de la protéine régulatrice Rgg alors que celles codant les peptides isolés identifiés chez *S. mutans* et *S. equi* sont en aval du régulateur. Les séquences peptidiques de ces deux derniers peptides sont quant à elles

Tableau 14 – Peptides$_{SHOW}$ « souche spécifique » présentant des caractéristiques intéressantes.

	Peptide détecté par le logiciel SHOW		
	Localisation[b]		Séquence
Streptococcus pneumoniae 4 souches séquencées sur 12	-1	amont	MGFKKYLKNLPKNSGFLIWSWIQLIWFETWFWG
Streptococcus sanguinis SK36 1 souche séquencée	-1	amont	MKKIVYNLILLAVTSIVTTSVFPWWWLWW
Streptococcus mutans NN2025 1 souche séquencée sur 2	1	aval	MKLRKLLIAIVTFWFICGGC
Streptococcus equi subsp. zooepidemicus MGCS10565 1 souche séquencée sur 3	1	aval	MKLLHLLKIAAVVFLLIPPIIF

Tableau 15 – Identification d'une famille de peptides « sans » critère au sein de l'espèce *Streptococcus pneumoniae*.

	Peptide détecté par le logiciel SHOW		
	Localisation[b]		Séquence
R6	1	amont	MSRKDVTVTIF
D39	1	amont	MSRKDVTVTIF
TIGR4	1	amont	MSRKDVAVTIF
JJA	1	amont	MSRKDVAVTIF
Taiwan19F-14	1	amont	MSRKDVAVTIF
G54	1	amont	MSRKDVTVTIF
CGSP14	1	amont	MSRKDVAVTIF
ATCC 700669	1	amont	MSRKDVAVTIF
TCH8431/19A	1	amont	MSRKDVAVTIF
70585	1	amont	MSRKDVAVTIF
Hungary19A-6	1	amont	MSRKDVTVTIF
P1031	1	amont	MSRKDVAVTIF

riches en acides aminés hydrophobes et trois acides aminés aromatiques sont présents dans la partie C-terminale du peptide de *S. mutans*.

La localisation sur l'arbre phylogénétique des protéines Rgg et Rgg-like de l'ensemble des familles ou peptides « souche specifique » décrits dans cette partie est présentée en Annexe 3.

4.4 Discussion

L'identification de nouvelles familles de peptides est rendue difficile par la complexité du traitement des données, basé dans notre cas sur certains critères tels que la similarité de séquence ou la présence de caractères communs.

La première contrainte résulte en la bonne annotation des démarrages de traduction. En effet, il n'est pas impossible, au vu de la famille détectée au sein du groupe mitis 1 ou du genre *Listeria*, d'être passé à côté de familles présentant des critères d'intérêt. De manière générale, l'annotation des démarrages de traduction est un problème récurrent de l'annotation de gène. Dans le cadre de cette analyse *in silico*, cela avait auparavant déjà été constaté concernant le XIP de *Streptococcus uberis*. Dans ce cas, une séquence peptidique (allongeant la séquence finale du peptide) avait été ajoutée à la séquence connue et publiée (Mashburn-Warren *et al.*, 2010). Aussi, le traitement des données issues de systèmes automatisés requiert une certaine prudence et justifie la nécessité de confirmer à « l'œil » et de manière méthodique l'ensemble des données fournies. Il est important de préciser, par ailleurs, que d'autres familles de peptides – ne présentant pas de critère défini sur la base des SHP et XIP – ont été détectées lors de cette analyse. Un exemple est présenté dans le Tableau 15.

Une deuxième contrainte à prendre en compte est la notion de seuil très nettement mise en évidence au cours de l'analyse des peptides SHP et XIP. Aussi, malgré la détection de la majorité des peptides SHP et XIP connus à des seuils « stringents », l'application de seuils moins sélectifs est nécessaire à la détection de l'ensemble de ces peptides. Aussi, la complexité du répertoire de données augmente significativement, résultant en un nombre important de protéines Rgg-like bornées par une ou plusieurs spCDS pouvant être localisées à la fois en amont et en aval (cf. exemples présentés en Annexe 4). La difficulté consiste alors à savoir si toutes les spCDS détectées sont premièrement transcrites (notion de faux-positifs) et deuxièment codent des peptides possédant de réelles fonctions physiologiques. Aussi, la définition *in silico* de familles de peptides sans critère défini au préalable s'avère être un travail audacieux.

Enfin, la pertinence dans le choix des peptides à étudier et la validation expérimentale de la production et du rôle de ces peptides reste indispensable à la réelle définition d'une nouvelle famille de peptides.

Dans le cadre de cette analyse, seulement des critères communs et spécifiques aux familles SHP et XIP ont été utilisés. Néanmoins, afin d'être le plus exhaustif possible dans la recherche de peptides potentiellement sécrétés ou « phéromones », d'autres critères décrits dans la littérature aurait pu être pris en compte. En particulier, la recherche de signaux d'exportation (Chandler and Dunny, 2008; Lanigan-Gerdes *et al.*, 2007; Vickerman *et al.*, 2010), la présence d'un motif double glycine (Annexe 5) (Håvarstein *et al.*, 1994; Håvarstein *et al.*, 1996) ou encore la recherche dans les séquences signales de protéines secrétées.

Aucun peptide transversal à l'ensemble du taxon n'ayant été identifié, chaque partenaire peptide/protéine régulatrice semble s'associer de manière spécifique, ceci soit au niveau d'un genre comme décrit pour les SHP et XIP chez les streptocoques ou pour les $peptides_{SHOW_listeria}$, soit au niveau de l'espèce. Au sein de cette dernière configuration *S. pneumoniae* semble être la bactérie ayant accumulé le plus grand nombre de systèmes impliquant des protéines Rgg-like au cours de son l'évolution. Néanmoins, le nombre de souches séquencées n'étant pas homogène entre les différentes espèces de streptocoques, cette observation sera probablement remise en cause à plus ou moins long terme.

Une dernière interrogation serait de savoir si le terme « Rgg » est utilisé dans cette étude à bon escient. La proximité des protéines Rgg avec les régulateurs de la famille RNPP et de la famille ComR (Rgg-like), laisse présager de nombreuses difficultés d'annotation et d'identification. En effet, il est intéressant de relever que les protéines « Rgg-like » associées à la famille de peptides, présentée dans le Tableau 15, sont annotées PlcR. De plus, les antiphéromones d'*E. faecalis* associées au régulateur PrgX ont également été détectées par cette analyse. Ceci pause alors clairement la question : comment définir et classer précisément les régulateurs Rgg, ComR et de la famille RNPP ? Des solutions seront proposées dans la discussion générale.

Chapitre 2 : Etude de l'interaction SHP/Rgg

Fleuchot B.[1,2], Besset C.[1,2], Bruneau E.[3], Vidic J.[4], Gitton C.[1,2], Guillot A.[1,2], Monnet V.[1,2], Gardan R.[1,2]

[1]INRA, UMR1319 MICALIS, F-78352 Jouy en Josas, France.
[2]AgroParistech, UMR MICALIS, F-78352 Jouy en Josas, France.
[3]INRA, UMR1128, IFR110, Nancy-Université, Laboratoire de Génétique et Microbiologie, BP239, 54506 Vandœuvre-lès-Nancy, France.
[4]INRA, UR892 Virologie Immunologie Moléculaire, F-78352 Jouy en Josas, France.

1. Introduction

La construction d'un arbre phylogénétique des protéines Rgg et Rgg-like a fait ressortir l'existence de trois groupes de systèmes SHP/Rgg. Dans cette étude, nous nous sommes attachés, dans un premier temps, à démontrer que d'autres loci (appartenant aux groupes I et II) fonctionnent selon le même mécanisme de QS que celui décrit sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9. Aussi, la fonctionnalité de trois nouveaux systèmes appartenant à des espèces de streptocoques pathogènes (*S. agalactiae* NEM316 et *S. mutans* UA159) et GRAS (*S. thermophilus* LMG18311) a été analysée dans un système hétérologue. Pour ceci, des fusions transcriptionnelles couplant le promoteur du gène *shp* de chaque locus à un système rapporteur *luxAB* ont été construites et intégrées dans le chromosome de *S. thermophilus* LMD-9, exprimant ou non de manière endogène les gènes *shp* et *rgg* du locus apparenté. L'activité des promoteurs *shp* a ensuite été mesurée dans un contexte sauvage et délété du système de transport AmiCDEF (contrôle négatif). La détermination des formes matures des phéromones SHP de ces trois systèmes a par ailleurs été entreprise par LC-MS/MS.

L'arbre phylogénétique des protéines Rgg-like, présenté dans l'article 1, a mis en évidence une coévolution des protéines régulatrices Rgg et des phéromones SHP. Aussi, la similarité des séquences en acides aminés, des protéines Rgg et des phéromones SHP associées, soulève la question de phénomènes de cross-talk ou de cross-inhibition entre les streptocoques. Ces phénomènes ont déjà été décrits pour d'autres genres bactériens, tels que *Bacillus* (Bouillaut *et al.*, 2008 ; Slamti et Lereclus, 2005) ou *Staphylococcus* (Wright *et al.*, 2005). Dans ces études, l'analyse des séquences des phéromones et régulateurs

Tableau 16 – Description des souches utilisées dans cette étude.

Souches et génotypes		Résistance[a]	Description[b]	Références
***Streptococcus thermophilus* LMD-9 et dérivés**				
LMD-9	*Wild-type*			Makarova *et al.*, 2006
TIL1345	LMD-9 *blp*::P*shp::*wt_*agalactiae::aphA3*	Km	pGICB004::*aphA3*::P*shp::*wt_*agalactiae* → LMD-9	Cette étude
TIL1380	LMD-9 *blp*::P*shp::*Δ*rgg_agalactiae*::*aphA3*	Km	pGICB004::*aphA3*::P*shp::*Δ*rgg_agalactiae* → LMD-9	Cette étude
TIL1381	*amiCDE::erm blp*::P*shp::*wt_*agalactiae::aphA3*	Km/Erm	TIL1389 DNA → TIL1345	Cette étude
TIL1382	LMD-9 *blp*::P*shp::*Δ*shp_agalactiae::aphA3*	Km	pGICB004::*aphA3*::P*shp::*Δ*shp_agalactiae* → LMD-9	Cette étude
TIL1383	LMD-9 *blp*::P*shp::*wt_*mutans::aphA3*	Km	pGICB004::*aphA3*::P*shp::*wt_*mutans* → LMD-9	Cette étude
TIL1384	LMD-9 *blp*::P*shp::*Δ*shp_mutans::aphA3*	Km	pGICB004::*aphA3*::P*shp::*Δ*shp_mutans* → LMD-9	Cette étude
TIL1385	*amiCDE::erm blp*::P*shp::*wt_*mutans::aphA3*	Km/Erm	TIL1389 DNA → TIL1383	Cette étude
TIL1386	LMD-9 *blp*::P*shp::*Δ*rgg_mutans*::*aphA3*	Km	pGICB004::*aphA3*::P*shp::*Δ*rgg_mutans* → LMD-9	Cette étude
TIL1389	*amiCDE::erm*	Erm	PCR fragment *amiCDE::erm* → LMD-9	Cette étude
TILA	LMD-9 *blp*::P*shp::*wt_*thermophilus*LMG*::aphA3*	Km	pGICB004::*aphA3*::P*shp::*wt_*thermophilus*LMG → LMD-9	Cette etude
TILB	LMD-9 *blp*::P*shp::*Δ*rgg_thermophilus*LMG::*aphA3*	Km	pGICB004::*aphA3*::P*shp::*Δ*rgg_thermophilus*LMG → LMD-9	Cette etude
TILC	LMD-9 *blp*::P*shp::*Δ*shp_thermophilus*LMG*::aphA3*	Km	pGICB004::*aphA3*::P*shp::*wt_*thermophilus*LMG → LMD-9	Cette etude
TILD	*amiCDE::erm blp*::P*shp::*wt_*thermophilus*LMG*::aphA3*	Km/Erm	TIL1389 DNA → TILA	Cette etude
***Streptococcus thermophilus* LMG18311 et dérivés**				
LMG18311	*Wild-type*			BCCM LMG[c]
TIL1320	Δ*rgg0182*			Henry *et al.*, 2011
TIL1322	LMG18311 *blp::Pshp0182*			Henry *et al.*, 2011
TIL1324	Δ*rgg0182 blp::Pshp0182*			Henry *et al.*, 2011
***Streptococcus mutans* UA159**				
UA159	*Wild-type* ATCC 700610			ATCC
***Streptococcus agalactiae* NEM316**				
NEM316	*Wild-type*			Gaillot *et al.*, 1997
Escherichia coli				
TG1			*supE hsd*Δ*5 thi* Δ(*lac-proAB*) F'[*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15]	Gasson, 1983
Rosetta		Cm	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* pRARE	Novagen
TIL1318	Rosetta pET28a::Rgg1358	Km, Cm		Article 1
TIL1387	Rosetta pET28a::Gbs1555	Km, Cm	pET28a::Rgg1358 → Rosetta	Cette etude
TIL1388	Rosetta pET28a::Smu.1509	Km, Cm	pET28a::Rgg1358 → Rosetta	Cette étude
C41DE3			Souche dérivée de *E. coli* BL21(DE3), F- *ompT hsdS*B (rB-mB) *gal dcm met* (DE3)	Miroux et Walker, 1996
TIL1035	C41DE3 pET155::Rgg0182	Amp	pET15b::Rgg0182 → C41DE3	Henry *et al.*, 2011

a. Km, Erm et Cm indiquent respectivement une résistance à la kanamycine, l'érythromycine et le chloramphénicol.
b. Les flèches indiquent des constructions réalisées par transformation naturelle en présence d'ADN chromosomique ou plasmidique.
c. BCCM Belgian Coordinated Collections of Microorganisms, LMG Laboratory of Microbiology and Genetics, Université de Gent.

associés dans différentes souches révèle une diversité qui permet la définition de phérotypes. Aussi, au sein d'un même phérotype, les systèmes sont compatibles alors qu'ils ne le sont pas ou sont inhibiteurs entre phérotypes. Le deuxième enjeu de ce travail est alors d'explorer la spécificité des systèmes SHP/Rgg en évaluant, par SPR, l'affinité de différentes paires SHP/Rgg. Aussi, dans un premier temps, l'affinité de quatre paires SHP/Rgg légitimes (*S. thermophilus* (2), *S. agalactiae* (1) et *S. mutans* (1)) est mesurée afin d'être comparée, dans un second temps, aux affinités de paires illégitimes, associant le Rgg d'un locus avec le SHP d'un autre locus. En parallèle, une validation *in vivo* est entreprise, pour chaque locus *shp/rgg* considéré, par ajout des formes matures synthétiques des peptides phéromones légitimes et illégitimes dans des cultures de souches possédant une fusion rapportrice mais ne produisant pas de SHP.

Dans cette étude, une contrainte non négligeable a concerné la difficulté de synthèse des différentes formes matures des peptides phéromones SHP. En effet, en regard des expériences à réaliser, un niveau de pureté minimum (90%), nous paraissait nécessaire. L'ensemble des entreprises contactées pour la synthèse des peptides a, de manière systématique, rencontré une réelle difficulté de purification de ces derniers. Aussi, les dernières formes matures synthétiques de ces peptides venant d'être réceptionnées, après 7 mois d'attente, les résultats concernant le deuxième objectif de ce travail ne pourront être présentés dans cette partie. Les expériences devraient néanmoins être réalisées d'ici la date de soutenance.

2. Matériel et méthode

- **Souches bactériennes et conditions de croissance**

Les souches utilisées dans cette étude sont listées dans le Tableau 16. Pour chacun des quatre loci *shp/rgg* étudiés, quatre souches ont été construites, permettant de suivre l'expression des promoteurs P_{shp} dans un contexte sauvage, délété du gène *shp*, *rgg* ou de l'opéron *ami*. Les souches de *S. thermophilus* ont été cultivées à 30, 37 ou 42°C et celles de *S. agalactiae* et *S. mutans* à 30°C, en milieu M17 complémenté avec 10 $g.L^{-1}$ de lactose (M17lac) ou en MCD sans peptide (Letort and Juillard, 2001). Les souches de *E. coli* ont été cultivées à 30 ou 37°C en milieu Luria-Bertani (LB) sous agitation. De l'agar à 1,5% a été ajouté aux différents milieux quand nécessaire. De même, lorsque cela a été requis, les cultures ont été supplémentées en antibiotiques aux concentrations suivantes : kanamycine, 30 µg/mL pour *E. coli* et 1000 µg/mL pour *S. thermophilus* ; érythromycine, 5 µg/mL pour *S. thermophilus*. Le suivi de croissance a été réalisé par mesure de la densité optique des

Tableau 17 – Oligonucléotides utilisés dans cette étude.

Oligonucléotides	Séquence (5' → 3')[a,b]
AphA3-F	CCAGCGAACCATTTGAG
AphA3-R	GTTGCGGATGTACTTCAG
Erm-F	GGGACCTCTTTAGCTCCTTGG
Erm-R	GGAGATAAGACGGTTCGTGTTCG
BF1	CGGCTTTGACCTACACG
BF4	AAGTGACCTGGGCGTAC
BF21	GCATCAATCAGTTTATCTC
BF22	GAGACGGTCAGATTTTT
BF25-EcoRI	GAAGAATTCTTATCCACCAACGATAAT
BF26-SpeI	AACTACTAGTGCACTCATTAGCATTCTT
BF30-SpeI	AACTACTAGTATCTTCTTAGTGTTTTTCCT
BF31-EcoRI	GAAGAATTCATTTTTTTCATGACTATCTCCT
BF32-EcoRI	GAAGAATTCCTACCCACCGCCGATA
BF33-SpeI	AACTACTAGTAAGAGAGATTTTTATCCC
BF34-EcoRI	GAAGAATTCTTATTTCTCATGTTTTTCTC
BF35-SpeI	AACTACTAGTCCTTCATTTTTAGACCGC
BF36-NdeI	CCACCACATATGGAAAAAGAATTAG
BF37-XhoI	CCGCTCGAGCTATTAGCATTCTTTATTAT
BF38-NdeI	CCACCACATATGACTGAATATGGAA
BF39-XhoI	CCGCTCGAGCTATTATCCTTCTAAAATTT
BF40	CTACTGACAGCTTCCAAGGAGCTAAAGAGGTCCCAAGGCGATGGTTGGTAGTG
BF41	GCAAGTCAGCACGAACACGAACCGTCTTATCTCCTGAGGAAAAAGCACCAC
EB01-EcoRI	GAAGAATTCCATAAACGGGATTAACCT
EB02-EcoRI	GAAGAATTCATATGCAACCTTCTTTTAAA
EB03-EcoRI	GAAGAATTCAACAATTTCATATGATTTCC
EB04-EcoRI	GAAGAATTCTTAGACACCACCAACTAT

a. Les sites de restrictions des enzymes sont soulignés.
b. Les séquences complémentaires à la cassette kanamycine sont soulignées en pointillés.

Tableau 18 – Plasmides utilisés dans cette étude.

Plasmide	Description[a]	Références
pGICB004	Erm, Plasmide à origine de réplication thermosensible permettant l'intégration de fusions transcriptionnelles *luxAB* par double recombinaison homologue au locus *blp* de *S. thermophilus*.	Article 1
pGICB004::*aphA3*	Erm, Km, Plasmide pGICB004 contenant un insert conférant une résistance à la Km en aval des gènes luxAB et permettant l'intégration par transformation naturelle de fusions transcriptionnelles *luxAB* par double recombinaison homologue au locus *blp* de *S. thermophilus*.	Cette étude
pET28a	Km, Vecteur permettant la production de protéines recombinante présentant une queue 6-histidine.	Novagen
pET28a::Rgg1358	Dérivé du plasmide pET28 permettant la surproduction de la protéine Rgg1358 de *S. thermophilus* LMD-9	Article 1
pET28a::Gbs1555	Dérivé du plasmide pET28 permettant la surproduction de la protéine Gbs1555 de *S. agalactiae* NEM316	Cette étude
pET28a::Smu.1509	Dérivé du plasmide pET28 permettant la surproduction de la protéine Smu.1509 de *S. mutans* UA159	Cette étude
pET15b	Amp, Vecteur permettant la production de protéines recombinantes présentant une queue 6-histidine.	Novagen
pET15b::Rgg.0182	Dérivé du plasmide pET15b permettant la surproduction de la protéine Rgg.0182 de *S. thermophilus* LMG18311	Henry *et al.*, 2011

a. Km, Erm et Amp indiquent respectivement une résistance à la kanamycine, à l'érythromycine et à l'ampicilline.

cultures à 600 nm (DO_{600}) grâce à un spectrophotomètre Uvikon 931 (Kontron Instruments, UK).

- **Manipulation de l'ADN et séquençage**

Les enzymes de restriction, la T4 DNA ligase, la T4 DNA polymérase (New England Biolabs), la Taq DNA polymérase (Qbiogen) et la Phusion High-Fidelity DNA polymérase (Finnzyme) ont été utilisées selon les instructions du fournisseur. La souche *E. coli* TG1 a été utilisée comme souche hôte pour la réalisation des clonages. Les oligonucléotides nécessaires aux différentes amplifications PCR sont présentés dans le Tableau 17. Les cellules naturellement compétentes de *S. thermophilus* ont été transformées suivant le protocole décrit dans l'article 1. L'ADN chromosomique de *S. thermophilus* a été extrait par précipitation selon le protocole décrit par Gardan *et al.* en 2009.

- **Construction de souches et plasmides**

Les plasmides utilisés dans cette étude sont présentés dans le Tableau 18. Le plasmide pGICB004 permet l'introduction de fusions transcriptionnelles *luxAB* au locus *blp* de *S. thermophilus* LMD-9 par simple recombinaison homologue. Dans cette étude, un dérivé de ce plasmide contenant une cassette de résistance à la kanamycine (pGICB004::*aphA3*), a été construit dans le but d'intégrer des fusions transcriptionnelles par double recombinaison homologue au locus *blp* de *S. thermophilus* LMD-9. Les cellules de *S. thermophilus* LMD-9 naturellement compétentes ont été transformées suite à la linéarisation des plasmides pGICB004::$P_{intérêt}$::*aphA3* par l'enzyme de restriction *Sca*I. Pour ceci, la cassette *aphA3* a été amplifiée par PCR à l'aide des oligonucléotides AphA3-F et AphA3-R et intégrée au pGICB004 au niveau du site de restriction *Sma*I (Figure 46).

Les plasmides pET28a::Gbs1555 et pET28a::Smu.1509 ont été construits dans le but de surproduire puis de purifier les protéines Rgg recombinantes des systèmes de *S. agalactiae* NEM316 (Gbs1555) et *S. mutans* UA159 (Smu.1509) en présence d'une queue 6-histidine en N-terminal. Les gènes *rgg*, *gbs1555* et *smu.1509*, ont respectivement été amplifiés par PCR à l'aide des oligonucléotides BF36-*Nde*I / BF37-*Xho*I et BF38-*Nde*I / BF39-*Xho*I. Les fragments obtenus ont ensuite été digérés par les enzymes de restriction *Nde*I et *Xho*I et clonés dans le plasmide pET28a préalablement digéré par ces mêmes enzymes de restriction. Le produit de ligation a alors servi à transformer la souche d'*E. coli* Rosetta résultant en la construction des souches TIL1387 (*E. coli* Rosetta pET28a::Gbs1555) et TIL1388 (*E. coli* Rosetta pET28a::Smu.1509).

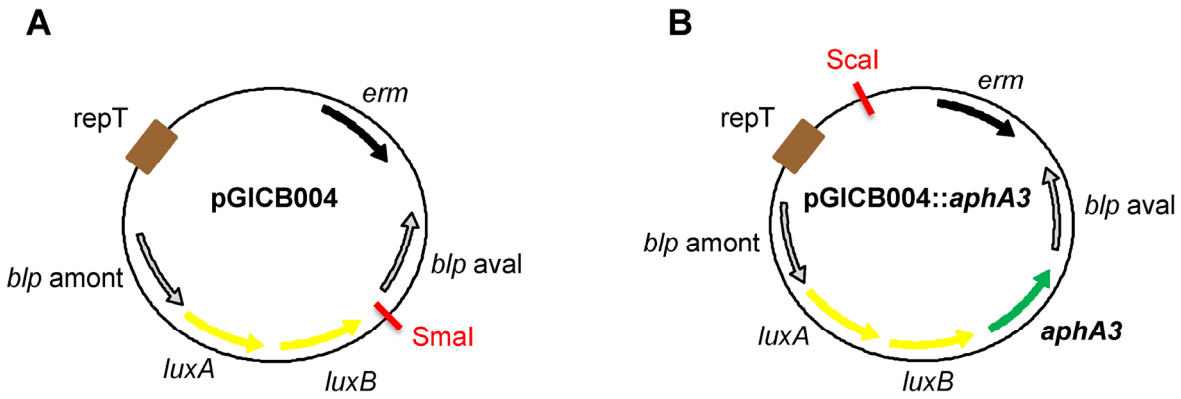


Figure 46 – Représentation schématique des plasmides pGICB004 (A) et pGICB004::*aphA3* (B).

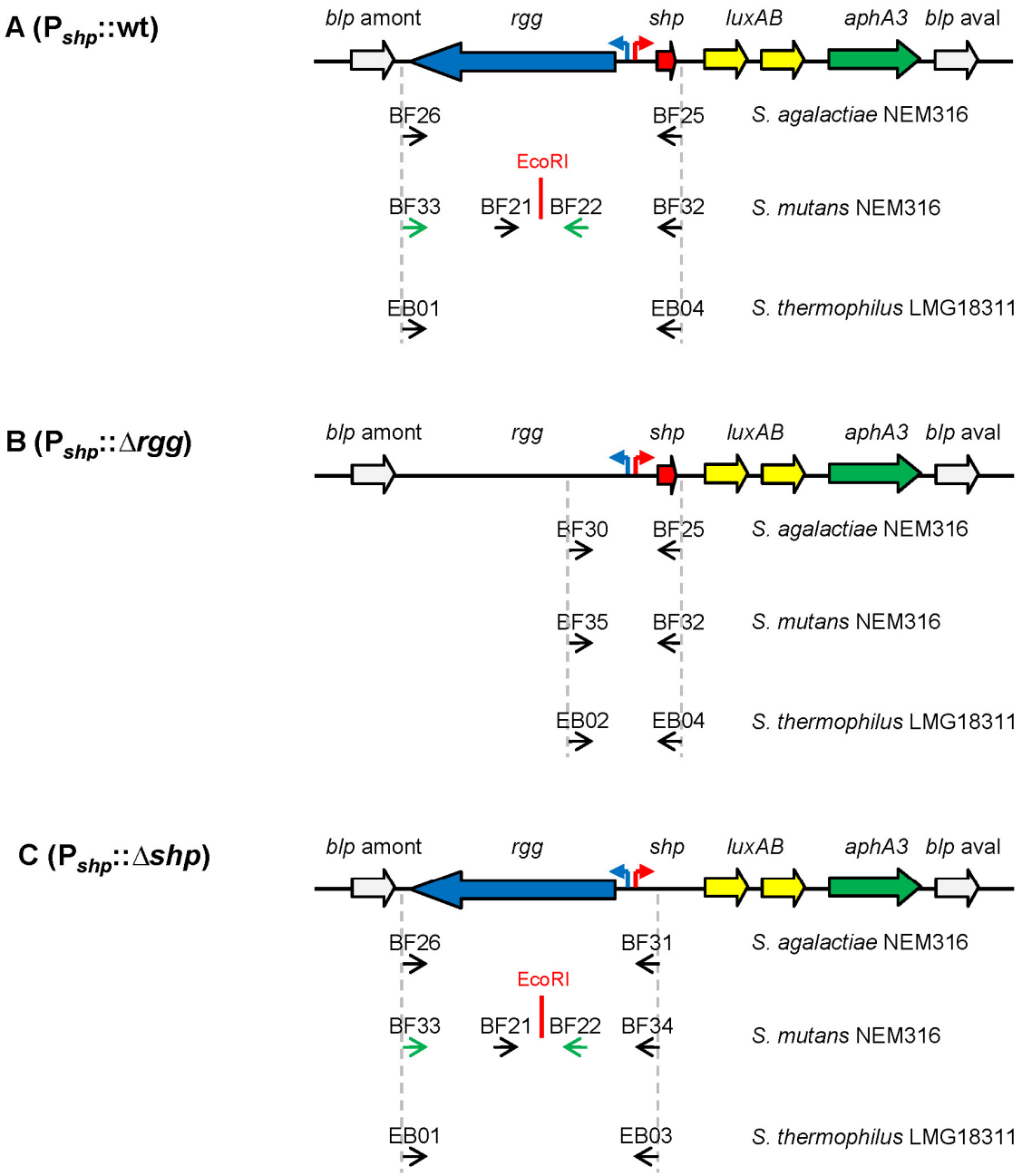


Figure 47 – Fusions transcriptionnelles construites pour l'étude de la fonctionnalité des systèmes SHP/Rgg.

La souche TIL1389 (*amiCDE::erm*) a été construite par PCR chevauchante à trois fragments. La cassette érythromycine a été amplifiée par PCR à partir du pG$^+$host9 en utilisant les oligonucléotides Erm-F / Erm-R et les fragments amont et aval du locus *amiCDEF* à partir des oligonucléotides BF1 / BF40 et BF41 / BF4. Une PCR trois fragments a ensuite été réalisée à l'aide des oligonucléotides BF1 / BF4. Le produit de PCR a alors servi à transformer des cellules de *S. thermophilus* LMD-9 naturellement compétentes.

Afin d'étudier la fonctionnalité des systèmes SHP/Rgg de *S. agalactiae* NEM316 (SHP/Gbs1555), *S. mutans* UA159 (SHP/Smu.1509) et *S. thermophilus* LMG18311 (SHP/Rgg.0182), trois fusions transcriptionnelles couplant le promoteur du gène *shp* à un système rapporteur *luxAB* et exprimant ou non de manière endogène les gènes *shp* et *rgg* du locus apparenté ont été construites. Ces fusions transcriptionnelles, appelées A (P_{shp}::wt), B (P_{shp}::Δ*rgg*), et C (P_{shp}::Δ*shp*), sont présentées dans la Figure 47. La construction des plasmides pGICB004::*aphA3*::*agalactiae* a été réalisée comme suit. Les différents fragments du promoteur *shp* ont été amplifiés par PCR à l'aide des oligonucléotides BF26-*Spe*I / BF25-*Eco*RI (A), BF30-*Spe*I / BF25-*Eco*RI (B) et BF26-*Spe*I / BF31-*Eco*RI (C), digérés par les enzymes de restriction *Eco*RI et *Spe*I et clonés dans le plasmide pGICB004::*aphA3* préalablement digéré par ces mêmes enzymes de restriction. L'intégration de ces plasmides dans la souche de *S. thermophilus* LMD-9 a conduit à l'obtention des souches TIL1345 (LMD-9 *blp*::P*shp::*wt_*agalactiae::aphA3)*, TIL1380 (LMD-9 *blp*::P*shp::*Δ*rgg_agalactiae*:: *aphA3*) et TIL1382 (LMD-9 *blp*::P*shp::*Δ*shp_agalactiae::aphA3*). La souche TIL1381 (*amiCDE::erm blp*::P*shp::*wt_*agalactiae::aphA3*) a, quant à elle, été obtenue par transformation de la souche TIL1345 avec de l'ADN chromosomique de la souche TIL1389. Les plasmides pGICB004::*aphA3::mutans* ont été construits selon la même procédure. La présence d'un site *Eco*RI dans le gène *rgg* a néanmoins nécessité une étape supplémentaire de clonage. Aussi, un premier fragment a tout d'abord été amplifié par PCR à l'aide des oligonucléotides BF33-*Spe*I / BF22, digéré par les enzymes de restriction *Eco*RI et *Spe*I puis cloné dans le plasmide pGICB004::*aphA3*. Les différents fragments du promoteur *shp* ont ensuite été amplifiés par PCR à l'aide des oligonucléotides BF21 / BF32-*Eco*RI (A), BF35 / BF32-*Eco*RI (B) et BF21 / BF34-*Eco*RI (C), digérés et clonés soit dans le pGICB004::*aphA3* (B), soit dans le plasmide pGICB004::*aphA3* contenant le fragment 1 préalablement intégré (A et C). L'intégration de ces plasmides dans la souche de *S. thermophilus* LMD-9 a conduit à l'obtention des souches TIL1383 (LMD-9 *blp*::P*shp::*wt_*mutans::aphA3*), TIL1386 (LMD-9 *blp*::P*shp::*Δ*rgg_mutans*::*aphA3*) et TIL1384 (LMD-9 *blp*::P*shp::*Δ*shp_mutans::aphA3*). La souche TIL1385 (*amiCDE::erm blp*::P*shp::*wt_*mutans::aphA3*) a ensuite été obtenue par transformation de la souche TIL1383 avec de l'ADN chromosomique de la souche TIL1389. La construction des plasmides pGICB004::*aphA3::thermophilus*LMG a également été réalisée selon la même procédure. La présence d'un site *Spe*I dans le gène *rgg*, nous a

néanmoins conduit à entreprendre l'intégration de ces fusions transcriptionnelles au niveau du site *Eco*RI. Aussi, les différents fragments du promoteur *shp* ont été amplifiés par PCR à l'aide des oligonucléotides EB01-*Eco*RI / EB04-*Eco*RI (A), EB02-*Eco*RI / EB04-*Eco*RI (B) et EB01-*Eco*RI / EB03-*Eco*RI (C), digérés par l'enzymes de restriction *Eco*RI et clonés dans le plasmide pGICB004::*aphA3* préalablement digéré par l'enzyme de restriction *Eco*RI. L'intégration de ces plasmides dans la souche *S. thermophilus* LMD-9 a conduit à l'obtention des souches TILA (LMD-9 *blp*::P*shp::*wt_*thermophilus*LMG*::aphA3)*, TILB (LMD-9 *blp*::P*shp::*Δ*rgg_thermophilus*LMG::*aphA3*) et TILC (LMD-9 *blp*::P*shp::*Δ*shp_thermophilus* LMG*::aphA3*). La souche TILD (*amiCDE::erm blp*::P*shp::*wt_*thermophilus* LMG*::aphA3*) a, quant à elle, été obtenue par transformation de la souche TILA avec de l'ADN chromosomique de la souche TIL1389.

L'ensemble des constructions a été vérifié par PCR et validé par séquençage.

- **Mesure de l'activité luciférase**

A partir d'une colonie isolée sur boite, une pré-culture sur la journée a été ensemencée dans 5 mL de milieu M17lac avec ou sans antibiotique et incubée à 42°C. Un repiquage à 2% a ensuite été effectué dans 5 mL de milieu MCD sans antibiotique et incubé sur la nuit à 42°C. Cette deuxième préculture a ensuite été utilisée pour ensemencer, à t=0, une culture de 50 mL de MCD à DO_{600nm} = 0,05 incubée à 42, 37 ou 30°C. A partir de t=1h, des aliquotes de 1 mL ont été prélevés à intervalle de temps régulier jusqu'en phase stationnaire de croissance et analysés selon le protocole décrit dans l'article 1. La bioluminescence émise est exprimée en unité relative de luminescence (RLU, **R**elative **L**ight **U**nit) et les résultats sont rapportés à la DO_{600nm} (RLU/DO_{600nm}).

Lorsque nécessaire, les peptides synthétiques (solubilisés dans du DMSO 100% dilué ensuite à 5% dans du tampon HBS 20 mM pH 7,4) ont été ajoutés aux cultures, à t=0, à une concentration finale de 1 µM. Dans ce cas, l'activité luciférase a été mesurée en système automatisé Synergy HT multimode reader (Bioteck, Winooski, VT) en microplaque PS blanche 96 puits à fond plat transparent (VWR). Une solution de nonylaldéhyde à 0,01% est alors préparée extemporanément dans de l'huile de paraffine. 50 µL de cette solution sont déposés au niveau des zones inter-puits et l'analyse est réalisée sur des cultures de 250 µL.

- **Analyse nanoLC-MS/MS**

Afin de déterminer les formes matures des phéromones SHP des systèmes de *S. mutans* et *S. agalactiae*, une première analyse en système hétérologue chez *S. thermophilus* LMD-9

a été entreprise. Dans un deuxième temps, la présence des peptides identifiés en système hétérologue a été recherchée dans les surnageants de culture des souches sauvages *S. mutans* UA159 et *S. agalactiae* NEM316.

Locus *shp/gbs1555* de *S. agalactiae* NEM316 : Les surnageants de cultures des souches TIL1345 (dérivé de *S. thermophilus* LMD-9 produisant la phéromone d'intérêt) et LMD-9 ont été récupérés par centrifugation et analysés selon le protocole décrit dans l'article 1. A partir de 5 µL de surnageant de culture, les peptides natifs maturés par hydrolyse à partir de la séquence du précurseur MKKINKALLFTLIMDILIIVGG ont été recherchés. Les peptides du surnageant ont été séparés par nanoHPLC et analysés en ligne par spectrométrie de masse sur LTQ-Orbitrap. L'analyse des données de spectrométrie de masse a été réalisée manuellement par extraction du signal de masse (XIC) correspondant aux séquences en acides aminés possédant la glycine C-terminale du précurseur et allant de l'hexapeptide LIIVGG au tridecapeptide FTLIMDILIIVGG. La validation manuelle de la présence de(s) peptide(s) a été réalisée en se référant à trois critères principaux ; les peptides détectés au sein du surnageant de culture de la souche TIL1345 devaient (i) posséder un temps de rétention chromatographique compatible avec l'indice GRAVY d'hydrophobicité du peptide, (ii) être absents du surnageant de culture de la souche LMD-9 et (iii) générer après fragmentation en CID (Colision Induce Dissociation) des ions fils compatibles avec la séquence.

Locus *shp/smu.1509* de *S. mutans* UA159 : Les surnageants de cultures des souches TIL1383 (dérivé de *S. thermophilus* LMD-9 produisant la phéromone d'intérêt), LMD-9 et UA159 (*S. mutans*) ont été récupérés par centrifugation et analysés selon le protocole décrit dans l'article 1, à partir de 5 µL de surnageant de cuture. Une analyse par spectrométrie de masse (identique à celle décrite précédemment) a été entreprise afin d'identifier la/les forme(s) mature(s) issue(s) du précurseur MRNKIFMTLIVVLETIIIIGGG, en recherchant au sein des surnageants TIL1383 et LMD9 la présence de peptides allant de l'hexapeptide IIIGGG au tridecapeptide IVVLETIIIIGGG. Dans ces conditions d'analyse aucun signal de masse compatible avec les séquences testées n'a été mise en évidence. Dans un deuxième temps, une seconde approche d'identification ciblée et plus sensible a été mise en place pour valider par spectrométrie de masse au sein des différents surnageants la présence du peptide ETIIIIGGG en recherchant la transition m/z>b7 (739 > 741) en condition d'acquisition de type SRM (Single Reaction Monitoring) sur LTQ-Orbitrap. La transition m/z > ions fils b7 a été choisie sur la base du comportement à la fragmentation sur LTQ-Orbitrap CID de la phéromone SHP1358 qui génére un ion fils N-terminal dominant correspondant à un clivage au niveau de la liaison peptidique résultant de la perte de 2 acides aminés coté C-terminal (article 1).

- **Production et purification des protéines Rgg**

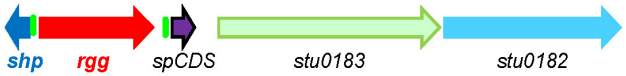
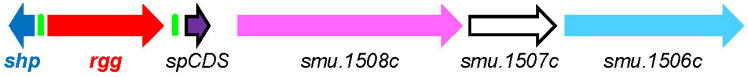
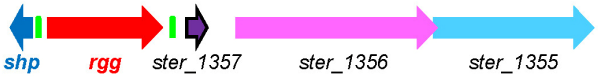
La production et la purification des protéines Rgg1358, Gbs1555, Smu.1509 et Rgg.0182, ont été réalisées selon le protocole décrit dans l'article 1, à partir de cultures des souches TIL1318, TIL1387, TIL1388 et TIL1035 respectivement.

3. Résultats

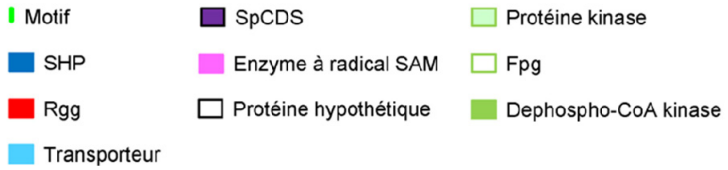
- **Sélection des loci *shp/rgg***

Afin d'élargir, à d'autres loci, le mécanisme de QS décrit sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9 et d'étudier la spécificité des systèmes SHP/Rgg, des loci modèles ont tout d'abord été choisis. Bien que la construction d'un arbre phylogénétique des protéines Rgg ait fait ressortir trois groupes de systèmes SHP/Rgg (article 1), nous nous sommes focalisés dans cette étude sur les systèmes appartenant aux groupes I et II, présentant de plus fortes similarités en termes de séquences protéiques et d'organisation génétique. Afin d'évaluer la spécificité des systèmes SHP/Rgg et de définir si cette dernière est souche-spécifique, espèce-spécifique ou groupe-spécifique ; deux systèmes SHP/Rgg ont été sélectionnés au sein de chaque groupe. L'analyse des différents loci appartenant au groupe I (dont l'environnement génétique est généralement peu conservé) a souligné la présence d'un système *shp/rgg* présentant un environnement génétique identique entre trois streptocoques pathogènes : *S. agalactiae*, *S. dysgalactiae* et *S. pyogenes* (cf. résultats chapitre 3). En effet, pour ces trois loci, l'environnement génétique est composé d'une glycosylase (Fpg), d'une CoaE-kinase et d'un transporteur. Aussi, afin d'étudier ce locus, nous avons choisi de travailler sur le locus *shp/gbs1555* de *S. agalactiae* NEM316, seul locus *shp/rgg* détecté au sein de l'espèce. Dans un second temps, le système *shp/rgg.0182* de *S. thermophilus* LMG18311 présentant une particularité au niveau de la séquence peptidique de la phéromone SHP a été sélectionné (parmi les 5 loci présents au sein de cette souche). En effet, alors que la majorité des peptides SHP appartenant au groupe I, présente un résidu aspartate conservé dans leur séquence, ce peptide SHP possède un résidu cystéine (article 1, cf. résultats chapitre 3). Au sein du groupe II, l'environnement génétique des loci *shp/rgg* étant conservé et composé principalement d'une spCDS, d'une enzyme à radical SAM et d'un transporteur (cf. résultats chapitre 3), le choix s'est alors porté sur la proximité des bactéries entre elles et leur possibilité de se retrouver dans un même écosystème. Aussi, le locus *shp/smu.1509* de *S. mutans* UA159 a été sélectionné en regard du locus *shp/rgg1358* de *S. thermophilus* LMD-9 déjà étudié (article 1) et sélectionné comme deuxième locus au sein de ce groupe de systèmes SHP/Rgg. En effet, ces deux

Tableau 19 – Loci *shp/rgg* sélectionnés dans cette étude.

		Locus	Séquence du peptide SHP[a]	Environnement génétique
Groupe I	*S. agalactiae* NEM316	*shp/gbs1555*	MKKINKALLFTLIM**D**ILIIVGG	
	S. thermophilus LMG18311	*shp/rgg.0182*	MKLLKIIVLLT**C**IYIIVGGV	
Groupe II	*S. mutans* UA159	*shp/smu.1509*	MRNKIFMTLIVVL**E**TIIIIGGG	
	S. thermophilus LMD-9	*shp/rgg.1358*	MKKQILLTLLLVVF**E**GIIVIVVG	

a. Les formes matures des phéromones déterminées par LC-MS/MS sont indiquées en couleur sur les séquences des précurseurs.



streptocoques font partie de la flore buccale de l'Homme et des animaux, le premier étant résidant, le second ne faisant que transiter au sein de celle-ci. Le locus *shp/rgg* de *S. mutans* est le seul locus *shp/rgg* détecté au sein de l'espèce. *A contrario*, *S. thermophilus* LMD-9 possède également 3 autres loci dans son génome. Le Tableau 19 récapitule les différents loci sélectionnés.

- **Etude de la fonctionnalité des systèmes SHP/Rgg : stratégie**

L'étude du locus *shp/rgg1358* de *S. thermophilus* LMD-9 a révélé que le gène *shp* code une phéromone dont l'expression auto-induite nécessite la présence du régulateur transcriptionnel Rgg1358 et d'un transporteur d'oligopeptides AmiCDEF (article 1). Nous émettons l'hypothèse que dans tout système SHP/Rgg, l'expression du gène *shp* est contrôlée de la même manière. Aussi, afin d'étudier la fonctionnalité de systèmes SHP/Rgg de streptocoques pathogènes (*S. agalactiae* NEM316 et *S. mutans* UA159) et GRAS (*S. thermophilus* LMG18311), différentes fusions transcriptionnelles P_{shp}::*luxAB* ont été construites (Figure 47). Ces fusions couplent le promoteur du gène *shp* de chaque locus à un système rapporteur *luxAB* et, expriment ou non de manière endogène les gènes *shp* et *rgg* du locus apparenté. Ces fusions, introduites dans le chromosome de *S. thermophilus* LMD-9 au locus *blp*, ont été dénommées : P_{shp}*::wt_strepto*, lorsque les gènes *shp* et *rgg* sont exprimés de manière endogène et, P_{shp}*::*Δ*rgg_strepto* et P_{shp}*::*Δ*shp_strepto*, lorsque seulement le gène *shp* ou *rgg*, respectivement, est exprimé de manière endogène (Figure 47). Le terme « *strepto* » étant remplacé par « *mutans* », « *agalactiae* » ou « *thermophilus*LMG » en fonction du locus étudié. Le rôle du système de transport AmiCDEF ayant été démontré dans la réimportation de la phéromone SHP1358 du locus *shp/rgg1358* de *S. thermophilus* LMD-9 (article 1), la fusion transcriptionnelle P_{shp}*::wt_strepto* a également été étudiée dans un contexte génétique Δ*ami*. Cette fusion a été nommée P_{shp}*::*Δ*ami_strepto*.

Aussi, si la fonctionnalité des systèmes *shp/rgg* de *S. mutans* UA159, *S. agalactiae* NEM16 et *S. thermophilus* LMG18311 est identique à celle du système *shp/rgg1358* de *S. thermophilus* LMD-9, l'activité des promoteurs P_{shp} doit être dépendante des trois partenaires, c'est-à-dire de la phéromone SHP, du régulateur transcriptionnel Rgg et du système de transport d'oligopeptides AmiCDEF.

- **Influence de la température sur l'activité des systèmes SHP/Rgg**

Récemment, les travaux de Henry *et al.* ont montré l'influence de la température sur l'expression du gène *shp0182* du locus *shp/rgg0182* de *S. thermophilus* LMG18311 (Henry

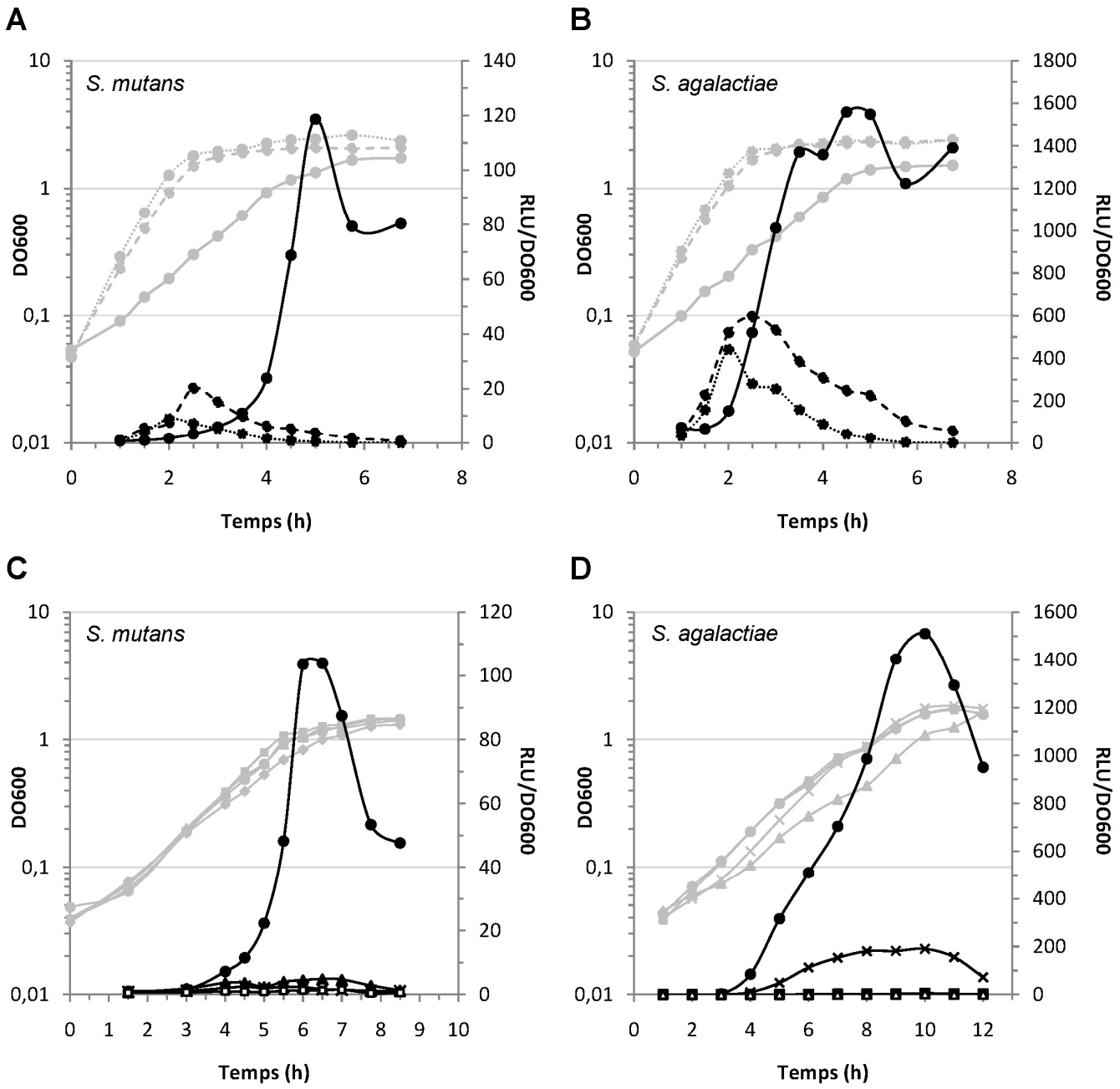


Figure 48 – Croissance et activité luciférase de souches présentant une fusion P*shp::luxAB* dans leur génome et dans différentes conditions de cultures.

Les courbes de croissance sont de couleur grise et les activités luciférases en noires. Les données présentées sont représentatives de trois expériences indépendantes.

A et **B**, Influence de la température sur la fusion transcriptionnelle P_{shp}*::wt_strepto*. Les températures de croissances sont indiquées par le style de trait utilisé : 30°C (—), 37°C (- - -) et 42°C (⋯).

C et **D**, Etude de l'activité des P_{shp} dans différents contextes génétiques à 30°C. Les fusions transcriptionnelles utilisées sont indiquées de la manière suivante : P_{shp}*::wt_strepto* (●), P_{shp}*::Δrgg_strepto* (■), P_{shp}*::Δshp_strepto* (▲) et P_{shp}*::Δami_strepto* (✕).

et al., 2011). L'activité de la fusion $P_{shp0182}$*::luxAB* introduite au locus *blp* de *S. thermophilus* LMG18311 s'est révélée 6 fois plus importante à 30°C versus 42°C en MCD. 42°C représentant la température optimale de croissance de *S. thermophilus*, le système SHP/Rgg0182 a alors été proposé comme intervenant dans la réponse au stress thermique chez *S. thermophilus* LMG18311 (Henry *et al.*, 2011). Aussi, afin de définir si la température est une condition modulant l'activité d'autres systèmes SHP/Rgg, l'expression des fusions transcriptionnelles P_{shp}*::wt_strepto* des loci de *S. agalactiae* et *S. mutans* a tout d'abord été testée à ces deux températures. Par ailleurs, ce travail étant réalisé en système hétérologue, il nous a semblé opportun d'étudier également l'expression de ces fusions transcriptionnelles à la température optimale de croissance de ces deux bactéries, c'est-à-dire 37°C.

Quelque soit la fusion testée – P_{shp}*::wt_mutans* ou P_{shp}*::wt_agalactiae* –, un profil similaire d'activité des promoteurs P_{shp} a été observé (Figure 48A, B). En effet, bien qu'une activité luciférase ait été détectée pour chaque température testée, le niveau d'expression de ces dernières est d'autant plus élevé que la température est faible. Ainsi, à 30°C, une activité luciférase équivalente à 120 RLU/DO_{600nm} ou à 1600 RLU/DO_{600nm} a été obtenue avec les fusions P_{shp}*::wt_strepto* de *S. mutans* et *S. agalactiae*, respectivement. Ces activités luciférases sont respectivement 6 ou 2,6 fois plus importantes en comparaison avec les profils d'activité des promoteurs P_{shp} à 37°C. Ceci suggère un effet température sur l'expression de ces deux systèmes SHP/Rgg et éventuellement leur implication dans l'adaptation de la bactérie à un stress froid. En effet, à 30°C, le taux de croissance de *S. thermophilus* LMD-9 est deux fois plus faible ($G_{30°C}$ = 69 min / $G_{37/42°C}$ = 33 min) et la biomasse finale obtenue moins importante. Enfin, quelque soit la température testée, les niveaux d'activité luciférase des fusions P_{shp}*::wt_strepto* atteignent leur optimum toujours en fin de phase exponentielle / début de phase stationnaire de croissance (Figure 48A, B).

Sur la base de ces résultats, l'étude de la fonctionnalité des systèmes SHP/Rgg de *S. agalactiae* NEM316, *S. mutans* UA159 et *S. thermophilus* LMG18311 a été entreprise à 30°C.

- **L'activité des promoteurs P_{shp} est contrôlée par SHP, Rgg et AmiCDEF chez *S. agalactiae* et *S. mutans***

Afin de définir si les produits des gènes *shp* et *rgg* sont nécessaires à l'activation des promoteurs P*shp*, la comparaison de la fusion transcriptionnelle P_{shp}*::wt_strepto* avec les fusions transcriptionnelles P_{shp}*::*Δ*rgg_strepto* et P_{shp}*::*Δ*shp_strepto* a tout d'abord été entreprise. Une activité luciférase, similaire à celles obtenues lors des essais température

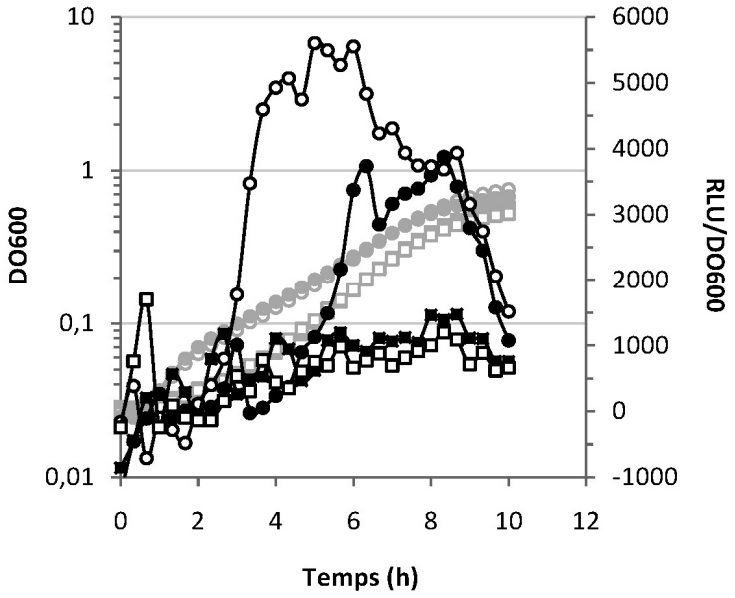


Figure 49 – Croissance et activité luciférase de différentes souches de *Streptococcus thermophilus* LMG18311 présentant une fusion $P_{shp0182}$*::luxAB* dans leur génome.
Les courbes de croissance sont de couleur grise et les activités luciférases en noires. L'utilisation des symboles rond et carré fait respectivement référence aux souches TIL1322 (LMG18311 *blp::Pshp0182*) et TIL1324 (Δ*rgg0182 blp::Pshp0182*). Lorsque du peptide synthétique CIYIIVGGV a été ajouté aux cultures, l'intérieur des symboles est de couleur blanche. Les données présentées sont représentatives de trois expériences indépendantes.

(Figure 48A, B), a bien été mesurée avec les fusions P_{shp}*::wt_strepto* de *S. mutans* et *S. agalactiae* (Figure 48C, D). En revanche, aucune activité luciférase (ou une activité luciférase considérée comme non significative) n'a été détectée avec les fusions transcriptionnelles n'exprimant pas de manière endogène le gène *shp* ou *rgg*. Aussi, les partenaires SHP et Rgg de chaque locus apparenté (*shp/smu.1509* ou *shp/gbs1555*) sont essentiels à la fonctionnalité de ces deux systèmes SHP/Rgg ; les protéines régulatrices Rgg contrôlant l'expression de leur gène *shp* apparenté et le produit du gène *shp* impliqué dans sa propre transcription. Dans un second temps, et afin d'étudier le rôle du transporteur d'oligopeptides sur la transcription des gènes *shp*, la fusion transcriptionnelle P_{shp}*::wt_strepto* a été comparée avec la fusion P_{shp}*::*Δ*ami_strepto*. Dans le cadre du locus de *S. mutans*, aucune activité du promoteur P_{shp} n'a été détectée en présence d'un contexte génétique délété du système de transport AmiCDEF (Figure 48C). Aussi, ce dernier est également impliqué dans le contrôle positif de la transcription du gène *shp* appartenant au locus *shp/smu.1509* de *S. mutans*. Dans le cadre du locus de *S. agalactiae*, un niveau d'activité luciférase de 200 RLU/DO_{600nm} a été détecté en présence de la fusion P_{shp}*::*Δ*ami_agalactiae* (Figure 48D). Ce niveau d'activité est néanmoins 7,3 fois plus faible que celui obtenu dans un contexte sauvage. Aussi, la différence d'activité significative du promoteur P_{shp} dans un contexte génétique délété du système de transport Ami en comparaison au contexte sauvage, nous permet de conclure quant à l'implication de ce transporteur dans la boucle de rétrocontrôle positif modulant la transcription du gène *shp* du locus de *S. agalactiae*. Néanmoins, il est possible que la transcription du gène *shp* de ce locus implique un autre élément de régulation.

Enfin, aucune activité luciférase n'a été détectée avec les fusions transcriptionnelles du locus *shp/rgg.0182* de *S. thermophilus* LMG18311 (données non présentées).

- **L'activation du promoteur $P_{shp0182}$ est fonctionnelle en système homologue**

L'absence d'activité luciférase du promoteur $P_{shp0182}$ du locus *shp/rgg.0182* de *S. thermophilus* LMG18311 à 30°C et en système hétérologue, nous a conduit, dans un second temps, à tester la fonctionnalité de ce système en système homologue. Aussi, sur la base des séquences des phéromones SHP connues, le nonapeptide CIYIIVGGV, correspondant à la partie C-terminale de la séquence précurseur (Tableau 19), a été synthétisé dans le but de complémenter des cultures des souches TIL1322 (LMG18311 *blp::*$P_{shp0182}$) et TIL1324 (Δ*rgg0182 blp::*$P_{shp0182}$) (Figure 49). L'ajout de la phéromone synthétique dans le surnageant de la souche sauvage augmente d'un facteur 1,5 l'activité luciférase de la fusion avec le promoteur $P_{shp0182}$. Par ailleurs, l'ajout du peptide synthétique

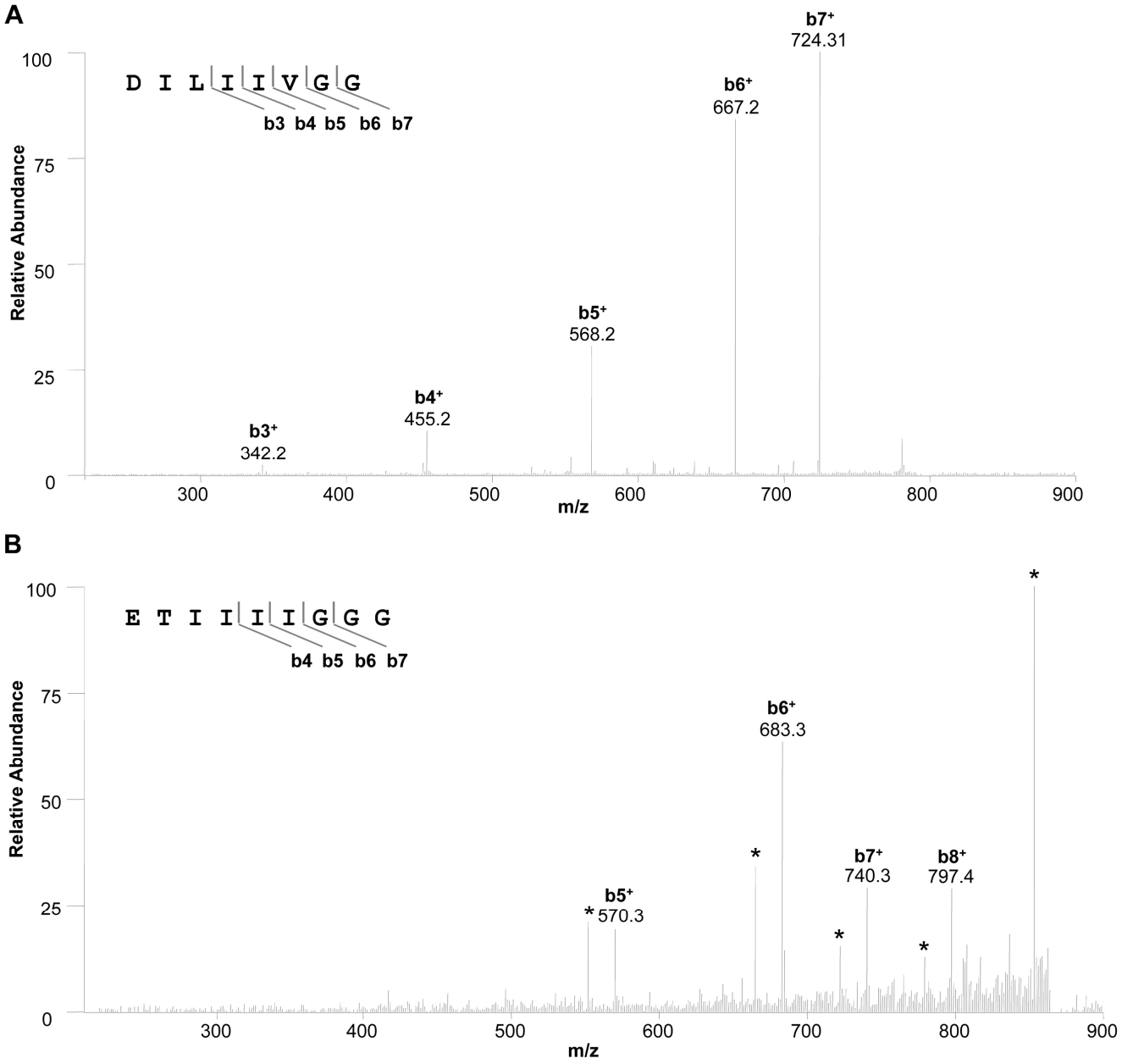


Figure 50 – Spectres de fragmentation des ions m/z (799,5 et 872,5) permettant la validation des séquences des phéromones SHP issues des loci *shp/gbs1555* de *Streptococcus agalactiae* NEM316 (A) et *shp/smu.1509* de *Streptococcus mutans* UA159 (B).
Le symbole « * » correspond à la perte d'une molécule d'H_2O.

dès le début de la culture impacte le déclenchement de l'activation du promoteur qui se produit 2h plus tôt. En revanche, dans un contexte génétique délété du gène *rgg0182,* une activité similaire du promoteur $P_{shp0182}$ est observée quelque soit la condition testée (en absence ou en présence de la phéromone synthétique) (Figure 49). Aussi, ces résultats soulignent l'implication du régulateur transcriptionnel Rgg.0182 sur la transcription du gène *shp0182* (Henry *et al.*, 2011) et démontrent que le nonapeptide contenant une cystéine est une forme mature capable d'activer l'expression du système *shp/rgg.0182* de *S. thermophilus* LMG18311.

- **Les formes matures des phéromones SHP sont générées à partir du résidu conservé au sein de leur séquence**

Identification de la phéromone du système SHP/Gbs1555 de *S. agalactiae* NEM316 : La comparaison des surnageants de cultures issus des souches TIL1345 et LMD-9 a permis de mettre en évidence un peptide (non présent dans le surnageant de la souche LMD-9) de masse m/z = 799,4329 et présentant un profil d'hydrophobicité compatible avec la séquence peptidique putative de la phéromone. Cette masse, une fois fragmentée et analysée sur LTQ-Orbitrap, a permis d'identifier un octapeptide de séquence DILIIVGG, correspondant aux 8 acides aminés issus d'un clivage côté C-terminal de la séquence précurseur de la phéromone SHP. La masse de ce peptide a ensuite été recherchée dans le surnageant de culture de la souche sauvage *S. agalactiae* NEM316. En parallèle, et afin de s'affranchir de la présence d'éventuels produits de dégradation pouvant générer un fragment compatible avec les masses identifiées, la séquence de l'octapeptide a été recherchée dans la base de données « non-redundant protein sequences » de GenBank. Un fragment heptapeptidique de séquence ILIIVGG localisé dans la partie N-terminale d'une protéine membranaire putative, codée par le gène *gbs0830*, a été identifié. Bien que similaire, ce fragment étant plus court que celui correspondant à la masse observée, ne pouvait influencer la détection de l'octapeptide identifié. Aussi, la détection de la masse de l'octapeptide dans le surnageant de culture de la souche sauvage et sa validation par fragmentation a permis de confirmer la forme mature de la phéromone du locus *shp/gbs1555* de *S. agalactiae* NEM316 (Figure 50A ; Tableau 19).

Identification de la phéromone du système SHP/Smu.1509 de *S. mutans* UA159 : L'analyse manuelle des signaux de masse de type SRM a permis de mettre en évidence 6 zones chromatographiques pouvant correspondre à la transition recherchée (et correspondant à un peptide de séquence ETIIIIGGG) en condition de production hétérologue chez *S. thermophilus* LMD-9. Une analyse manuelle des spectres de fragmentation moyens

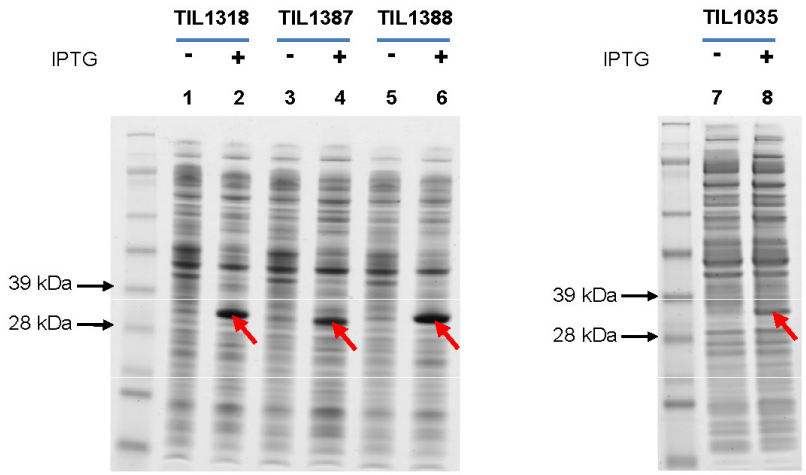


Figure 51 – SDS-PAGE des fractions protéiques issues des cultures d'*Escherichia coli* permettant la surproduction de protéines Rgg recombinantes.
Les souches TIL1318, TIL1387, TIL1388 et TIL1035 permettent respectivement la surproduction des protéines Rgg1358 (34,1 KDa), Gbs1555 (33,9 KDa), Smu.1509 (33,1 KDa) et Rgg.0182 (34,1 KDa) indiquées par des flèches rouges.

associés a permis de vérifier que ce peptide est bien produit et relargué dans le surnageant en condition de production hétérologue mais également chez la souche sauvage *S. mutans* UA159 (Figure 50B ; Tableau 19). Aucun fragment compatible avec une séquence protéique chez *S. mutans* n'a, par ailleurs, été détecté suite à la réalisation d'un Blastp dans la base de données « non-redundant protein sequences » de GenBank.

- **Surproduction des protéines Rgg**

La surproduction des protéines Rgg (Rgg1358, Gbs1555, Smu.1509, Rgg.0182) et leur purification sont nécessaires à l'étude *in vitro* des affinités des couples SHP/Rgg par SPR. Aussi, après construction des plasmides, la surproduction des protéines Rgg recombinantes chez *E. coli* a été vérifiée sur gel SDS-PAGE (Figure 51). L'induction à l'IPTG, conduisant à la surexpression des quatre protéines Rgg, est visible dans les pistes 2, 4, 6 et 8. Aussi, les différentes constructions étant validées, les derniers peptides synthétiques correspondant aux formes matures des différentes phéromones SHP identifiées dans cette étude reçus et le protocole de purification des protéines mis au point (article 1), les expériences de SPR ont pu être planifiées pour fin octobre.

4. Discussion

La fonctionnalité de systèmes SHP/Rgg appartenant aux groupes I et II a été évaluée au sein de plusieurs espèces de streptocoques. L'étude des loci *shp/gbs1555* de *S. agalactiae* NEM316 et *shp/smu.1509* de *S. mutans* UA159, en système hétérologue chez *S. thermophilus* LMD-9, a tout d'abord permis d'élargir la fonctionnalité du mécanisme SHP/Rgg (décrit initialement sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9) à deux streptocoques pathogènes. En effet, au même titre que le locus *shp/rgg1358* de *S. thermophilus* LMD-9, l'activité des promoteurs P_{shp} de ces deux systèmes a été décrite comme contrôlée par SHP, Rgg et AmiCDEF. Ces résultats, soulignant l'auto-induction de la transcription des gènes *shp* de ces deux systèmes, confortent le phénomène d'auto-induction des phéromones largement décrit dans des mécanismes de QS (article1, Chang *et al.*, 2011; Fontaine *et al.*, 2010a; Mashburn-Warren *et al.*, 2010; Novick, 2003; Pestova *et al.*, 1996a). L'identification des formes matures des phéromones SHP de ces deux systèmes a également été entreprise. Aussi, un nonapeptide de séquence ETIIIIGGG et un octapeptide de séquence DILIIVGG ont été identifiés comme correspondant aux formes matures des systèmes de *S. mutans* et *S. agalactiae*, respectivement. Dans tous les cas, et en particulier pour *S. mutans* où une recherche ciblée d'une séquence prédite comme fonctionnelle a été réalisée, on ne peut exclure que d'autres formes matures actives existent. En outre, de

récentes études soulignant le rôle essentiel des résidus conservés – aspartate ou glutamate – identifiés au sein des séquences des phéromones SHP, corroborent le fait que ces résidus soient inclus dans les séquences matures des phéromones SHP (article 1) car nécessaire à leur activité (Chang *et al.*, 2011). Il est alors envisageable que les formes matures des phéromones SHP soient, dans leur ensemble, générées suite à un clivage côté C-terminal de la forme précurseur effectué à partir du résidu conservé au sein de leur séquence. L'identification de ces formes matures ayant été réalisée à partir de surnageants de culture par LC-MS/MS, ceci indique que les phéromones SHP des systèmes SHP/Rgg de *S. mutans* et *S. agalactiae*, au même titre que la phéromone SHP1358 de *S. thermophilus* LMD-9, sont libérées – au moins en partie – dans le milieu extracellulaire au cours de la croissance. Récemment, l'étude de deux loci *shp/rgg* chez *S. pyogenes* a également démontré ce phénomène (Chang *et al.*, 2011).

Le rôle de la peptidase membranaise Eep a été démontré dans la maturation de diverses phéromones, telles que les phéromones SHP (article 1, Chang *et al.*, 2011) ou encore les anti- et sex-phéromones chez *E. faecalis* (Thoendel et Horswill, 2010). Néanmoins, les séquences peptidiques de ces différentes molécules de signalisation ne permettent pas la reconnaissance d'un ou de plusieurs site(s) protéolytique(s) de clivage. Aussi, il est possible que la spécificité des protéines Eep ait évoluée au sein des différents genres bactériens ou que les protéines Eep présentent une large spécificité de substrat.

Dans cette étude, nous avons montré que la température est un paramètre influençant l'expression des systèmes SHP/Rgg de *S. agalactiae* et *S. mutans*. Aussi, un niveau maximal d'activité luciférase ayant été obtenu lors de cultures réalisées à 30°C, l'application de conditions non optimales de croissance semble favoriser l'expression de ces systèmes. Néanmoins, nous ne pouvons affirmer pour le moment que ces systèmes répondent de manière spécifique à un stress froid. En effet, les conditions de culture utilisées dans cette étude ne correspondent pas à l'application d'un stress à proprement parler. Aussi, l'application d'un saut de température passant de 42 à 30°C permettrait de définir précisément le lien entre ces systèmes SHP/Rgg et la réponse au stress thermique. Néanmoins, quelque soit la température utilisée lors des expériences, un niveau maximal d'activité a été observé au moment de la transition phase exponentielle / phase stationnaire de croissance. Ce changement physiologique correspondant à un état de stress pour les bactéries (carence en nutriments, pH acide, etc.), l'action combinée de différents facteurs de « stress » pourrait alors expliquer l'augmentation brutale de l'activité des promoteurs P_{shp}*::wt_mutans* ou P_{shp}*::wt_agalactiae* dans ces conditions drastiques de croissance pour *S. thermophilus* LMD-9. Aussi, ces deux systèmes SHP/Rgg pourraient jouer un rôle dans la

Tableau 20 – Pourcentage d'identité des protéines affines AmiA chez *Streptococcus thermophilus* LMD-9 et CNRZ1066.

Séquence 1	Séquence 2	Identité
LMD-9_AmiA1	LMG18311_AmiA2	99 %
	LMG18311_AmiA3	86 %
LMD-9_AmiA3	LMG18311_AmiA2	86 %
	LMG18311_AmiA3	98 %

Les alignements ont été réalisés sous ClustalW2, paramètres standards.

réponse au stress des bactéries et répondre de manière additionnelle et relative à la quantité de stress générés sur les cellules. Enfin, l'effet visible de la température sur l'activité des promoteurs P_{shp} de *S. mutans* et *S. agalactiae*, nous a conduit à tester l'activité du $P_{shp1358}$ du locus *shp/rgg1358* de *S. thermophilus* LMD-9 à 30 et 37°C, celui-ci ayant auparavant été étudié à 42°C seulement (article 1). Dans ces conditions, un niveau d'activité luciférase inférieur à celui obtenu à 42°C a été observé (R. Gardan, communication personnelle). Aussi, l'impact d'un changement de température semble être propre à un nombre spécifique de systèmes SHP/Rgg et par conséquent, cette variation ne conditionne pas nécessairement une forte expression de la totalité de ces systèmes.

De manière inattendue, l'étude du locus *shp/rgg.0182* de *S. thermophilus* LMG18311 (en système hétérologue chez *S. thermophilus* LMD-9) n'a révélé aucune activité du promoteur $P_{shp0182}$ dans les conditions testées. Les souches LMD-9 et LMG18311, bien que possédant toutes deux la protéine affine AmiA3, se différencient au niveau de leur seconde protéine affine présente dans leur génome ; AmiA2 pour *S. thermophilus* LMG18311 et AmiA1 pour *S. thermophilus* LMD-9. Aussi, il est envisageable que l'import de la phéromone SHP du système *shp/rgg.0182*, présentant la particularité de posséder un résidu cystéine dans sa séquence, soit spécifique de la protéine affine AmiA2. L'alignement de la séquence protéique de la protéine affine AmiA2 avec celles de *S. thermophilus* LMD-9 souligne de forts pourcentages d'identité pouvant aller jusqu'à 99% (Tableau 20). Aussi, les résidus les différenciant seraient hautement spécifiques des peptides réimportés. De manière intéressante, la souche *S. thermophilus* CNRZ1066, présentant dans son génome un système *shp/rgg* paralogue au système *shp/rgg.0182* de *S. thermophilus* LMG18311, possède également la protéine affine AmiA2. Par ailleurs, dans la littérature, aucun lien n'est encore clairement établi entre la présence de certains résidus dans une séquence protéique de protéine affine et leur spécificité vis-à-vis de peptides donnés. Aussi, deux protéines affines présentant un fort pourcentage d'identité, peuvent ne pas partager exactement la même spécificité de substrat (Picon *et al.*, 2005). Des études de transport ou d'activité du promoteur $P_{shp0182}$ dans la souche sauvage et dans un mutant nul des protéines affines complémenté par un plasmide permettant la surproduction de chacune d'entre elles (AmiA1, AmiA2 et AmiA3) permettraient d'évaluer cette hypothèse. Le rôle éventuel de certains résidus dans la spécificité de liaison du substrat pourrait ensuite être étudié par mutagénèse dirigée. Par ailleurs, l'hypothèse d'une éventuelle différence de spécificité de la protéase membranaire Eep, dans ces différentes souches de streptocoques contenant le locus *shp/rgg.0182* (LMG18311 et CNRZ1066), a également été soulevée. Néanmoins, l'absence d'activité du promoteur $P_{shp0182}$ en système hétérologue, après ajout de peptide synthétique (CIYIIVGGV) dans les surnageants de cultures (données non présentées), ne permet pas de

valider cette hypothèse. Enfin, il est envisageable que différents systèmes SHP/Rgg ne soient pas compatibles au sein d'une même souche et que, par conséquent, un phénomène d'inhibition du système SHP/Rgg.0182 de *S. thermophilus* LMG18311 par un autre système SHP/Rgg de *S. thermophilus* LMD-9 se produise dans la cellule.

Tableau 21 – Récapitulatif des trois groupes de systèmes SHP/Rgg décrit dans l'article 1.

	Groupe I	Groupe II	Groupe III
Nombre de loci *shp/rgg* (sans redondance*)	42 (13)	19 (9)	7 (6)
Organisation génétique			
Séquence peptidique des SHP (sans redondance*)	MKKINKALLFTLIMDILIIVGG MKKINKALLLTLIMDILIIVGG MKKVNKALLFTLIMDILIIVGG MEKVSKILPILILVMDIIIIVGG MKKISKFLPILVLVMDIIII MKKISKFLPILILAMDIIIIVGGEVT MKKISKFLPILILAMDIIIIVGG MKKISKFLPILFLVMDIIIIVGG MKKISKFFPILMLVMDIIIIVGG MRKSYKLLKLILDIIIIIGLCQ** MKKYYQIFLLLFDIIIIIGLYQ** MKLLKIIVLLTCIYTIVGGV MKLLKIIVLLTCIYIIVGGV	MKKHGLLTLLLIILESIIVIGIG MNKRHFLMLLLMVLEEIIVIVVGYL MKKQKLLPILFLLFEGIIIIVVG MNKRHFLMLLLMVLEEIIVIGVGYL MKKQKLLLLVVLVCEGIIVILVG MNKESFLAILLLIFESIIVIAVG MKKQILLTLLLVVFEGIIVIVVG MRNKIFMTLIVVLETIIIIGGG MKQNYLIANITIVLILLISILKDIPPIIVIK	MKKVIAIFLFIQTVVVIDIIIFPPFG MKKQVLTLLTIVADIIIFFPFLTNR MKKQVLTLLTIVAEIIIFFPFLTNR MKKQILTLLKIVAEIIIILPFLTNL MKKQILTLLKIVAEIIIILPFLTNR MKKRKIQLILLLISEWVIVIPFLTNL

* Le terme « sans redondance » se réfère à la séquence peptidique des phéromones SHP.
** Séquences des phéromones SHP appartenant aux clusters 11 et 12.

.

Chapitre 3 : Prédiction des régulons SHP/Rgg

Fleuchot B.[1,2], Nicolas P.[3], Chambellon E.[1,2], Monnet V.[1,2], Gardan R.[1,2]

[1]INRA, UMR1319 MICALIS, F-78352 Jouy en Josas, France.
[2]AgroParistech, UMR MICALIS, F-78352 Jouy en Josas, France.
[3]INRA, UR1077 Mathématique, Informatique et Génome, F-78352 Jouy en Josas, France.

1. Contexte

Soixante huit systèmes SHP/Rgg ont été identifiés dans les génomes de streptocoques et classés en trois groupes selon deux principaux critères : la coévolution des protéines Rgg et des phéromones SHP ainsi que l'organisation génétique des gènes *shp* et *rgg* (Tableau 21) (article 1). Alors que le mécanisme de régulation de ces systèmes a été étudié sur différents loci *shp/rgg* (article 1 ; (Chang *et al.*, 2011), peu d'informations concernant les fonctions physiologiques qu'ils contrôlent sont encore disponibles dans la littérature. En effet, seulement les systèmes SHP/Rgg chez *S. pyogenes* ont été décrits, très récemment, comme impliqués dans la formation de biofilms (Chang *et al.*, 2011). Par conséquent, la prédiction des régulons des systèmes SHP/Rgg est un enjeu important que l'on a voulu aborder à travers cette étude. Pour ceci, une analyse *in silico* a été menée afin de définir un motif de fixation des protéines Rgg à l'ADN et de rechercher sa présence dans le génome des streptocoques séquencés. En parallèle, une analyse des environnements génétiques des loci *shp/rgg* a été réalisée. Actuellement, la validation de cibles prédites, *in vitro*, sur certains loci de streptocoques pathogènes ou GRAS est en cours au laboratoire.

2. Définitions

Pour faciliter la compréhension du vocabulaire utilisé dans ce chapitre, nous précisons ici le sens que nous donnons à quelques termes. On désigne par « squelette ADN » le motif consensus identifié dans les régions promotrices des gènes *shp* en ne prenant en compte que les nucléotides strictement conservés entre les différents motifs identifiés. Le terme « motif », quant à lui, fait référence à la séquence ADN présente dans les régions promotrices des gènes *shp* (et cibles putatives) prenant aussi en compte les nucléotides spécifiques du système SHP/Rgg auquel il se rapporte.

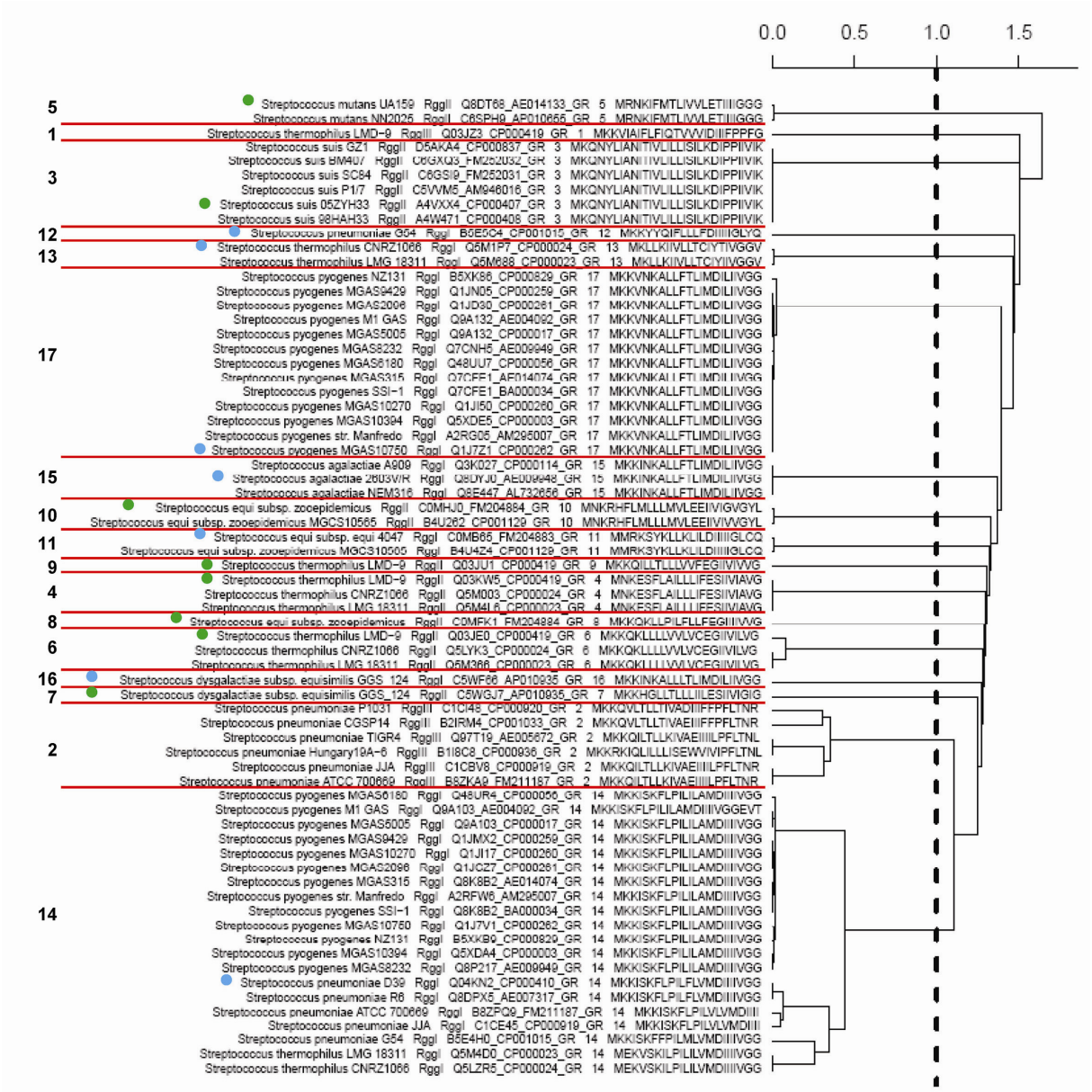


Figure 52 – Arbre de classification reflétant la divergence des sites synonymes au sein des protéines Rgg.

Les références arbitraires sont indiquées par un rond bleu ou vert, indiquant respectivement l'appartenance de la référence au groupe I ou II. Seulement deux clusters étant disponible au sein du groupe III, l'ensemble des régions inter-géniques a été utilisé lors des alignements ; par conséquent aucune référence n'a été désignée dans ce groupe de système SHP/Rgg. La ligne pointillée indique le seuil sélectionné.

Le nombre de « mismatches » correspond au nombre de différences, aux positions non déterminées par le squelette, entre l'occurrence du motif dans la région promotrice du gène *shp* et l'occurrence en amont d'un gène cible putatif.

3. Matériel et méthode

- **Identification de « squelettes » pour les motif(s) de fixation des protéines Rgg à l'ADN**

Afin d'éliminer les redondances entre les 68 systèmes SHP/Rgg identifiés dans l'article 1, susceptibles de perturber la recherche de motifs dans les séquences promotrices, une classification des systèmes SHP/Rgg en fonction de leur proximité évolutive a tout d'abord été construite. Il s'agit en effet d'éviter autant que possible les ressemblances en dehors des régions fonctionnelles soumises à pression de sélection (telles que les sites de fixations de facteurs de transcription). Pour contourner la difficulté de mesurer la distance évolutive directement entre les séquences intergéniques, dont l'alignement est difficile et pour lesquelles on ne connait pas la proportion de sites non soumis à sélection, nous avons utilisé une estimation du nombre de substitutions synonymes par site synonyme au sein des séquences nucléotidiques des protéines Rgg comme mesure de divergence (Yang and Nielsen, 2000). Finalement, 17 clusters de protéines Rgg, tels que le taux estimé de mutations synonymes par site soit au minimum de 1 entre deux clusters, ont été définis et une référence au sein de chaque cluster a été désignée de manière arbitraire pour l'analyse (Figure 52). Un alignement local des 17 régions promotrices des gènes *shp* associés aux références a alors été entrepris avec MEME (Bailey and Elkan, 1994) et comparés à un alignement global avec MUSCLE (Edgar, 2004) présenté au format ClustalW2. L'ensemble des références a tout d'abord été traité simultanément puis de manière séparée, c'est-à-dire en fonction de leur groupe SHP/Rgg d'appartenance. Les résultats obtenus ont permis de définir un squelette des motifs de site de fixation sous la forme d'un consensus « IUPAC » pour chacun des trois grand groupes de systèmes SHP/Rgg (I, II et III).

- **Recherche de sites de fixation des protéines Rgg à l'ADN dans les génomes**

La recherche de motifs a été entreprise sur la totalité des génomes des streptocoques séquencés et disponibles dans « Genome Reviews » à la date du 05.10.2010. Afin d'identifier les sites de fixation putatifs des protéines régulatrices Rgg à l'ADN, un script dédié en langage Perl a été écrit. Les consensus (squelettes) identifiés avec MEME et MUSCLE ont alors été recherchés sur une distance de 300 pb à partir de l'ATG des

	Séquence
Motif shp1358	GCATATATGGGAATA
Motif ster_1357	GCAAATATGGGAATA

Figure 53 – Comparaison des motifs ADN identifiés dans les régions promotrices des gènes cibles du locus *shp/rgg1358* de *Streptococcus thermophilus* LMD-9.
Le squelette ADN défini pour les motifs des systèmes SHP/Rgg du groupe II apparait en vert.

différentes ORFs présentes dans chaque génome séquencé, ainsi que sur les deux brins d'ADN et dans les deux sens. La distance de 300 pb établie dans cette étude, recouvre à la fois des régions non codantes et codantes. De plus, l'utilisation du logiciel bactgeneSHOW a permis de mettre en évidence les spCDS (**s**hort **p**utative **C**oding **D**NA **S**equence) présentes dans les régions du chromosome analysées.

Les systèmes SHP/Rgg d'un même groupe pouvant être présents en multi-copies dans une souche de streptocoque, plusieurs critères ont été pris en compte afin de pouvoir affiner la détection des cibles et être capable d'attribuer une cible putative identifiée à un système SHP/Rgg donné. Lors de l'étude du locus *shp/rgg1358* de *S. thermophilus* LMD-9 (article 1), la fixation de la protéine Rgg1358 sur les régions promotrices des deux gènes cibles du système, *shp1358* et *ster_1357*, a été démontrée sur des sondes ADN contenant un motif conservé correspondant au squelette ADN identifié dans cette étude pour les systèmes SHP/Rgg du groupe II (Figure 53). Bien qu'un squelette de 12 nucléotides ait été détecté pour ce groupe (voir ci-dessous), il est intéressant de souligner qu'une similitude de séquence plus longue peut être identifiée entre le motif d'un gène *shp* de référence et celui d'un gène cible associé. De plus, les séquences identifiées dans les régions promotrices des gènes *shp1358* et *ster_1357* ne diffèrent qu'au niveau d'un nucléotide (en position 4), alors que le squelette défini pour le groupe II autorise 4 différences. Cette faible variabilité entre les motifs est probablement en lien avec une spécificité de liaison du Rgg d'un système sur le motif ADN du gène cible du même système. C'est pourquoi, plutôt que de simplement identifier les séquences conforment aux squelettes, nous avons pour chaque système SHP/Rgg recherché de manière spécifique des occurrences du squelette ressemblant à celle trouvée dans la région promotrice du gène *shp* correspondant. Pour ceci, la première étape a consisté en une recherche de similitudes entre les régions amont des gènes *shp* et le reste du génome (300 pb à partir de l'ATG des différentes ORFs) avec une fenêtre glissante en autorisant un nombre de mismatches limité : 2 mismatches pour fenêtre de taille 12 (groupe II), 3 mismatches pour fenêtre de taille 14 (groupe I). Les données ont ensuite été filtrées pour ne garder que les séquences qui satisfont les contraintes « IUPAC » définies par le squelette (Figure 53). Pour la recherche de cibles putatives apparentées aux systèmes SHP/Rgg du groupe III, seul le squelette a été utilisé.

4. Résultats

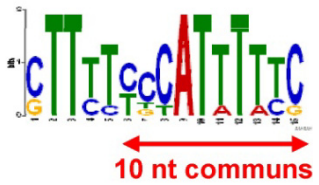
- **<u>Identification de trois motifs putatifs de fixation des protéines Rgg à l'ADN</u>**

Les gènes *shp* codent les phéromones nécessaires à l'activation des systèmes de QS SHP/Rgg. La caractérisation de plusieurs de ces systèmes (article1, résultats chapitre 2,

Tableau 22 – Alignements MEME et MUSCLE des références des groupes I, II et III.

MEME **MUSCLE**

Groupe I




```
RggI_11     AAGAAATAGAAATTCCTTTTTC----------TTCCCGCATTTTTCTATAGGTTTCACC
RggI_12     -----ATAGAAATCCCCTTTTC----------TTCTCTTATTTTCCTATAAGTTTCACC
RggI_13     ------------ATTTTTTTCAGTTCATTTTGTTTTCCCATATATGCAACCTTCTTTTA
RggI_14     ----ACACCAAGTTTTGATTTCA-------TAGTTTTCCCATTTTCCCAAC---------
RggI_16     ------------TCTTTTTCCA-------T--TTTTCCCACTTTCACAACA-------A
RggI_15     -----------ATTCTTTTTCCA-------T--TTTTCCCACTTTCACAAC---------
RggI_17     ------------TTCTTTTTCCA-------T--TTTTCCCACTTTCACAAC--------A
                              ** *           **  *  *  *    *
                                                 10 nt communs

RggI_11     TAAGATCATA-----GATTATCTCCGTT-TGTAGT-----TAAAAAGT---GGAAAGCGT
RggI_12     TAAAATCATA------CTTATCTCCATT-TACCAT--AAACAATCATT---AGAAAATGT
RggI_13     AAAAATCAAATCAAATGTTA--TCC--TATGTAAC----AAACTAAGTTGGAGGAAATCA
RggI_14     AAAAATTAACTTATTTCTTA--TTTATTGTACAAT--AAATAAAAAATGAAAGGAAGTCA
RggI_16     AAAAATTCAA--ATCCGTCAAATCCATTATATAATACAAGTACAAAAT---AGAAAATCA
RggI_15     --AATTTAAAAGACCCTTTAAATCCATTATATAATACAAGTACAAAAT---AGAAAATCA
RggI_17     AAAAATTAAA--ATCTGCTAAATCCATTATATAATACAAGTACAAAAT---AGAAAATCA
              *  *             *  *    * *           *   * *    * **

RggI_11     GACAA---------
RggI_12     GACCGATT------
RggI_13     T-------------
RggI_14     GAGT----------
RggI_16     GAAAGGAAGCTGTC
RggI_15     GAAAGGAGATAGTC
RggI_17     GAAAGGAGACAGTC
```

Suivi de l'ATG du gène *shp*

Groupe II



```
RggII_5     ---TATTCAGTCATCAATAGTACCTCAATTGTCTATATGGGATTTATAT-----TCAGTT
RggII_4     CCATAATCTTTCATAAAAAGCACCTCTTTTGCATATACGAGAATTTCCCAGA-TAAACTT
RggII_3     --------------------CCTCCTTTTGTATATAAGGGATTTGTTTAAAGCATAACT
RggII_10    ----AATGGTTCAAAACTACCACCTCACTTGTACATACGGGAAAGACTT-----ATCATT
RggII_6     CCAAATTCTTTCATGG-AAGCCCCCAGTTTGCATATATGGGAATA---T-----GTATTT
RggII_8     ----AATTTTTCATATCACACCTCAATCACGCATATATGGGAATATTTT-----TTGATT
RggII_7     ----AATTTTTCATTTTAAACCTCTATCACGCATATATGGGAATAAATT-----TAAATT
RggII_9     ----ATTTTTTCATACTACACCTCCATCACGCATATATGGGAATAATTT-----TCATTT
                                          *   *** * **                *

RggII_5     AATTGCAATTTATTATAACATATAAGAAAA--------------GTGAGTGAGGAGAAAA
RggII_4     AAAAGTATTATATCATATAGTGAGATAGGA---------------------GGTTAATT
RggII_3     TAATAGATGTTATCATGATTTTAAAATAAATATTATCTATATTCAACAATGGGGAATATC
RggII_10    AGTTCTATCATAACATACACTATAGAAAAA-------------CAGAAGTGAGGATTATC
RggII_6     TTTTGTATCTTAACATAAAATACAGTAAAA-------------CAAAAGTGAGGTAATTC
RggII_8     TTTAACATTTTAGCATAAAATAGAAAAAAA-------------TAAAAGTGAGGTATTTC
RggII_7     AACTTCATAATAGCATACAATATGAAAAAA-------------TAAAAGTGAGGATTATC
RggII_9     GCTTTCATCATATCATACAATAAAGAAAAA-------------CAAAAGTGAGGATTATC
                  *   **  **    *        *                      **

RggII_5     AC-
RggII_4     TCT
RggII_3     A--
RggII_10    AC-
RggII_6     AT-
RggII_8     AC-
RggII_7     AC-
RggII_9     AC-
```

Suivi de l'ATG du gène *shp*

Tableau 22 (suite) – Alignements MEME et MUSCLE des références des groupes I, II et III.

	MEME	MUSCLE
Groupe III		RggIII_1 CACATATGCAAAAATTAAGAAAAAGTGGTAATAGCTATATACCTCT---TTTTCTTGTTC RggIII_2 ----------ACACCACTAGAAACATAATCTTGTTCAGTTGCA--ACTT RggIII_2 --TTCCCAAATTCCCCTCCATAAACTGTCTCCAACTCAGCTTCTGTCAATTCCATTGTTT RggIII_2 TATTCCCAAATTCCCCTCCATAAACTATCTCCAACTCAGCTTCTGTCAATTCCATTGTTT RggIII_2 TATTCCCAAATTCCCCTCCATAAACTATCTCCAACTCAGCTTCTGTCAATTCCATTGTTT RggIII_2 AATTTACGAGGTTGCCTCCTCTAACATCTTGCAACTCAGCTTCTGTCAATTCCATTGTTT RggIII_2 AATTTACGAGGTTGCCTCCTCTAACATCTTGCAACTCAGCTTCTGTCAATTCCATTGTTT * * * * * RggIII_1 TTAGATTAACAAGCAATAAAGCAACCGCCATTACATACTTGTTCTT-----------CTA RggIII_2 CAGTCACATCAAGTTCCAATGAAA-----ATTCTGTATTTTCCATAATAAACGATTCCTC RggIII_2 CTG------CAAATTGTAA----------ATTTAACATCTTTTACA-----------CTC RggIII_2 CTG------CAAATTGTAA----------ATTTAACATCTTTTACA-----------CTC RggIII_2 CTG------CAAATTGTAA----------ATTTAACATCTTTTACA-----------CTC RggIII_2 CTG------CAAATTGTAA----------ATTTAACATCTTTTACA-----------CTC RggIII_2 CTG------CAAATTGTAA----------ATTTAACATCTTTTACA-----------CTC *** ** *** * * ** RggIII_1 CTTCGA-TATCTTCGATAAACATATGCCTGTTCTCCTTTCGTTATGATATCC--CAA--- RggIII_2 CTCCAATTATTTTTACT---TATAAACCACTTCTGCGACC--TAGGATTTGCTTCAAGTG RggIII_2 CTTCAATTATCTTCATT---TATAAACCACTTCTGCGACC--TAGGATTTGCTTCAAGTG RggIII_2 CTTCAATTATCTTCATT---TGTAAACCACTTCTGCGACC--TAGGATTTGCTTCAAGTG RggIII_2 CTTCAATTATCTTCATT---TGTAAACCACTTCTGCGACC--TAGGATTTGCTTCAAGTG RggIII_2 CTTCAATTATCTTCATT---TGTAAACCACTTCTGCGACC--TAGGATTTGCTTCAAATG RggIII_2 CTTCAATTATCTTCATT---TGTAAACCACTTCTGCGACC--TAGGATTTGCTTCAAGTG ** * * *** ** * ** ** **** * * ** *** * * *** RggIII_1 --TTATAACTTATTTATGTCACAAAA----------TTAAAAAATTGAAGATATGAAAT RggIII_2 CTTTATAAGTACAGTATAACACGAAAATTGGCTTATTTTAGAAAATCGCATATTTGATAT RggIII_2 CTTTACAAGTACAGTATAACACGAAAATTGGCTTATTTTAAAAAATCGCATATTTGATAT RggIII_2 CTTTACAAGTACAGTATAACACGAAAATTGGCTTATTTTAGAAAATCGCATATTTGATAT RggIII_2 CTTTACAAGTACAGTATAACACGAAAATTGGCTTATTTTAGAAAATCGCATATTTGATAT RggIII_2 TTTTACAAGGACAGTATAACACGGAAATTGGCTTATTTTAGAAAATCGCATATTTGATAT RggIII_2 CTTTACAAGTACAGTATAACACGAACATTGGCTTATTTTAGAAAATCGCATATTTGATAT *** ** *** *** * *** ***** * * ** *** ** RggIII_1 TTTTTAATAAATTTATTTCATATCTTCAATTTTGCTGAA-TTGATAAATTTTATGTTAGT RggIII_2 TTTTTCTTATAAAAATTTCACATTTGCGATTTTAGTAGATTTGATCATTTCCCTG---GT RggIII_2 TTTTTCTTATAGAAATTTCTTATTTGCGATTTTGATAGATTTGATTATTTCCCTG---GT RggIII_2 TTTTTCTTATAGAAATTTCAGATTTGCGATTTTGGTGAATTTGATTACTTCTCTG---GT RggIII_2 TTTTTCTTATAGAAATTTCAGATTTGCGATTTTGGTGAATTTGATTACTTCTCTG---GT RggIII_2 TTTTTCTTATAGAAATTTCAGATTTGCGATTTTGGTAGATTTGATTATTTCCCTG---GT RggIII_2 TTTTTCTTATAGAAATTTCAGATTTGCGATTTTGGTGAATTTGATTACTTCTCTG---GT ***** ** * ***** ** * * ***** * * ***** * ** ** ** RggIII_1 ATAACAGTCAAACTATTGATTAGGAG--GATACAGC-- RggIII_2 ATAATAAAGTTATTACTAATGAGGAGTGGAAGTCAT-- RggIII_2 ATAATAAAGTTATTACTAACAAGGAGG-AATATAATAG RggIII_2 ATAATAAAGTTACTACTAATGAGGAGT-GGAGAAAT-- RggIII_2 ATAATAAAGTTACTACTAATGAGGAGT-GGAGAAAT-- RggIII_2 ATAATAAAGTTATTACTAATGAGGAGG-GAAGTAAT-- RggIII_2 ATAATAAAGTTACTACTAATGAGGAGT-GGAGAAAT-- **** * * ** * * ***** Suivi de l'ATG du gène *shp*

Tableau 23 – Squelettes ADN des motifs putatifs de fixation des régulateurs Rgg à l'ADN identifiés dans les régions promotrices des gènes *shp*.

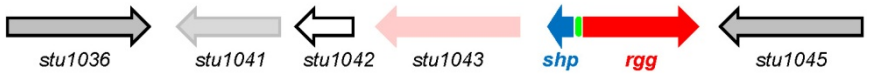
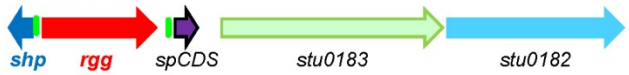
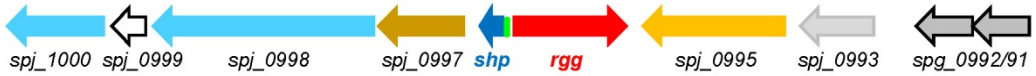
	Groupe I	Groupe II	Groupe III
Séquence (5'→ 3')	CCCAnnTnnnCAAC	GnnnATAnGGGA	AAAATnGnAnATnnGAnAT

(Chang *et al.*, 2011) a souligné l'existence d'une boucle de régulation positive de ces gènes nécessaire à la fonctionnalité du mécanisme décrit sur certains systèmes SHP/Rgg. Nous avons alors émis l'hypothèse que le gène *shp* était une cible de chaque système SHP/Rgg. Aussi, dans le but de définir un motif de fixation des protéines Rgg à l'ADN, un alignement des 17 régions promotrices des gènes *shp* (références) a été entrepris en utilisant MEME et MUSCLE.

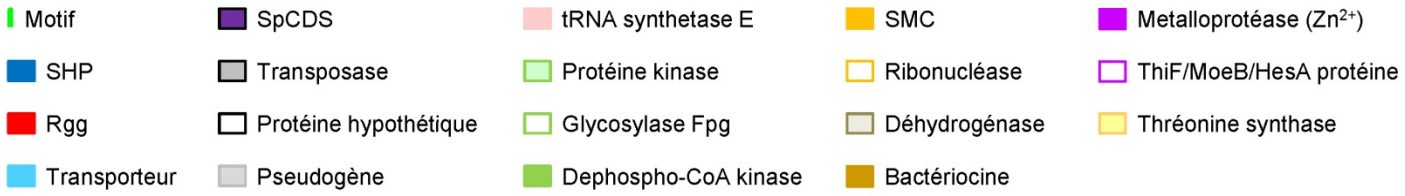
L'alignement des 17 régions promotrices de référence n'a pas permis d'identifier un motif de fixation putatif, des protéines Rgg à l'ADN, commun aux trois groupes de systèmes SHP/Rgg (données non présentées). Cette observation laissant présager des spécificités dans les régions promotrices des gènes *shp*, une deuxième analyse des régions promotrices des gènes *shp* a alors été réalisée en fonction du groupe d'appartenance des références : groupe I (7 références), groupe II (8) et groupe III (Figure 52). La complémentarité des alignements réalisés sous MEME et MUSCLE a rapidement permis d'identifier un motif ADN conservé dans les régions promotrices des gènes *shp* du groupe II (Tableau 22), dont le squelette ADN est présenté dans le Tableau 23. L'alignement des régions promotrices des gènes *shp* de référence du groupe I, sous MEME et MUSCLE, a fait ressortir deux séquences ADN partageant en commun 10 nt (Tableau 22). L'alignement sous MUSCLE a, par ailleurs, fait ressortir une divergence des séquences ADN des références 11 et 12 par rapport aux autres références du groupe. Ces séquences étant similaires entre elles, il nous a semblé opportun, pour la suite de l'analyse, de les traiter séparément des autres références du groupe et de ne pas les prendre en compte dans la définition du squelette ADN apparenté à ce groupe de systèmes SHP/Rgg. Par ailleurs, les séquences des phéromones SHP associées à ces systèmes présentent une extrémité C-terminale atypique (Tableau 21). Aussi, seulement le squelette ADN majoritairement identifié au sein des régions promotrices des références du groupe I est présenté dans le Tableau 23. Enfin, le groupe III ne contenant que deux références, les alignements ont été effectués sur l'ensemble des séquences appartenant à ce groupe et s'élevant au nombre de sept. L'analyse des alignements réalisés sous MEME a fait ressortir un motif ADN conservé au sein de l'ensemble des séquences (Tableau 22). Le report de ces motifs sur les alignements de MUSCLE a alors fait ressortir un deuxième motif strictement conservé pour la référence Rgg_1 ou se différenciant au niveau de 2 ou 3 nucléotides pour les références appartenant au cluster Rgg_2. L'alignement sous MUSCLE souligne également que le motif ADN, dont le squelette ADN est présenté dans le Tableau 23, est conservé sur 19 nucléotides.

L'ensemble de ces résultats a permis de déterminer un squelette ADN spécifique au sein de chaque groupe SHP/Rgg (Tableau 23). Quelque soit le groupe étudié, une tendance à la

Tableau 24 – Environnement génétique des loci *shp/rgg* associés au groupe I.

Espèce	Nb[a]	Souche[b]	Gène *rgg*	Séquence SHP	Environnement génétique
S. thermophilus	2/3*	LMG18311	*stu1044*	MEKVSKILPILILVMDIIIIVGG	
	2/3	LMG18311 CNRZ1066	*stu0182* *str0182*	MKLLKIIVLLTCIYIIVGGV MKLLKIIVLLTCIYTIVGGV	
S. agalactiae	3/3	NEM316	*gbs1555*	MKKINKALLFTLIMDILIIVGG	
S. dysgalactiae	1/1	subsp. equisimilis GGS_124	*sdeg_0529*	MKKINKALLLTLIMDILIIVGG	
S. pyogenes	13/13	M1GAS	*spy_0496*	MKKVNKALLFTLIMDILIIVGG	
	13/13	M1GAS	*spy_0533*	MKKISKFLPILILAMDIIIIVGGEVT	
S. pneumoniae	3/9	R6 G54	*spr0960* *spg_0976*	MKKISKFLPILILAMDIIIIVGG MKKISKFFPILMLVMDIIIIVGG	
	2/9	JJA	*spj_0996*	MKKISKFLPILVLVMDIIII	
	1/9	G54	*spg_1268*	MKKYYQIFLLLFDIIIIIGLYQ	
S. equi	2/3	subsp. zooepidemicus MGCS10565	*sez_0050*	MRKSYKLLKLILDIIIIIGLCQ	

a. Nb représente le nombre de souche(s) possédant le locus représenté sur l'ensemble des souches séquencées appartenant à l'espèce apparenté.
b. Lorsque plusieurs souches présentent le même environnement génétique, celui-ci est présenté sur un locus choisi de manière arbitraire.
* Le locus est tronqué dans la LMD-9 (DIIIGG).



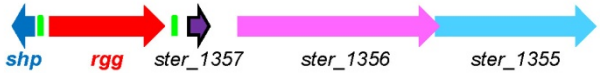
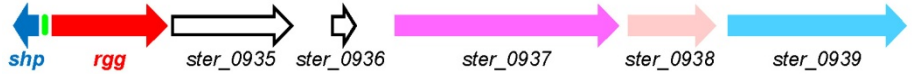
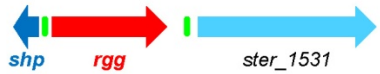
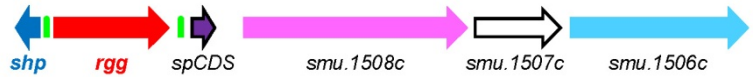
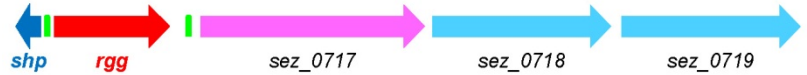
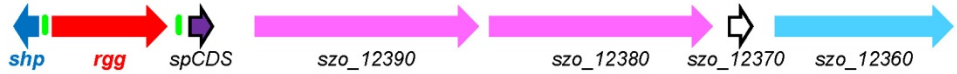
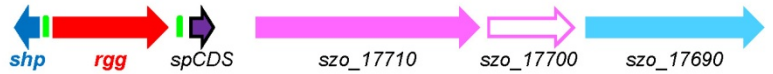
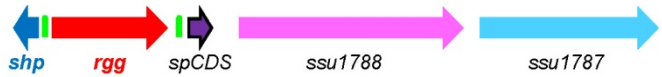
conservation de ces différents motifs en amont du promoteur des gènes *shp*, au niveau des boites -10 et -35, a été observée. Par ailleurs, l'analyse approfondie des motifs des groupes I et II laissent entrevoir une structure de base inversée commune de type « CCC-n_7-C » pour le groupe I et « G-n_7-GGG » pour le groupe II (Tableau 23). Aussi, bien que l'organisation génétique des systèmes SHP/Rgg diffère entre les groupes I/II et III, leur mécanisme de régulation semble au minimum présenter un caractère conservé.

- **Les systèmes SHP/Rgg semblent être impliqués dans la régulation de cibles distales et proximales**

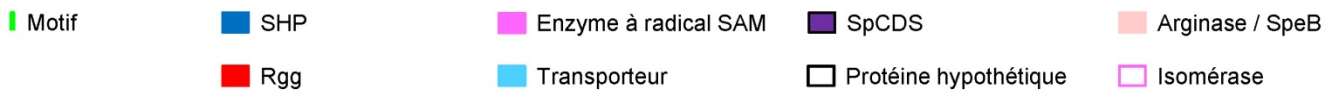
L'analyse des données issues de la recherche de cibles a fait ressortir deux grandes catégories de cibles putatives associées aux systèmes SHP/Rgg : (i) des cibles proximales, correspondant au gène localisé en aval du gène *rgg* pour les systèmes du groupe I ou II ou en amont du gène *shp* pour les systèmes du groupe III et/ou (ii) des cibles distales ont pu être détectées. L'ensemble des gènes cibles putatifs des systèmes SHP/Rgg des groupes I, II et III sont respectivement présentés en Annexe 6A, B et C. Il convient par ailleurs de préciser, que bien qu'ils aient été systématiquement détectés lors de l'analyse, les gènes *shp* et *rgg* du locus étudié ont été retirés de la liste des cibles putatives.

Un premier constat de cette analyse concerne la classification de ces deux catégories de cibles au sein des différents groupes de systèmes SHP/Rgg. Alors que des cibles distales ont été détectées pour les systèmes du groupe I et II (voir ci-dessous), la régulation de cibles proximales semble être, quant à elle, spécifique aux systèmes des groupes II et III. Quelques exceptions sont cependant observées. En effet, pour deux loci paralogues (*stu0182 et str0182*) chez *S. thermophilus* appartenant au groupe I (Tableau 24), un motif a été identifié en amont de la spCDS détectée par le logiciel BactgeneSHOW. Ces deux systèmes sont associés à une phéromone SHP atypique possédant un résidu cystéine à la place du résidu aspartate conservé. Une deuxième exception, concerne le locus *ster_0934* (groupe II) de *S. thermophilus* LMD-9 (et ses paralogues au sein des autres souches de l'espèce) où aucun motif n'a été détecté dans la région promotrice du gène proximal (Tableau 25). Cette exception peut néanmoins s'expliquer par la présence d'une région intergénique « gène *rgg* / gène aval proximal » plus courte que celle identifiée pour les autres systèmes SHP/Rgg du groupe II chez *S. thermophilus* LMD-9 (59 pb vs 82, 83, 84 et 116 pb). Bien qu'un motif ne soit pas été détecté en amont du gène *ster_0934*, une co-transcription du gène *ster_0934* avec le gène *rgg* peut néanmoins être envisagée. L'analyse de l'environnement génétique des loci *shp/rgg* – des systèmes appartenant aux groupes II et III – souligne, par ailleurs, une homogénéité des environnements génétiques au sein de ces deux groupes (Tableaux 25 et

Tableau 25 – Environnement génétique des loci *shp/rgg* associés au groupe II.

Espèce	Nb[a]	Souche[b]	Gène *rgg*	Séquence SHP	Environnement génétique
S. thermophilus	1/3	LMD-9	*ster_1358*	MKKQILLTLLLVVFEGIIVIVVG	
	2/3	LMG18311	*stu1950*	MNKKALFSLLFVILEGIIVIGVG	
	3/3	LMD-9	*ster_0934*	MNKESFLAILLLIFESIIVIAVG	
	3/3	LMD-9	*ster_1530*	MKKQKLLLLVVLVCEGIIVILVG	
S. mutans	2/2	UA159	*smu.1509*	MRNKIFMTLIVVLETIIIIGGG	
S. dysgalactiae	1/1	subsp. equisimilis GGS_124	*sdeg_1013*	MKKHGLLTLLLIILESIIVIGIG	
S. equi	1/3	subsp. zooepidemicus MGCS10565	*sez_0716*	MNKRHFLMLLLMVLEEIIVIVVGYL	
	1/3	subsp. zooepidemicus	*szo_12400*	MNKRHFLMLLLMVLEEIIVIGVGYL	
	1/3	subsp. zooepidemicus	*szo_17730*	MKKQKLLPILFLLFEGIIIIVVG	
S. suis	6/6	P1/7	*ssu1789*	MKQNYLIANITIVLILLISILKDIPPIIVIK	

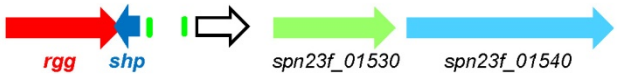
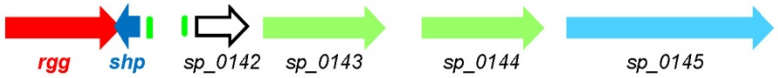
a. Nb représente le nombre de souche(s) possédant le locus représenté sur l'ensemble des souches séquencées appartenant à l'espèce apparenté.
b. Lorsque plusieurs souches présentent le même environnement génétique, celui-ci est présenté sur un locus choisi de manière arbitraire.



26). La présence de cibles proximales putatives est corrélée à la présence d'un environnement génétique conservé. En effet, la comparaison de ces environnements génétiques met en évidence l'association de spCDS ou de petits gènes avec des métallo-enzymes impliquées dans des modifications post-traductionnelles (enzymes à radical SAM (Fontecave *et al.*, 2004), CAAX (Pei and Grishin, 2001), oligoendopeptidase F et des transporteurs (perméases ou ABC-transporteurs). Enfin, la comparaison des environnements génétiques (Tableau 25) et des séquences peptidiques des peptides codés par les spCDS (Annexe 6B) souligne, que les loci *smu.1509* de *S. mutans* UA159 (et son paralogue) et *sez_17730* de *S. equi* subps. *zooepidemicus* (et son paralogue) seraient orthologues.

Une autre particularité au sein de ces différents systèmes contrôlant la transcription de gènes cibles proximaux putatifs est la présence d'une distance relativement conservée « motif / codon start du gène *shp* » et « motif / codon start du gène cible proximal putatif » (≈ 77 pb) (Annexe 6B et C). Cette distance s'avère également conservée pour certains gènes cibles distaux, en particulier chez *S. thermophilus* et dans une moindre mesure chez *S. equi* (Annexe 6B). Aussi, il est envisageable que les systèmes acquis au sein des souches, aient évolué dans le but de contrôler d'autres fonctions. L'environnement génétique de ces cibles distales est représenté dans le Tableau 27. De manière intéressante, le locus *ster_1923* de *S. thermophilus* LMD-9 (et ses paralogues) présente un environnement génétique similaire à celui décrit pour la présence de cibles proximales, c'est-à-dire composé de gènes codant un peptide, une métallo-enzyme et un transporteur. Les environnements génétiques correspondant aux deux autres cibles distales de *S. thermophilus* sont composés (i) d'une protéine de liaison des acides nucléiques, d'un régulateur de réponse et d'un pseudogène correspond à une histidine kinase et (ii) de gènes codant des déoxy et ribonucléases formant probablement une voie de dégradation de l'ADN et de l'ARN. De plus, en amont du gène *ster_0353*, un environnement génétique composé également d'une métallo-enzyme et d'un transporteur est visible (Tableau 27). Au sein de l'espèce *equi*, les environnements génétiques des cibles distales impliquent tous deux, au vu des annotations dans GenBank, un régulateur transcriptionnel de la famille RNPP ou apparenté. En effet, ce régulateur appartient soit à la famille des Rgg (MutR) et est suivi de deux gènes codant un transporteur (locus *sez_0631*), soit à la famille des PlcR (locus *sez_0517*). Dans ce dernier cas, une spCDS en amont du gène *plcR* a été détectée. De plus, des gènes codant des protéines impliquées dans la maturation et la sécrétion d'un peptide phéromone de compétence chez certains streptocoques constituent l'environnement aval du gène codant PlcR. Aussi, ce locus pourrait représenter une voix de régulation de la compétence où le gène annoté *plcR* serait activé par le peptide en amont afin de contrôler l'expression de la phéromone, nécessaire au déclenchement de la compétence chez *S. equi*.

Tableau 26 – Environnement génétique des loci *shp/rgg* associés au groupe III.

Espèce	Nb[a]	Souche[b]	Gène *rgg*	Séquence SHP	Environnement génétique
S. thermophilus	1/3	LMD-9	*ster_1299*	MKKVIAIFLFIQTVVVIDIIIFPPFG	
S. pneumoniae	3/9	ATCC700669	*spn23f_01510*	MKKQILTLLKIVAEIIIILPFLTNR MKKRKIQLILLLISEWVIVIPFLTNL	
	2/9	TIGR4	*sp_0141*	MKKQVLTLLTIVADIIIFFPFLTNR MKKQILTLLKIVAEIIIILPFLTNL	
	1/9	CGSP14	*spcg_0143*	MKKQVLTLLTIVAEIIIFFPFLTNR	

a. Nb représente le nombre de souche(s) possédant le locus représenté sur l'ensemble des souches séquencées appartenant à l'espèce apparenté.
b. Lorsque plusieurs souches présentent le même environnement génétique, celui-ci est présenté sur un locus choisi de manière arbitraire.

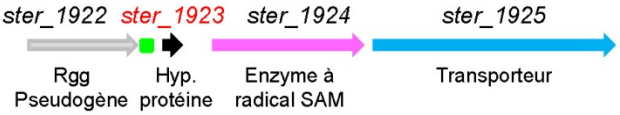


De manière transversale aux groupes de systèmes SHP/Rgg I et II, des cibles distales, où la distance « motif / codon start du gène cible putatif » n'est pas conservée avec celle « motif / codon start du gène *shp* » et varie de 3 à 288 pb, ont été détectées (Annexe 6A et B). Les gènes cibles putatifs sont décrits comme codant des protéines sécrétées ou de surface (pouvant correspondre à des facteurs de virulence), des régulateurs transcriptionnels, des protéines impliquées dans la division cellulaire, la maintenance cellulaire ou la biosynthèse du peptidoglycane, des composants de la machinerie de traduction, des enzymes impliquées dans différentes voies métaboliques (glucidiques, lipidiques et acides aminés), des composants de systèmes de transport ou des éléments de transposition (transposases, phages). De nombreuses méthyltransférases, différentes des enzymes à radical SAM précédemment décrites, semblent également être contrôlées. Enfin, un grand nombre de protéines de fonctions inconnues a aussi été détecté.

Par ailleurs, aucune cible n'a pu être détectée au sein des clusters 11 et 12, dont les références utilisées lors de la recherche de motifs ADN présentaient une variabilité nucléotidique au niveau de la région promotrice des gènes *shp*. D'autres motifs, spécifiques aux régions promotrices des gènes *shp* de ces deux références, ont par ailleurs été testés mais se sont également avérés sans résultats (données non présentées).

Certains streptocoques présentent la particularité d'accumuler plusieurs systèmes SHP/Rgg dans leur génome. Il est intéressant de relever que lorsque plusieurs systèmes SHP/Rgg sont présents au sein d'une même souche, les gènes *shp* et les cibles proximales des autres systèmes sont en partie également détectés. En effet, les exceptions s'expliquent par un nombre de mismatches supérieur à celui fixé pour l'analyse. Ce phénomène, visible en particulier au sein de l'espèce *thermophilus* (espèce ayant accumulée le plus grand nombre de systèmes SHP/Rgg), souligne une forte conservation des motifs au sein d'une souche. De plus, bien que plusieurs espèces (*thermophilus*, *dysgalactiae* et *equi*) possèdent des systèmes SHP/Rgg appartenant aux groupes I et II, la détection d'un système SHP/Rgg du groupe I par un motif du groupe II est seulement visible au sein de l'espèce *thermophilus*. Ceci est explicable, pour l'espèce *equi*, par la non-détection de cible lors de l'analyse des systèmes SHP/Rgg du groupe I au sein de cette espèce. Bien que ce résultat était envisageable étant donné les paramètres de l'analyse (autorisant la recherche de motifs sur les deux brins et dans les deux sens), le fait qu'un motif du groupe I ait pu être détecté *via* une recherche à partir d'un motif du groupe II chez *S. thermophilus* suggère l'évolution des motifs I et II à partir d'un motif « ancestral ». Par conséquent, la structure de base inversée de type « CCC-n_7-C » pour le groupe I et « G-n_7-GGG » pour le groupe II, et donc l'orientation du motif, pourrait intervenir dans la régulation des gènes cibles.

Tableau 27 – Environnement génétique des gènes cibles distaux présentant une distance « motif / codon start » conservée.

Espèce	Gène *rgg*	Environnement génétique
S. thermophilus	*ster0353* *stu0316* *str0316*	
	ster_1774 *stu1801* *str1801*	
	ster_1923 *stu1951* *str1951*	
S. equi	*sez_0631* *szo_13060*	
	sez_0517 *szo_14750*	

- **Localisation des motifs de fixation des protéines Rgg à l'ADN dans des régions codantes**

La détection des motifs au sein des trois groupes de systèmes SHP/Rgg a également mis en évidence une localisation de ces derniers au sein de nombreuses régions codantes (Annexe 7A, B, C). En effet, les motifs ont été recherchés sur une distance de 300 pb en amont des démarrages de traduction et dans certains cas, ils ont pu être localisés dans l'ORF localisé en amont du gène en question. De manière intéressante, la majorité des gènes dans lesquels un motif a été détecté codent des fonctions similaires à celles identifiées pour les gènes dont les motifs ont été détectés dans les régions promotrices ou intergéniques, qui seront à présents appelées non-codantes. A titre d'exemple, le gène *upp* impliqué dans la synthèse du peptidoglycane est détecté chez *S. thermophilus* (vs *S. pneumoniae* pour région non-codante) et des gènes impliqués dans la division cellulaire *fts* et l'opéron *lac* ont été détectés chez *S. pyogenes* et *S. pneumoniae* (vs *S. suis*). De plus, de nombreuses méthyltransférases et éléments de transposition ont pu être détectés. Enfin, des gènes dont les produits sont impliqués dans la machinerie d'internalisation de l'ADN, dans la réponse au stress et dans un composant de systèmes de transport (d'oligopeptides, ATPases) ont également été détectés lors de l'analyse.

5. Discussion

Ce travail a été entrepris dans le but de définir un motif de fixation, des protéines régulatrices Rgg associées à des phéromones SHP, à l'ADN et d'identifier le régulon de chacun des systèmes SHP/Rgg.

- **Motifs de fixation des protéines Rgg à l'ADN**

L'alignement des régions promotrices des gènes *shp* a conduit, dans un premier temps, à l'identification de trois motifs putatifs de fixation des protéines Rgg à l'ADN pour lesquels, un squelette de base spécifique à chaque groupe de systèmes SHP/Rgg a été défini (Tableau 23). La difficulté de faire ressortir un motif conservé, au sein de chaque groupe, présage d'une évolution de ces derniers au sein des streptocoques. Aussi, il est probable que les nucléotides strictement conservés au sein de ces motifs soient requis et essentiels à la fixation du Rgg à l'ADN alors que, la variabilité de certaines positions joue un rôle dans la spécificité et l'affinité de la liaison Rgg/ADN. Des expériences de mutagénèse dirigée couplée à de la SPR – afin de mesurer les affinités de fixation des protéines Rgg à l'ADN (complexées ou non à leur phéromone SHP) – permettraient de confirmer cette hypothèse.

Au premier abord, une étroite spécificité de fixation d'un Rgg sur son motif ADN, empêchant possiblement tout phénomène de cross-activation entre deux Rgg d'un même groupe présent au sein d'une même souche peut être suggérée. En effet, toujours sur la base du locus *shp/rgg1358* de *S. thermophilus* LMD-9 (article 1), l'absence d'activité du promoteur $P_{shp1358}$ dans des mutants respectifs des gènes *shp1358* et *rgg1358* ne présage d'aucune cross-activation *via* un autre système au sein de cette souche. Aussi, bien que les cibles d'un système SHP/Rgg peuvent être détectées *via* le motif d'un autre système SHP/Rgg, la similarité de deux motifs ADN ne semble pas être suffisante à l'expression des cibles. Ceci est d'autant plus vrai que le motif du locus *shp/rgg1358* est identique à celui du locus *shp/ster_1530* chez *S. thermophilus* LMD-9 (Annexe 6B). Cette observation suggère alors (au minimum pour *S. thermophilus* LMD-9) que la spécificité de liaison de la phéromone SHP avec son régulateur Rgg apparenté est essentielle à l'activation du régulon associé. Néanmoins, on ne peut exclure que l'expression de ces systèmes, dans le temps, ne soit pas coordonnée et ne puisse par conséquent permettre une éventuelle cross-activation ou cross-inhibition.

Il conviendrait alors, dans un premier temps, de construire une fusion transcriptionnelle au promoteur $P_{shp1530}$ afin de mesurer son activité au cours du temps et de la comparer à celle du promoteur $P_{shp1358}$. Dans un deuxième temps, et en fonction des résultats, il pourrait être intéressant de complémenter les mutants respectifs des gènes *rgg* de ces deux loci, présentant une fusion transcriptionnelle au promoteur P_{shp} apparenté, avec un plasmide surproducteur de chaque protéine Rgg (Rgg1358 et Rgg1530). Par ailleurs, un motif du groupe I (CCC-n_7-C) pouvant être détecté par un motif du groupe II (G-n_7-GGG) chez *S. thermophilus* LMD-9, il serait également intéressant d'étudier l'activation des promoteurs en inversant le sens des différents motifs ADN.

La distance « motif / ATG du gène cible putatif » souligne une étroite régulation des gènes cibles proximaux, présentant une distance équivalente à celle de la distance « motif / ATG du gène *shp* apparenté ». Néanmoins, la régulation de cibles putatives ne semblent pas corréler, dans tous les cas, à une distance « motif / ATG » conservée. Dans cette étude, la recherche de motifs étant limitée à une distance de 300 pb à partir des démarrages de traduction, des distances comprises entre 3 à 288 pb ont pu être mises en évidence. Le régulon PlcR est composé de 138 gènes dont les distances « motif / ATG » ont été décrites comme pouvant aller de 35 à 700 pb (Agaisse *et al.*, 1999; Gohar *et al.*, 2008; Økstad *et al.*, 1999). Aussi, il serait intéressant d'allonger cette distance afin de voir si d'autres cibles des systèmes SHP/Rgg pourraient être identifiées. De plus, l'étude de Gohar *et al.* (2008) a permis de corréler la détection de faux-positifs à un environnement amont/aval de la PlcR box moins riche en AT. Il serait intéressant de voir dans notre cas, si cet environnement peut

également influencer la fixation des protéines régulatrices Rgg à l'ADN. Enfin, le motif pouvant être localisé entre la région promotrice et le RBS des gènes cibles putatifs, un rôle répresseur des protéines régulatrices Rgg (associées aux phéromones SHP) peut être proposé dans certains cas.

- **Prédiction des régulons SHP/Rgg**

L'analyse des données concernant la recherche de cibles putatives indique que les régulateurs Rgg des groupes I et II pourraient être pléiotropes. En effet, les cibles détectées sont susceptibles d'être impliquées dans de nombreuses fonctions physiologiques dont certaines sont vitales pour la cellule (*e.g.* division cellulaire). En revanche, les régulateurs Rgg appartenant au groupe III semblent être plus spécifiques, aucune cible distale (motif dans région non-codante) n'ayant été détectée. Chang *et al.* (2011) ont démontré que les deux systèmes SHP/Rgg chez *S. pyogenes* sont impliqués de manière antagoniste dans la formation de biofilms chez cette espèce. Aussi, il était envisageable que les cibles de ces deux systèmes soient identiques. Dans cette étude, et bien que certaines cibles putatives soient communes aux deux systèmes SHP/Rgg, d'autres en revanche semblent être spécifiques d'un seul des deux systèmes. Par conséquent, ces deux systèmes, bien qu'antagonistes pour certaines fonctions ou phénotypes, pourraient réguler de manière spécifique et indépendante un certain nombre de cibles. Par ailleurs, le régulon putatif identifié au sein de chaque souche d'une même espèce semble présenter des particularités intra-souches. Cette variabilité du régulon, plus facilement visible chez *S. pyogenes* au vu des nombreuses souches séquencées au sein de cette espèce, peut s'expliquer par le nombre important d'éléments de transposition et de protéines phagiques qui seraient contrôlés par ces systèmes (Annexes 6 et 7) (Anbalagan *et al.*, 2011). Cette variabilité du régulon a déjà été décrite pour RopB (Rgg stand-alone) chez *S. pyogenes* (Anbalagan *et al.*, 2011; Dmitriev *et al.*, 2008). Cependant, on ne peut exclure, dans cette étude, l'influence du seuil sur la détection des cibles. Enfin, il est important de souligner, qu'aucune relation entre l'étude réalisée sur la protéine régulatrice « RovS » (Samen *et al.*, 2006) et notre analyse ne peut être établie. En effet, Samen et ses collaborateurs ont montré la fixation de RovS sur les régions promotrices de plusieurs gènes cibles présentant un motif ADN conservé différant de celui décrit dans notre étude (Introduction bibliographique Chapitre 3, § 2.2.2). De plus, bien que les données actuelles associent une phéromone SHP à ce régulateur Rgg, ce dernier avait, par conséquent, été étudié en temps que « stand-alone ». Bien que notre étude ne remette pas en cause les résultats de cet article, elle souligne néanmoins la possibilité que des régulateurs Rgg se fixent sur différents motifs. En effet, les régulateurs Rgg pourraient être capables de se fixer à un premier motif en absence de leur phéromone

SHP et d'en fixer un second en présence de ces dernières. Cette constatation est concordante avec les résultats de l'analyse ChIP-Chip effectuée sur le Rgg stand-alone « RopB » de *S. pyogenes*, n'ayant pas permis de définir *in fine* un motif consensus de fixation de ce Rgg à l'ADN (Anbalagan *et al.*, 2011).

L'analyse des environnements génétiques des loci appartenant aux groupes II et III a fait ressortir une homogénéité au sein de ces derniers composés de spCDS ou petites ORF, métallo-enzymes (enzymes à radical SAM, CAAX, endopeptidases F) et transporteurs. La présence de cet environnement conservé avait déjà été souligné, sur les loci connus en 2007 au sein du groupe II, lors de la découverte des systèmes SHP/Rgg par les travaux de Ibrahim *et al.*, (2007b). Aussi, ce travail élargit cette tendance aux nouveaux loci appartenant au groupe II mais également à ceux du groupe III détectés par l'analyse *in silico* présentée dans l'article 1. La présence de gènes codant des enzymes à radical SAM et CAAX – au sein d'opérons impliqués dans des voies de biosynthèse de bactériocines – a par ailleurs déjà été décrite dans la littérature (Pei. 2001 ; Diep 1996 ; Rea 2011) et s'illustre, dans cette étude, par un système SHP/Rgg appartenant au groupe III (locus *spcg_0143* de *S. pneumoniae* CGSP14) (Tableau 26). Aussi, la détection de spCDS ou la présence de gènes codant des protéines hypothétiques en amont des gènes codant des métallo-enzymes et transporteurs suggère que ces systèmes contrôlent l'expression de gènes dont les produits, après modification, seraient destinés à l'export et à un devenir extracellulaire. Le locus *shp/rgg1358* de *S. thermophilus* LMD-9 a, par ailleurs, déjà été décrit comme régulant le gène *ster_1357*, dont le produit – un peptide cyclique appelé Pep1357C – est présent dans les surnageants de culture (Ibrahim *et al.*, 2007a). En outre, la synthèse d'une nouvelle classe de phéromones dont la maturation requiert une enzyme CAAX a également été démontrée chez les champignons (Schmoll *et al.*, 2010).

Alors que des gènes cibles proximaux et/ou distaux ont été détectés pour la majorité des systèmes SHP/Rgg (tous groupes confondus), trois systèmes restent quant à eux sans cible associée. Pour deux d'entre eux (*shp/spg_1268* de *S. pneumoniae* G54 et *shp/sez_0050* de *S. equi* subsp. *zooepidemicus* MGCS10565), ceci coïncide avec une absence d'identification d'un motif conservé (références Rgg_11 et Rgg_12). De plus, les phéromones SHP associées à ces systèmes présentent une particularité au niveau de leur séquence au C-terminale de type « GLYQ » ou « GLCQ » (Tableau 24). En revanche, le troisième locus, *shp/gbs1555* de *S. agalactiae* (et ses paralogues), ne présente pas de particularité au niveau de la séquence de la phéromone SHP. Aussi, le développement d'approches globales semble être nécessaire à l'identification des régulons de ces différents systèmes. Une

Tableau 28 – Cibles putatives sélectionnées pour validation *in vitro*.

	Protéine Rgg	Cible P	Cible D	Motif NC	Motif C
***S. thermophilus* LMD-9**	Rgg1358		*shp_1530*	×	
			amiE		×
***S. thermophilus* LMG18311**	Rgg0182		*stu1251*	×	
			stu0665		×
***S. mutans* UA159**	Smu.1509	*smu_1508*		×	
			smu_1031	×	
			smu_1711		×
			smu_1879		×
***S. agalactiae* NEM316**	Gbs1555		*gbs1125*		×

Cible P, cible proximale ; Cible D, cible distale ; Motif NC, motif dans région non codante ; Motif C, motif dans région codante.

analyse protéomique est en cours au laboratoire concernant le locus *shp/gbs1555* de *S. agalactiae* NEM316.

Enfin, de nombreux motifs de fixation des protéines Rgg, associées à une phéromone SHP, ont été détectés au niveau de régions codantes. Bien que ce phénomène soit de plus en plus décrit dans la littérature par l'utilisation d'approches de type ChIP-Chip et Chip-Seq, le rôle de la fixation de régulateurs transcriptionnels au sein de régions codantes n'est pas encore bien compris et diverses spéculations sont proposées concernant ce phénomène (MacQuarrie *et al.*, 2011). Néanmoins, au regard de l'analyse ChIP-Chip réalisée sur le Rgg stand-alone « RopB » de *S. pyogenes*, qui décrit la fixation de ce régulateur au niveau de 65 régions incluant pour plus de la moitié des régions codantes (>55%) (Anbalagan *et al.*, 2011), le fort pourcentage (variable suivant les systèmes SHP/Rgg) dans cette étude de motifs identifiés dans des régions codantes ne peut être considéré comme négligeable.

Actuellement, la validation *in vitro* de cibles putatives détectées dans cette étude est en cours au laboratoire. Cette validation, concernant à la fois l'étude des motifs au sein des régions codantes et non-codantes, devrait permettre de confirmer certaines hypothèses émises dans cette discussion concernant (i) les systèmes *shp/rgg1358* et *shp/ster_1530* chez *S. thermophilus*, (ii) la fixation des protéines Rgg dans les régions codantes et (iii) la régulation de gènes codant des transposases, un système PTS et une protéine de ségrégation (Tableau 28).

Discussion générale et perspectives

Mon travail de thèse a, d'une part, permis d'identifier une nouvelle famille de régulateurs associés à des peptides hydrophobes chez les streptocoques, et, d'autre part, a ouvert de nombreuses autres perspectives de travail dans le domaine du quorum-sensing (QS) et des petits gènes codants chez les bactéries. Ce sont ces deux aspects que je discuterai ici en abordant sept points qui me paraissent importants.

1. Détection de petits gènes codant potentiellement des peptides et identification de familles de petits gènes dans les génomes bactériens.

L'annotation des petits gènes (définis généralement par une taille inférieure à 50 codons) codant potentiellement des peptides a été largement sous estimée et réalisée de manière inégale lors de l'annotation classique des génomes. Ces petites protéines ou peptides font cependant l'objet d'un intérêt croissant du fait de leurs rôles potentiels (Boekhorst *et al.*, 2011 ; outil BactSHOW, (Ibrahim *et al.*, 2007b) ; outil BAGEL, (De Jong *et al.*, 2006). En effet, les peptides peuvent (i) être toxiques et être impliqués dans des mécanismes de mort cellulaire, (ii) jouer un rôle d'adaptateur ou de régulateur en favorisant la dégradation, la stabilisation ou l'activation de certaines protéines et, (iii) être impliqués dans des phénomènes de signalisation et par conséquent dans la régulation de gènes *via* les phénomènes de type QS (Alix and Blanc-Potard, 2009; Zuber, 2001). Chez *E. coli*, il a par ailleurs été observé que de nombreuses petites protéines s'accumulaient en réponse à des conditions spécifiques ou en réponse à un stress (Hemm *et al.*, 2010).

Si les protéines ribosomales et les bactériocines sont connues depuis longtemps, les recherches systématiques récentes de ces phases ouvertes de lectures courtes (ORFs pour Open Reading Frames) ont permis d'identifier des peptides hydrophobes régulateurs membranaires chez les bactéries à Gram négatif (Alix and Blanc-Potard, 2009) et des phéromones impliquées dans le QS chez les bactéries à Gram positif (Ibrahim *et al.*, 2007b). Mieux appréhender leur nombre et leur rôle chez les bactéries, rend le développement d'approches permettant leur détection de manière la plus exhaustive possible nécessaire.

Dans ce travail de thèse, la mise à jour (par rapport au travail de Ibrahim *et al.*, 2007b) des données concernant les systèmes SHP/Rgg dans les nouveaux génomes de streptocoques séquencés, a généré un important jeu de données de petits gènes concernant plusieurs genres de bactéries à Gram positif (ordre des *Lactobacillales* et famille des *Listeriaceae*). En effet, afin de définir si les systèmes SHP/Rgg étaient bien spécifiques des streptocoques, une analyse de type Psi-Blast a tout d'abord été entreprise afin de déterminer

un taxon Rgg, au sein duquel, la recherche de petites ORFs a été effectuée. Bien que l'analyse des données ait permis de confirmer l'appartenance des systèmes SHP/Rgg aux streptocoques, un nombre important d'autres petites ORFs associées à des gènes codant des régulateurs transcriptionnels Rgg-like a été détecté lors de cette analyse. Nous avons cherché à partir de ce jeu de données, jusqu'alors partie invisible de l'iceberg, à définir de nouvelles familles de peptides phéromones. La première contrainte consiste à prévenir, lors de l'analyse, la prédiction de faux-positifs, c'est-à-dire de petites ORFs ne codant pas de protéines. Au sein des génomes bactériens à bas GC%, la fréquence élevée des codons STOP prévient considérablement ce phénomène (Boekhorst *et al.*, 2011). Néanmoins, alors que le risque de détection des faux-positifs reste limité pour des gènes codant des protéines supérieures à 50 acides aminés, il augmente lorsque la taille des gènes recherchés diminue. Ces prédictions doivent donc être considérées avec prudence pour des petites ORFs dont la taille est comprise entre ou inférieure à 20/25 acides aminés, comme c'est le cas pour les phéromones SHP, d'autres peptides identifiés (résultats chapitre 1) et données non présentées. La question de la taille minimale raisonnable des ORFs à prendre en compte est donc posée. Pour limiter (au moins en partie) le risque de détection de faux-positifs, un deuxième critère à prendre à compte concerne la conservation d'une petite ORF au sein de plusieurs espèces. Aussi, une famille de petites ORFs détectées au sein de différentes espèces présente, *a posteriori*, une probabilité plus grande d'être codante. A ces critères généraux d'analyse s'ajoute la difficulté liée à des erreurs d'annotation des démarrages de traduction, qui au vu de certaines familles de peptides (mitis1 et Listeria) identifiées au cours de ce travail alourdit considérablement la recherche et l'identification des familles dans la base de données générées. Aussi, l'amélioration des algorithmes semble nécessaire dans le but de pouvoir rechercher de manière automatisée et plus fiable des ORFs sous certains critères établis au préalable. Pour autant, la validation expérimentale des prédictions n'en reste pas moins nécessaire. L'expression des petits gènes n'est pas facile *via* des approches de puces classiques ou de PCR quantitatives. Les peptides et petites protéines sont également plus difficiles à détecter et à identifier que les protéines de taille plus importante. Des approches globales telles que le « Tiling array » constituent actuellement le moyen le plus efficace pour la validation des prédictions. L'utilisation du « Tiling array » pourrait s'avérer intéressante dans le cadre des systèmes SHP/Rgg. En effet, le rôle physiologique de ces systèmes n'étant pas connu, l'expression de ces systèmes sous différentes conditions pourrait alors renseigner et orienter les recherches vers des conditions optimales d'expression et par conséquent de caractérisation de ces différents systèmes. Néanmoins, des approches ciblées utilisant des techniques moins couteuses, telles que des fusions transcriptionnelles pour la mise en évidence de l'expression de petits gènes ou l'HPLC pour la recherche des peptides codés par les petites ORFs dans les surnageants de cultures,

peuvent également confirmer, à plus petite échelle, la prédiction de petits gènes codant des peptides. Nous avons utilisé avec succès ces approches dans le cadre de l'étude du locus *shp/rgg1358* de *S. thermophilus* LMD-9.

2. Mécanistique des systèmes de QS « SHP/Rgg » et comparaison avec d'autres systèmes de QS

2.1. Le peptide

Nous avons montré que les systèmes SHP/Rgg répondent à un mécanisme de QS. La régulation de tels phénomènes, bien que dépendante d'un ratio molécules de signalisation / biomasse bactérienne, est également décrite comme répondant à divers facteurs pouvant être intrinsèques ou extrinsèques. En effet, la production des molécules de signalisation est généralement décrite comme auto-induite et dépendante d'une boucle de rétrocontrôle positive, mettant en jeu l'effecteur central (le régulateur transcriptionnel) assurant la modulation de l'expression des gènes cibles (Fontaine *et al.*, 2010a; Mashburn-Warren *et al.*, 2010; Novick, 2003; Pestova *et al.*, 1996a). Nous avons également montré l'existence d'un tel phénomène dans la mécanistique des systèmes SHP/Rgg. Par ailleurs, certains critères extrinsèques ont également été décrits comme contrôlant le déclenchement d'un mécanisme de QS. Tel est le cas, à titre d'exemple, pour le mécanisme de la compétence chez *S. thermophilus* qui ne se déclenche que dans un milieu dépourvu de peptide (MCD) (Gardan *et al.*, 2009). De manière similaire, l'absence d'activité de la fusion $P_{shp1358}$*::luxAB* en milieu M17lac dans un contexte sauvage (données non publiées), indique également que pour le système SHP/Rgg1358 de *S. thermophilus* LMD-9, la présence de peptides inhibe le phénomène. Aussi, une compétition avec les peptides nutritionnels, réimportés par le transporteur d'oligopeptides Ami, semble avoir lieu.

Afin d'être détectées à partir du milieu extracellulaire, les molécules de signalisation y sont au préalable exportées. Aussi, la présence de signaux d'exportations dans la partie N-terminale des séquences peptidiques des précurseurs a été décrite dans de nombreux cas (Agaisse *et al.*, 1999; Håvarstein *et al.*, 1995; Lazazzera, 2001; Zhang *et al.*, 2004). Bien que des transporteurs aient été identifiés comme nécessaires à l'export de peptides phéromones associés à des systèmes à deux composants (Hui *et al.*, 1995; Novick and Geisinger, 2008), pour de nombreux autres, le mécanisme par lequel ils sont exportés reste inconnu. C'est le cas des peptides participant aux systèmes de QS impliquant un effecteur de la famille RNPP. Aussi, bien que l'export de PapR (*B. cereus*) (Agaisse *et al.*, 1999) et des peptides Phr (*B. subtilis*) (Lazazzera, 2001) *via* la voie de sécrétion Sec ait été proposé, l'essentialité de cette voie dans la survie des bactéries rend difficile la vérification de cette hypothèse.

Celle-ci est cependant très probable dans notre cas, puisque les séquences des précurseurs des phéromones SHP ressemblent, dans leur intégralité, à une séquence signale dépendante de la voie de sécrétion Sec (Hutchings *et al.*, 2009; Palmer *et al.*, 2005). De plus, la présence d'une glycine ou d'un motif double glycine en C-terminal, décrit comme site spécifique de clivage des séquences signales de certaines protéines ou lors de la maturation de phéromones (Håvarstein *et al.*, 1995; Michiels *et al.*, 2001) conforte l'hypothèse que les phéromones SHP (sous entendu leur forme mature) sont partie intégrante d'une séquence signale. Ce phénomène peut également être proposé pour les phéromones ComS chez les streptocoques, présentant généralement une glycine en C-terminal (Mashburn-Warren *et al.*, 2010). Aussi, il est possible de spéculer sur l'évolution des gènes codant ces peptides. En effet, ces derniers pourraient avoir évolué à partir de formes plus longues de manière à ne conserver que la partie strictement essentielle à l'activité du mécanisme de QS. Une observation faite chez *E. faecalis* conforte cette hypothèse. Chez cette bactérie, les sex-phéromones, impliquées dans les mécanismes de conjugaison et dont les séquences sont similaires à celles des phéromones SHP, sont des fragments de peptides signaux de précurseurs de lipoprotéines (Clewell *et al.*, 2000). Enfin, si la voie de sécrétion Sec n'était pas responsable de l'export des phéromones SHP, il serait possible d'envisager que, vu leur forte hydrophobicité, elles soient « passivement » transloquées à travers la membrane cytoplasmique.

La localisation des phéromones hors des bactéries n'est pas toujours très claire. Dans quelques cas, comme celui de cCF10 chez *E. faecalis* (Buttaro *et al.*, 2000), la sex-phéromone est décrite comme associée aux parois mais la plupart du temps, elles ont été isolées de surnageants de culture. Dans le cas des SHP, au moins une partie des phéromones est présente sous forme soluble dans le milieu extracellulaire mais on ne peut exclure, en particulier à cause de leur forte hydrophobicité, qu'une partie reste localisée au niveau de la paroi.

Au cours de leur export, les molécules de signalisation subissent une étape de maturation (impliquant une ou plusieurs coupures par des enzymes protéolytiques spécifiques ou pas et parfois, une modification post-traductionnelle du peptide) permettant *in fine* la libération de la forme mature (ou dite « active ») de la phéromone dans le milieu extracellulaire. La similarité des phéromones SHP et des anti-phéromones chez *E. faecalis*, nous a conduits à étudier le rôle, dans la maturation des phéromones SHP, de la protéase membranaire Eep impliquée dans la maturation des anti-phéromones d'entérocoques. Aussi, l'absence d'activation du promoteur $P_{shp1358}$ après ajout d'un surnageant de culture du mutant Δ*eep* contenant un plasmide surproducteur de la phéromone SHP1358, nous a permis de démontrer un rôle direct ou indirect de cette protéase dans la maturation d'une phéromone

SHP appartenant au groupe II. Au cours des travaux de Chang *et al.* (2011), la complémentation fonctionnelle d'un mutant Δ*eep* chez *S. pyogenes* a alors permis d'élargir ce phénomène à une phéromone SHP du groupe I. Aussi, afin de montrer de manière directe le rôle de la protéase membranaire Eep dans la maturation des phéromones SHP (et éventuellement d'autres protéases), le développement d'outils tels que l'utilisation d'un système acellulaire de type liposome peut-être entrepris. Par ailleurs, l'utilisation d'un système hétérologue pour l'étude de la fonctionnalité d'autres systèmes SHP/Rgg, suggère la maturation des phéromones SHP, de *S. agalactiae* NEM316 et *S. mutans* UA159, par la protéase membranaire Eep présente chez *S. thermophilus* LMD-9. Les protéases de *S. agalactiae* et *S. mutans* présentant respectivement 68 et 67% d'identité avec celle de *S. thermophilus* LMD-9, un lien évolutif entre les protéines Eep et les séquences des phéromones SHP semblent alors avoir été conservé. Si la protéase Eep de *S. thermophilus* est effectivement directement impliquée dans cette coupure, elle serait spécifique de résidus acides, alors que chez *Enterococcus*, Eep reconnait plutôt des acides aminés hydrophobes. Aussi, ces enzymes présenteraient une spécificité différente.

A nos connaissances, seulement des phéromones associées à des systèmes à deux composants ont été décrites comme ayant subi des modifications post-traductionnelles (Ji *et al.*, 1995; Ji *et al.*, 1997). La recherche et l'identification de plusieurs formes matures de phéromones SHP, à partir de surnageants de culture, permet actuellement de définir les phéromones SHP comme étant des peptides linéaires, issus d'un clivage côté C-terminal au niveau d'un résidu acide conservé (aspartate ou glutamate) au sein de leur séquence.

D'une manière générale, très peu de données concernant la stabilité des peptides phéromones (linéaires ou modifiés), à l'extérieur ou l'intérieur de la bactérie, sont disponibles dans la littérature. A nos connaissances, seulement un rôle de prévention contre la protéolyse est attribué aux modifications structurales de certaines phéromones (Chan *et al.*, 2004). Lors de la recherche des formes actives des phéromones SHP, certaines formes plus courtes dégradées à partir de l'extrémité N-terminale ont également pu être détectées par spectrométrie de masse (A. Guillot, communication personnelle), ce qui suggère que ces phéromones sont des substrats pour des enzymes protéolytiques extracellulaires ou membranaires. Chez *S. thermophilus* LMD-9, les candidates extracellulaires sont assez peu nombreuses : la protéase de paroi PrtS, la protéase de ménage HtrA et les protéases membranaires très mal connues. Cette protéolyse extracellulaire joue sans aucun doute un rôle important dans la régulation du QS. Bien qu'un rôle biologique puisse être attribué à différentes formes de phéromones par exemple pour les peptides PapR5 et 7 chez *B. cereus* (Bouillaut *et al.*, 2008; Slamti and Lereclus, 2005) ou encore ComS chez *S. thermophilus* (Fontaine *et al.*, 2010a) ; Chang *et al.* (2011), ont démontré l'importance, chez *S. pyogenes*,

de la taille d'une forme active de la phéromone SHP3 et du rôle du résidu acide conservé sur le déclenchement du mécanisme de QS. L'analyse *in silico* nous a permis de mettre en évidence des séquences atypiques de certaines phéromones SHP. En effet, certaines présentent un résidu cystéine (à la place de l'aspartate ou du glutamate conservé) ou une extrémité C-terminale variable. Il serait alors intéressant de définir si ces résidus jouent également un rôle sur l'expression des systèmes SHP/Rgg auxquels ils sont apparentés.

Pour permettre la détection des molécules de signalisation *via* leur senseur membranaire ou leur(s) protéine(s) affine(s) apparentés, les molécules de signalisation doivent atteindre une concentration critique dans le milieu extracellulaire. A l'heure actuelle, très peu de données sont disponibles dans la littérature concernant les concentrations requises au déclenchement des mécanismes de QS, en conditions de laboratoire (Declerck *et al.*, 2007; Dunny and Leonard, 1997). Chez *S. pneumoniae*, la concentration de la phéromone CSP nécessaire au déclenchement de la compétence a été évaluée à 10 ng/ml (Håvarstein *et al.*, 1995). Aussi, des travaux sont en cours afin de définir, sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9, la concentration en phéromone SHP1358 requise au déclenchement du mécanisme. Pour ceci, un peptide lourd a été synthétisé et permettra de doser, par spectrométrie de masse, la quantité de phéromone SHP1358 dans les surnageants de culture au cours de la croissance de *S. thermophilus* LMD-9. Néanmoins, bien que les phéromones SHP soient présentes dans les surnageants de culture, on ne peut exclure qu'une partie non-négligeable reste associée à la membrane des bactéries. En effet, ce phénomène a été observé pour la phéromone ComS de *S. thermophilus* LMD-9 (R. Gardan, communication personnelle).

Dans nos travaux, la détection de la phéromone SHP1358 a été montrée comme effective par les deux protéines affines associées au transporteur d'oligopeptides AmiCDEF. Bien que la protéine affine AmiA3 semble jouer un rôle majoritaire dans la détection de cette phéromone, nous n'en avons pas encore identifié la raison : soit la protéine AmiA3 présente une affinité plus forte pour les SHP, soit la protéine AmiA3 est plus abondante que la protéine AmiA1. Aussi, la surproduction de ces deux protéines dans un mutant nul des protéines affines a été entreprise au laboratoire afin d'étudier le rôle dans la spécificité d'import des phéromones SHP mais aussi ComS, également majoritairement importée dans la cellule par la protéine affine AmiA3 (Gardan *et al.*, 2009), chez *S. thermophilus* LMD-9.

1.2. Le régulateur

L'interaction d'un couple phéromone SHP / régulateur transcriptionnel Rgg a été démontrée par SPR lors de l'étude du système SHP/Rgg1358 de *S. thermophilus* LMD-9.

L'élargissement de la fonctionnalité du mécanisme de QS décrit sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9 à d'autres streptocoques, incluant des espèces pathogènes, nous amène actuellement à l'étude de l'existence de phénomène de cross-activation ou cross-inhibition de ces différents systèmes, à l'échelle inter-espèces ou inter-souches. En effet, la similarité des séquences des phéromones SHP peut être à la base de l'existence de phérotypes, comme déjà décrits pour plusieurs systèmes de QS (Bouillaut *et al.*, 2008; Slamti and Lereclus, 2005; Wright *et al.*, 2005). Aussi, la mesure de l'affinité de couples SHP/Rgg légitimes et illégitimes sont prévues, par SPR, et devrait permettre de définir ou non l'existence de phérotypes au sein de cette famille de phéromones.

La fixation du régulateur Rgg à l'ADN mérite d'être discutée. En effet, nous avons montré que la liaison de la protéine Rgg1358 à l'ADN n'est pas dépendante de la phéromone SHP, bien que la présence des deux partenaires soit nécessaire à l'activation des gènes cibles connus du système. Cette constatation diffère de ce qui a été observé lors de l'étude du système PapR/PlcR chez *B. cereus* par Slamti et Lereclus (2005) qui ont montré que la présence de la phéromone PapR est essentielle à la liaison du régulateur PlcR à l'ADN et par conséquent à l'activation du régulon. *A contrario*, les études portant sur le plasmide conjugatif pCF10 de *E. faecalis* ont montré que le régulateur PrgX est toujours lié à l'ADN et que la liaison de la sex-phéromone à PrgX traduit une levée de la répression par ce régulateur en provoquant un changement d'oligomérisation de ce dernier (Kozlowicz *et al.*, 2006b; Shi *et al.*, 2005). Bien que nous ne remettions pas en cause les résultats que nous avons publiés, nous ne pouvons exclure l'influence de la queue 6-His (localisée en N-terminale de la séquence de la protéine Rgg) sur la liaison de la protéine Rgg1358 à l'ADN ou, alors, qu'une seconde maturation de la phéromone SHP dans le cytoplasme soit non-caractérisée. Aussi, nous émettons l'hypothèse que le contrôle des gènes cibles des systèmes SHP/Rgg soit issu d'une régulation à l'interface de ces deux mécanismes. En effet, nous avons montré que la protéine Rgg1358 est un répresseur de son propre gène. Les analyses de retard sur gel confortent cette hypothèse et par conséquent le fait que Rgg1358 soit lié à l'ADN en absence de sa phéromone SHP apparentée. Aussi, nous proposons que la présence de la phéromone SHP induise un changement conformationnel de la protéine régulatrice Rgg1358 et/ou la modification de son état d'activation. Si tel est le cas, ce mode d'action phéromone / protéine régulatrice serait inédit et différent de ceux décrits dans la littérature pour les systèmes de QS PapR/PlcR et cCF10-iCF10/PrgX appartenant à la famille RNPP. Aussi, une étude cristallographique est actuellement entreprise afin de définir la structure de la protéine Rgg1358 complexée à l'ADN et liant ou non la phéromone SHP.

Une particularité des systèmes SHP/Rgg réside en leur possible accumulation au sein des souches de streptocoques. Dans ce travail, l'identification de motifs putatifs de fixation des protéines Rgg (associées à des phéromones SHP) à l'ADN, spécifique de chaque

groupe de systèmes SHP/Rgg, souligne l'évolution de ces systèmes entre ces différents groupes mais également à l'échelle intra-groupe. L'étude de ces motifs et l'identification des nucléotides impliqués dans la liaison des protéines Rgg à l'ADN ou dans l'affinité de liaison du régulateur à l'ADN permettrait certainement, à l'avenir, de mieux comprendre ces interactions et le rôle des phéromones SHP dans l'activation des régulons putatifs identifiés.

3. Les systèmes SHP/Rgg : transfert horizontal, évolution et fonction physiologique.

L'analyse des environnements génétiques des systèmes SHP/Rgg chez les streptocoques a révélé la présence de nombreuses transposases dans l'environnement proche de ces systèmes, en particulier pour les systèmes appartenant au groupe I. En effet, seulement deux transposases sont présentes dans l'environnement proche du système SHP/Rgg1358 chez *S. thermophilus* LMD-9, appartenant au groupe II. Aussi, il semblerait que les systèmes SHP/Rgg aient été acquis par transfert horizontal. De plus, l'étude de Eng *et al.* (2011), souligne le transfert horizontal de nombreux îlots génomiques chez *S. thermophilus* au sein desquels des régulateurs transcriptionnels de la famille Rgg ont été identifiés.

De manière intéressante, il semblerait, au vu des systèmes SHP/Rgg orthologues identifiés, que ces échanges aient pu se produire entre différentes espèces plus ou moins proches phylogénétiquement. *S. pyogenes*, *S. dysgalactiae* et *S. agalactiae* appartenant au groupe pyogenic, possèdent un système SHP/Rgg orthologue présentant 77% d'identité mais n'ayant pas évolué au sein des différentes espèces. Ces différents systèmes présentent 99% (*S. pyogenes*) ou 100% (*S. agalactiae*) d'identité à l'échelle de l'espèce. Néanmoins, *S. mutans* et *S. equi*, n'appartenant pas au même groupe phylogénétique et ne partageant pas le même habitat naturel, présentent également un système orthologue (67% d'identité) dans leur génome. De la même manière que pour le système orthologue précédemment décrit, une forte identité de ce système entre les différentes souches séquencées de *S. mutans* peut être observée (99%). Aussi, la stricte ou forte conservation de ces systèmes au sein d'une espèce suggère l'acquisition et le transfert récent de ces systèmes entre les souches d'une espèce apparentée.

Les régulateurs Rgg sont similaires aux régulateurs de prophages Cro/CI (Anbalagan *et al.*, 2011; McDonnell *et al.*, 1994). Ces derniers pourraient donc dériver d'un ancêtre commun ayant permis le transfert de ces systèmes. Bien que *S. thermophilus* présente une forte perte de gènes, c'est en revanche, l'espèce de streptocoque qui a accumulé le plus

grand nombre de systèmes SHP/Rgg dans son génome. Aussi, pourquoi *S. thermophilus* possède-t-il un nombre supérieur de systèmes SHP/Rgg en regard des autres streptocoques ? Et, pourquoi, une telle variabilité du nombre de ces systèmes est visible au niveau inter-espèces et parfois inter-souches ? Le fait que le rôle de ces différents systèmes ne soit pas encore connu, ne permet malheureusement pas de répondre, à l'heure actuelle, à ces interrogations.

4. Les systèmes « SHP/Rgg » versus « XIP/ComR ».

Au cours de mes travaux de thèse, le rôle des protéines ComR dans un phénomène de régulation impliquant un peptide et un régulateur transcriptionnel contrôlant le déclenchement de la compétence pour la transformation naturelle, a été démontré (Fontaine *et al.*, 2010a; Mashburn-Warren *et al.*, 2010). Les travaux de Fontaine *et al.* (2010) ont décrits les protéines ComR comme appartenant à la famille des régulateurs transcriptionnels Rgg, terme ensuite repris lors des travaux de Mashburn-Warren. Cette classification a alors été à la base d'avis divergents au sein de la communauté scientifique : d'un côté les « pro-XIP/ComR », définissant les protéines ComR comme étant des protéines Rgg ; et de l'autre côté, les « pro-SHP/Rgg », relatant que les protéines ComR et Rgg sont distinctes et par conséquent, à l'origine de deux systèmes de régulation différents.

A l'heure actuelle, l'arbre phylogénétique que nous avons réalisé (article 1) relate la proximité de ces deux systèmes par leur appartenance à un même genre bactérien, celui des streptocoques mais montre également qu'il s'agit bien de deux systèmes distincts. Ces deux systèmes SHP/Rgg et XIP/ComR fonctionnent tous deux sur la base de phéromones exportées, réimportées par AmiA3 chez *S. thermophilus* LMD-9 pour ensuite interagir avec un régulateur transcriptionnel. Dans mon travail de thèse, démontrer l'originalité des systèmes SHP/Rgg par rapport aux systèmes XIP/ComR (encore appelé ComRS) s'est alors avéré être un paramètre nécessaire à leur valorisation. Les différences suivantes entre les deux systèmes ont été identifiées :

- *Quorum-sensing versus Timing-sensing*

Bien que certaines étapes restent à valider pour les systèmes XIP/ComR (interaction phéromone / régulateur transcriptionnel et fixation du régulateur à l'ADN), les systèmes SHP/Rgg et XIP/ComR fonctionnent de manière différente. En effet, le système SHP/Rgg est clairement apparenté à du QS. Il s'allume en réponse à une certaine densité cellulaire, comme nous l'avons montré en utilisant des cultures ensemencées à différents niveaux qui fonctionnaient toutes à la même densité cellulaire. Le système XIP/ComR ne répond pas à la densité cellulaire mais s'apparente plutôt à ce que j'appellerai du « Timing Sensing ».

Quelque soit la densité cellulaire, le système s'allume après un temps défini ce qui correspond plus à un état physiologique des bactéries.

- *Organisation génétique des loci*

Les deux systèmes se distinguent par l'organisation génétique des gènes codant la phéromone et le régulateur apparenté. En effet, alors que le gène *comS* (codant la phéromone) est localisé en amont de son régulateur ComR et transcrit dans le même sens ; les gènes *shp* – bien que pouvant être localisés en amont ou en aval du gène *rgg* – sont toujours transcrits de manière divergente à ce dernier.

- *Structure des régulateurs et fixation à l'ADN*

Bien que l'annotation des génomes soit à prendre avec précaution, les protéines, actuellement désignées ComR, ne sont pas présentées comme possédant un domaine C-terminal Rgg-Cterm (TIGR01716, Genbank) identifié pour les protéines Rgg « stand-alone » ou associées aux phéromones SHP. Le fait que cette observation soit applicable à l'ensemble des protéines ComR – décrites dans la littérature (Fontaine *et al.*, 2010a; Mashburn-Warren *et al.*, 2010) – renforce l'idée que ces régulateurs bien que similaires présentent clairement des spécificités. De plus, leur annotation dans Genbank fait référence aux gènes *plcR*-like ce qui n'est pas le cas des Rgg. L'étude des régions promotrices des gènes codant les phéromones de ces systèmes a révélé l'existence de différents motifs de fixation 'putatifs' des protéines régulatrices à l'ADN. En effet, alors qu'une séquence palindromique a été identifiée dans la région promotrice des gènes *comS*, une séquence ADN conservée (ne présentant pas de motif particulier) a été identifiée au sein de chaque groupe de systèmes SHP/Rgg.

- *Séquence des phéromones*

La séquence des phéromones associées à ces systèmes présente également des spécificités non négligeables. Premièrement alors que la taille des peptides ComS est variable, celle des peptides SHP est conservée. Deuxièmement, les formes matures des phéromones de ces systèmes ont été caractérisées par la présence de résidus conservés : un motif double tryptophane pour les phéromones XIP et un acide aminé acide (aspartate ou glutamate) pour les phéromones SHP. Enfin, et comme précédemment décrit dans cette discussion, le comportement des phéromones dans le milieu extracellulaire (en MCD) est différent : les phéromones XIP n'étant pas libérée dans le milieu extracellulaire (R. Gardan, communication personnelle).

Aussi, au vu de ces premiers éléments justifiant l'étude parallèle de ces deux systèmes, il semblerait que les systèmes SHP/Rgg et XIP/ComR ne présentent finalement en commun, que le fait d'impliquer des peptides et des régulateurs transcriptionnels comme précédemment décrit au sein de la famille RNPP.

5. Les protéines régulatrices Rgg appartiennent-elles à la famille RNPP ?

De manière objective, aucune donnée ne permet actuellement de démontrer l'appartenance des protéines ComR à la famille Rgg et plus généralement de définir l'appartenance des protéines ComR et Rgg à la famille RNPP. Les effecteurs intracellulaires composants la famille RNPP (Rap, NprR, PlcR et PrgX) constituent une famille comprenant des domaines TPR (impliqués dans l'interaction protéine / protéine ou protéine / peptide) et un domaine hélice-tour-hélice de liaison à l'ADN appartenant à la famille XRE (Declerck *et al.*, 2007). Bien que la prédiction de la structure des protéines Rgg identifie certaines hélices alpha alignées à celles composant le domaine TPR de liaison de la molécule de signalisation des régulateurs PlcR et PrgX (Carroll *et al.*, 2011; Chang *et al.*, 2011; Declerck *et al.*, 2007; Shi *et al.*, 2005), aucun domaine TPR n'a été (à proprement parler) détecté lors de l'utilisation de certains logiciels (TPR-pred, Karpenahalli *et al.*, 2007) spécialisés dans la prédiction de ces domaines. Aussi, pour (i) établir le lien évolutif des protéines Rgg, ComR et RNPP et (ii) définir l'appartenance de ces deux régulateurs à la famille RNPP, une approche structurale est actuellement entreprise au laboratoire en collaboration avec une équipe de cristallographes. Par ailleurs, si l'appartenance des protéines Rgg à la famille RNPP est révélée, la dénomination RNPPR (Rap, NprR, PlcR, PrgX, Rgg) sera probablement plus adaptée (car respectant l'évolution des effecteurs de cette famille) que celle proposée par Federle et ses collaborateurs : RRNPP.

6. Les protéines régulatrices Rgg associées aux phéromones : une sous-famille des protéines Rgg « stand-alones » ?

Une question sous-jacente à l'étude des régulateurs transcriptionnels Rgg concerne la capacité de ces protéines à s'associer à différents signaux ou métabolites (Loughman and Caparon, 2006, 2007a); article 1). Aussi, les protéines Rgg associées à des phéromones SHP ne formeraient-elles pas une sous-famille de régulateurs transcriptionnels associant une certaine catégorie de signal ? La comparaison des structures d'un Rgg dit « stand-alone » c'est-à-dire non associé génétiquement à un petit peptide, à celle d'un Rgg associant une phéromone SHP permettrait de définir la conformation des poches (putatives) de liaison du substrat. Aussi, il pourrait être envisageable que les effecteurs de la famille RNPP et apparentés (Rgg et ComR) aient évolué au sein des différents genres bactériens (*Bacillus*,

Enterococcus et *Streptococcus*) et que au sein de ces différents genres, chaque régulateur associé ait évolué dans le but de répondre à une diversité de signaux.

7. Rôle des systèmes SHP/Rgg dans l'environnement naturel des streptocoques.

A contrario des molécules de signalisation AgrD, CSP et XIP qui sont décrites comme impliquées dans le contrôle d'une seule fonction physiologique (la virulence ou la compétence), l'accumulation des systèmes SHP/Rgg suggère la régulation d'un nombre varié de fonctions physiologiques. L'analyse des environnements génétiques des systèmes SHP/Rgg appartenant aux groupes II et III a suggéré que ces systèmes contrôlent l'expression de petits gènes codant des protéines / peptides destinés à un devenir extracellulaire et pouvant jouer le rôle de bactériocine (Diep *et al.*, 1996; Pei and Grishin, 2001; Rea *et al.*, 2010) ou encore de phéromones (Schmoll *et al.*, 2010). Au laboratoire, l'étude de la fonction d'un de ces peptides, le peptide cyclique Pep1357C de *S. thermophilus* LMD-9, est entreprise. Bien que ce peptide cyclique ne ressemble pas à des bactériocines connues, des tests antimicrobiens (portant sur l'activité bactéricide ou bactériostatique de ce peptide) sur différents micro-organismes, en particulier utilisés ou indésirables au cours des procédés de fabrication dans lesquels *S. thermophilus* participe, ont été réalisés. Néanmoins, aucun effet sur la croissance des micro-organismes testés n'a été détecté. Des tests à plus grande échelle sur un nombre plus important de micro-organismes seraient sans doute nécessaires pour écarter complètement cette hypothèse. L'étude du rôle du peptide Pep1357C dans la formation de biofilm n'a, par ailleurs, révélé aucune implication du peptide dans cette fonction (R. Briandet, R. Gardan, communication personnelle). Des analyses protéomiques d'un mutant du gène *ster_1357* (codant Pep1357C) sont en cours au laboratoire pour identifier d'autres pistes.

Le rôle des systèmes SHP/Rgg se pose de manière différente selon que le streptocoque est pathogène ou pas. Pour les pathogènes, les cibles évidentes sont les facteurs de virulence. La grande majorité des travaux menés dans le domaine du QS est d'ailleurs directement liée à la régulation de l'expression de facteurs de virulence. La prédiction des régulons Rgg a, par ailleurs, permis d'identifier des cibles putatives impliquées dans la virulence chez les streptocoques. Au laboratoire, nous explorons actuellement cette piste chez *S. agalactiae.* Chez *S. thermophilus* qui a accumulé ces systèmes et dont la seule niche écologique connue est le lait et les produits laitiers, la question est plus ouverte. Nous observons que ces systèmes ne se sont pas répandus chez d'autres espèces de bactéries lactiques comme *Lactococcus lactis* ou *Lactobacillus bulgaricus* qui cohabitent avec *S. thermophilus* dans les produits laitiers. Pour identifier le rôle de ces systèmes, il faudra

donc sans doute croiser les spécificités métaboliques de *S. thermophilus* et son environnement.

Les conditions dans lesquelles les systèmes SHP/Rgg sont fonctionnels c'est-à-dire dans des cultures planctoniques en MCD restent bien sûr très artificielles mais ont permis leur étude mécanistique. Aussi, qu'en est-il des communications qui s'établissent au sein d'écosystèmes (biofilms, micro-colonies) plus complexes, où les concentrations en bactéries et phéromones sont différentes ? Est-ce que les systèmes SHP/Rgg sont exprimés dans de telles conditions ? Le développement d'outils, au laboratoire, permettant de suivre l'expression de différents systèmes SHP/Rgg par fluorescence pourrait alors être entrepris. Cette technique permettrait, par ailleurs, d'avancer sur l'expression de différents systèmes SHP/Rgg dans le temps. Enfin, quel est le rôle des phéromones SHP et des systèmes SHP/Rgg au sein de tels environnements ? Jouent-ils une action inhibitrice sur d'autres systèmes de QS dans d'autres genres bactériens ? Ces questions ouvertes constituent actuellement l'enjeu de nombreuses perspectives auxquelles la communication scientifique « QS » s'attèle à répondre.

Références bibliographiques

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A.B., and Lereclus, D. (1999) PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol Microbiol* **32**: 1043-1053.

Al Safadi, R., Mereghetti, L., Salloum, M., Lartigue, M.F., Virlogeux-Payant, I., Quentin, R., and Rosenau, A. (2011) Two-component system RgfA/C activates the *fbsB* gene encoding major fibrinogen-binding protein in highly virulent CC17 clone group B *Streptococcus*. *PLoS One* **6**: e14658.

Alix, E., and Blanc-Potard, A.B. (2009) Hydrophobic peptides: novel regulators within bacterial membrane. *Mol Microbiol* **72**: 5-11.

Alloing, G., de Philip, P., and Claverys, J.P. (1994) Three highly homologous membrane-bound lipoproteins participate in oligopeptide transport by the Ami system of the gram-positive *Streptococcus pneumoniae*. *J Mol Biol* **241**: 44-58.

Alloing, G., Granadel, C., Morrison, D.A., and Claverys, J.P. (1996) Competence pheromone, oligopeptide permease, and induction of competence in *Streptococcus pneumoniae*. *Mol Microbiol* **21**: 471-478.

Alloing, G., Martin, B., Granadel, C., and Claverys, J.P. (1998) Development of competence in *Streptococcus pneumonaie*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Mol Microbiol* **29**: 75-83.

Ammann, A., Neve, H., Geis, A., and Heller, K.J. (2008) Plasmid transfer via transduction from *Streptococcus thermophilus* to *Lactococcus lactis*. *J Bacteriol* **190**: 3083-3087.

An, F.Y., Sulavik, M.C., and Clewell, D.B. (1999) Identification and characterization of a determinant (*eep*) on the *Enterococcus faecalis* chromosome that is involved in production of the peptide sex pheromone cAD1. *J Bacteriol* **181**: 5915-5921.

Anbalagan, S., McShan, W.M., Dunman, P.M., and Chaussee, M.S. (2011) Identification of Rgg Binding Sites in the *Streptococcus pyogenes* Chromosome. *J Bacteriol* **193**: 4933-4942.

Antiporta, M.H., and Dunny, G.M. (2002) *ccfA*, the genetic determinant for the cCF10 peptide pheromone in *Enterococcus faecalis* OG1RF. *J Bacteriol* **184**: 1155-1162.

Atkinson, S., and Williams, P. (2009) Quorum sensing and social networking in the microbial world. *J R S Interface* **6**: 959-978.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **79**: 137.158.

Bae, T., Clerc-Bardin, S., and Dunny, G.M. (2000) Analysis of expression of *prgX*, a key negative regulator of the transfer of the *Enterococcus faecalis* pheromone-inducible plasmid pCF10. *J Mol Biol* **297**: 861-875.

Bae, T., Kozlowicz, B., and Dunny, G.M. (2002) Two targets in pCF10 DNA for PrgX binding: their role in production of Qa and prgX mRNA and in regulation of pheromone-inducible conjugation. *J Mol Biol* **315**: 995-1007.

Bagnoli, F., Moschioni, M., Donati, C., Dimitrovska, V., Ferlenghi, I., Facciotti, C., Muzzi, A., Giusti, F., Emolo, C., Sinisi, A., Hilleringmann, M., Pansegrau, W., Censini, S., Rappuoli, R., Covacci, A., Masignani, V., and Barocchi, M.A. (2008) A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. *J Bacteriol* **190**: 5480-5492.

Bailey, T., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A.R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D.W., Normark, S., Rappuoli, R., and Henriques-Normark, B. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci U S A* **103**: 2857-2862.

Bassler, B.L. (2002) Small talk: Cell-to-cell communication in bacteria. *Cell* **109**: 421-424.

Bassler, B.L., and Losick, R. (2006) Bacterially speaking. *Cell* **125**: 237-246.

Bensing, B.A., Manias, D.A., and Dunny, G.M. (1997) Pheromone cCF10 and plasmid pCF10-encoded regulatory molecules act post-transcriptionally to activate expression of downstream conjugation functions. *Mol Microbiol* **24**: 285-294.

Blomqvist, T., Steinmoen, H., and Håvarstein, L.S. (2006a) Natural genetic transformation: A novel tool for efficient genetic engineering of the dairy bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* **72**: 6751-6756.

Blomqvist, T., Steinmoen, H., and Håvarstein, L.S. (2006b) Pheromone-induced expression of recombinant proteins in *Streptococcus thermophilus*. *Arch Microbiol* **186**: 465-473.

Boekhorst, J., Wilson, G., Siezen, R. J. (2011) Searching in microbial genomes for encoded small proteins. *Microbial biotechnol* **4**: 308-313.

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.

Bortoni, M.E., Terra, V.S., Hinds, J., Andrew, P.W., and Yesilkaya, H. (2009) The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology* **155**: 4123–4134.

Boto, L. (2010) Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **277**: 819-827.

Bouillaut, L., Perchat, S., Arold, S., Zorrilla, S., Slamti, L., Henry, C., Gohar, M., Declerck, N., and Lereclus, D. (2008) Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. *Nucleic Acids Res* **36**: 3791-3801.

Bouillaut L. Bases moléculaires d'un système de communication cellulaire chez *Bacillus cereus*. Sous la direction de D. Lereclus. AgroParisTech. (2007) Thèse.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002) The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**: 77-97.

Buttaro, B.A., Antiporta, M.H., and Dunny, G.M. (2000) Cell-associated pheromone peptide (cCF10) production and pheromone inhibition in *Enterococcus faecalis*. *J Bacteriol* **182**: 4926-4933.

Callegan, M.C., Kane, S.T., Cochran, D.C., Gilmore, M.S., Gominet, M., and Lereclus, D. (2003) Relationship of *plcR*-Regulated Factors to *Bacillus* Endophthalmitis Virulence. *Infect Immun* **71**: 3116-3124.

Campbell, E.A., Choi, S.Y., and Masure, H.R. (1998) A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* **27**: 929-939.

Carroll, R.K., and Musser, J.M. (2011) From transcription to activation: how group A streptococcus, the flesh-eating pathogen, regulates SpeB cysteine protease production. *Mol Microbiol* **81**: 588-601.

Carroll, R.K., Shelburne, S.A., Olsen, R.J., Suber, B., Sahasrabhojane, P., Kumaraswami, M., Beres, S.B., Shea, P.R., Flores, A.R., and Musser, J.M. (2011) Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J Clin Invest* **121**: 1956-1968.

Casino, P., Rubio, V., and Marina, A. (2010) The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* **20**: 763-771.

Chan, W.C., Coyle, B.J., and Williams, P. (2004) Virulence Regulation and Quorum Sensing in Staphylococcal Infections: Competitive AgrC Antagonists as Quorum Sensing Inhibitors. *J Med Chem* **47**: 4633-4641.

Chandler, J.R., and Dunny, G.M. (2004) Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides* **25**: 1377-1388.

Chandler, J.R., and Dunny, G.M. (2008) Characterization of the sequence specificity determinants required for processing and control of sex pheromone by the intramembrane protease Eep and the plasmid-encoded protein PrgY. *J Bacteriol* **190**: 1172-1183.

Chang, J.C., LaSarre, B., Jimenez, J.C., Aggarwal, C., and Federle, M.J. (2011) Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. *PLoS Pathog* **7**: e1002190.

Chaussee, M.A., Callegari, E.A., and Chaussee, M.S. (2004) Rgg regulates growth phase-dependent expression of proteins associated with secondary metabolism and stress in *Streptococcus pyogenes*. *J Bacteriol* **186**: 7091-7099.

Chaussee, M.S., Ajdic, D., and Ferretti, J.J. (1999) The *rgg* gene of *Streptococcus pyogenes* NZ131 positively influences extracellular SPE B production. *Infect Immun* **67**: 1715-1722.

Chaussee, M.S., Watson, R.O., Smoot, J.C., and Musser, J.M. (2001) Identification of Rgg-Regulated Exoproteins of *Streptococcus pyogenes. Infect Immun* **69**: 822-831.

Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Smoot, L.M., Graham, M.R., Watson, R.O., and Musser, J.M. (2002) Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*. *Infect Immun* **70**: 762-770.

Chaussee, M.S., Somerville, G.A., Reitzer, L., and Musser, J.M. (2003) Rgg coordinates virulence factor synthesis and metabolism in *Streptococcus pyogenes*. *J Bacteriol* **185**: 6016-6024.

Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2**: 241-249.

Chen, P., Qi, F., Novak, J., and Caufield, P.W. (1999) The specific genes for lantibiotic mutacin II biosynthesis in *Streptococcus mutans* T8 are clustered and can be transferred en bloc. *Appl Environ Microbiol* **65**: 1356-1360.

Claverys, J.P., Prudhomme, M., and Martin, B. (2006) Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* **60**: 451-475.

Claverys, J.P., and Håvarstein, L.S. (2007) Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol* **5**: 219-229.

Claverys, J.P., Martin, B., and Polard, P. (2009) The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol Rev* **33**: 643-656.

Clewell, D.B., An, F.Y., Flannagan, S.E., Antiporta, M., and Dunny, G.M. (2000) Enterococcal sex pheromone precursors are part of signal sequences for surface lipoproteins. *Mol Microbiol* **35**: 246-247.

Conte, A., Chinello, P., Civljak, R., Bellussi, A., Noto, P., and Petrosillo, N. (2006) *Streptococcus salivarius* meningitis and sphenoid sinus mucocele. Case report and literature review. *J Infect* **52**: 27-30.

Cunningham, M.W. (2000) Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **13**: 470-511.

Cunningham, M.W. (2008) Pathogenesis of group A streptococcal infections and their sequelae. *Adv Exp Med Biol* **609**: 29-42.

De Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J., and Kuipers, O.P. (2006) BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res* **34**: W273-W279.

De Vuyst, L., and Degeest, B. (1999) Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiol Rev* **23**: 153-177.

Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., Lereclus, D., and Arold, S.T. (2007) Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proc Natl Acad Sci U S A* **104**: 18490-18495.

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.

Delorme, C. (2008) Safety assessment of dairy microorganisms: *Streptococcus thermophilus*. *Int J Food Microbiol* **126**: 274-277.

Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., Guedon, E., Gibrat, J.F., and Renault, P. (2011) Complete Genome Sequence of the Pigmented *Streptococcus thermophilus* Strain JIM8232. *J. Bacteriol.* **193**: 5581-5582.

Denham, E.L., Ward, P.N., and Leigh, J.A. (2008) Lipoprotein signal peptides are processed by Lsp and Eep of *Streptococcus uberis*. *J Bacteriol* **190**: 4641-4647.

Di Cagno, R., De Angelis, M., Calasso, M., and Gobbetti, M. (2011) Proteomics of the bacterial cross-talk by quorum sensing. *J Proteomics* **74**: 19-34.

Diep, D., Håvarstein, L., and Nes, I. (1996) Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. *J Bacteriol* **178**: 4472-4483.

Dmitriev, A.V., McDowell, E.J., Kappeler, K.V., Chaussee, M.A., Rieck, L.D., and Chaussee, M.S. (2006) The Rgg regulator of *Streptococcus pyogenes* influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon. *J Bacteriol* **188**: 7230-7241.

Dmitriev, A.V., McDowell, E.J., and Chaussee, M.S. (2008) Inter- and intraserotypic variation in the *Streptococcus pyogenes* Rgg regulon. *FEMS Microbiol Lett* **284**: 43-51.

Doyuk, E., Ormerod, O.J., and Bowler, I.C. (2002) Native valve endocarditis due to *Streptococcus vestibularis* and *Streptococcus oralis*. *J Infect* **45**: 39-41.

Duan, K., Dammel, C., Stein, J., Rabin, H., and Surette, M.G. (2003) Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol Microbiol* **50**: 1477-1491.

Dunny, G.M., and Leonard, B.A. (1997) Cell-cell communication in gram-positive bacteria. *Annu Rev Microbiol* **51**: 527-564.

Dunny, G.M., Antiporta, M.H., and Hirt, H. (2001) Peptide pheromone-induced transfer of plasmid pCF10 in *Enterococcus faecalis*: probing the genetic and molecular basis for specificity of the pheromone response. *Peptides* **22**: 1529-1539.

Dunny, G.M. (2007) The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc Lond B Biol Sci* **362**: 1185-1193.

Dunny, G.M., and Johnson, C.M. (2011) Regulatory circuits controlling enterococcal conjugation: lessons for functional genomics. *Curr Opin Microbiol Cell regulation* **14**: 174-180.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Ehrmann, M., Ludwig, W., and Schleifer, K.H. (1994) Reverse dot blot hybridization: a useful method for the direct identification of lactic acid bacteria in fermented food. *FEMS Microbiol Lett* **117**: 143-149.

Elli, M., Callegari, M.L., Ferrari, S., Bessi, E., Cattivelli, D., Soldi, S., Morelli, L., Goupil Feuillerat, N., and Antoine, J.M. (2006) Survival of yogurt bacteria in the human gut. *Appl Environ Microbiol* **72**: 5113-5117.

Eng, C., Thibessard, A., Danielsen, M., Rasmussen, T.B., Mari, J.F., and Leblond, P. (2011) In silico prediction of horizontal gene transfer in *Streptococcus thermophilus*. *Arch Microbiol* **193**: 287-297.

Facklam, R. (2002) What happened to the *streptococci*: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**: 613-630.

Falugi, F., Zingaretti, C., Pinto, V., Mariani, M., Amodeo, L., Manetti, A.G., Capo, S., Musser, J.M., Orefici, G., Margarit, I., Telford, J.L., Grandi, G., and Mora, M. (2008) Sequence variation in group A *Streptococcus* pili and association of pilus backbone types with lancefield T serotypes. *J Infect Dis* **198**: 1834-1841.

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, A., Thibessard, A., Borges, F., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2004) Characterization of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368. *Arch Microbiol* **182**: 364-372.

Fernandez, A., Borges, F., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2006) The *rggC* locus, with a frameshift mutation, is involved in oxidative stress response by *Streptococcus thermophilus*. *Arch Microbiol* **186**: 161-169.

Fontaine, L., Boutry, C., Guedon, E., Guillot, A., Ibrahim, M., Grossiord, B., and Hols, P. (2007) Quorum-sensing regulation of the production of Blp bacteriocins in *Streptococcus thermophilus*. *J Bacteriol* **189**: 7195-7205.

Fontaine, L., Boutry, C., Henry de Frahan, M., Delplace, B., Fremaux, C., Horvath, P., Boyaval, P., and Hols, P. (2010a) A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol* **192**: 1444-1454.

Fontaine, L., Dandoy, D., Boutry, C., Delplace, B., de Frahan, M.H., Fremaux, C., Horvath, P., Boyaval, P., and Hols, P. (2010b) Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*. *Appl Environ Microbiol* **76**: 7870-7877.

Fontecave, M., Atta, M., and Mulliez, E. (2004) S-adenosylmethionine: nothing goes to waste. *Trends in Biochemical Sciences* **29**: 243-249.

Fujiwara, T., Hoshino, T., Ooshima, T., Sobue, S., and Hamada, S. (2000) Purification, characterization, and molecular analysis of the gene encoding glucosyltransferase from *Streptococcus oralis*. *Infect Immun* **68**: 2475-2483.

Fuqua, C., and Greenberg, E.P. (2002) Listening in on bacteria: acyl-homoserine lactone signalling. *Nat Rev Mol Cell Biol* **3**: 685-695.

Gaillot, O., Poyart, C., Berche, P., and Trieu-Cuot, P. (1997) Molecular characterization and expression analysis of the superoxide dismutase gene from *Streptococcus agalactiae*. *Gene* **204**: 213-218.

Gao, R., and Stock, A.M. (2009) Biological Insights from Structures of Two-Component Proteins. *Ann Rev Microbiol* **63**: 133-154.

Garault, P., Le Bars, D., Besset, C., and Monnet, V. (2002) Three oligopeptide-binding proteins are involved in the oligopeptide transport of *Streptococcus thermophilus*. *J Biol Chem* **277**: 32-39.

Gardan, R., Besset, C., Guillot, A., Gitton, C., and Monnet, V. (2009) The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. *J Bacteriol* **191**: 4647-4655.

Giraffa, G., Paris, A., Valcavi, L., Gatti, M., Neviani, E. (2001) Genotypic and phenotypic heterogeneity of *Streptococcus thermophilus* strains isolated from dairy products. *J Appl Microbiol* **91**: 937-943.

Goebl, M., and Yanagida, M. (1991) The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* **16**: 173-177.

Gohar, M., Økstad, O.A., Gilois, N., Sanchis, V., Kolsto, A.B., and Lereclus, D. (2002) Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. *Proteomics* **2**: 784-791.

Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Økstad, O.A., Gominet, M., Kolstø, A.B., and Lereclus, D. (2008) The PlcR virulence regulon of *Bacillus cereus*. *PLoS ONE* **3**: e2793.

Gominet, M., Slamti, L., Gilois, N., Rose, M., and Lereclus, D. (2001) Oligopeptide permease is required for expression of the *Bacillus thuringiensis plcR* regulon and for virulence. *Mol Microbiol* **40**: 963-975.

Griffith, F. (1928) The Significance of Pneumococcal Types. *J Hyg (Lond)* **27**: 113-159.

Guarner, F., Perdigon, G., Corthier, G., Salminen, S., Koletzko, B., and Morelli, L. (2005) Should yoghurt cultures be considered probiotic? *Br J Nutr* **93**: 783-786.

Guiral, S., Mitchell, T.J., Martin, B., and Claverys, J.-P. (2005) Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: Genetic requirements. *Proc Natl Acad Sci USA* **102**: 8710-8715.

Gutekunst, H., Eikmanns, B.J., and Reinscheid, D.J. (2003) Analysis of RogB-controlled virulence mechanisms and gene repression in *Streptococcus agalactiae*. *Infect Immun* **71**: 5056-5064.

Hanzelka, B.L., Parsek, M.R., Val, D.L., Dunlap, P.V., Cronan, J.E., Jr., and Greenberg, E.P. (1999) Acylhomoserine lactone synthase activity of the *Vibrio fischeri* AinS protein. *J Bacteriol* **181**: 5766-5770.

Håvarstein, L.S., Holo, H., and Nes, I.F. (1994) The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. *Microbiology* **140 (Pt 9)**: 2383-2389.

Håvarstein, L.S., Gaustad, P., Nes, I.F., and Morrison, D.A. (1996) Identification of the streptococcal competence-pheromone receptor. *Mol Microbiol* **21**: 863-869.

Håvarstein, L.S., Coomaraswamy, G., and Morrison, D.A. (1995) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **92**: 11140-11144.

Håvarstein, L.S., Gaustad, P., Nes, I.F., and Morrison, D.A. (1996) Identification of the streptococcal competence-pheromone receptor. *Mol Microbiol* **21**: 863-869.

Håvarstein, L.S., Martin, B., Johnsborg, O., Granadel, C., and Claverys, J.-P. (2006) New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Mol Microbiol* **59**: 1297-1037.

Hemm, M.R., Paul, B.J., Miranda-Rios, J., Zhang, A., Soltanzad, N., and Storz, G. (2010) Small Stress Response Proteins in *Escherichia coli*: Proteins Missed by Classical Proteomic Studies. *J Bacteriol* **192**: 46-58.

Henry, R., Bruneau, E., Gardan, R., Bertin, S., Fleuchot, B., Decaris, B., and Leblond-Bourget, N. (2011) The *rgg0182* gene encodes a transcriptional regulator required for the full *Streptococcus thermophilus* LMG18311 thermal adaptation. *BMC Microbiol* **11**: 223.

Hense, B.A., Kuttler, C., Muller, J., Rothballer, M., Hartmann, A., and Kreft, J.-U. (2007) Does efficiency sensing unify diffusion and quorum sensing? **5**: 230-239.

Higgins, C.F. (2001) ABC transporters: physiology, structure and mechanism--an overview. *Res Microbiol* **152**: 205-210.

Higgins, C.F. (2007) Multiple molecular mechanisms for multidrug resistance transporters. *Nature* **446**: 749-757.

Hoch, J.A. (2000) Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology* **3**: 165-170.

Hollands, A., Aziz, R.K., Kansal, R., Kotb, M., Nizet, V., and Walker, M.J. (2008) A Naturally Occurring Mutation in *ropB* Suppresses SpeB Expression and Reduces M1T1 Group A Streptococcal Systemic Virulence. *PLoS ONE* **3**: e4102.

Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Dusko Ehrlich, S., Guédon, 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.

Hughes, D.T., and Sperandio, V. (2008) Inter-kingdom signalling: communication between bacteria and their hosts. *Nature Rev* **6**: 111-120.

Hui, F.M., Zhou, L., and Morrison, D.A. (1995) Competence for genetic transformation in *Streptococcus pneumoniae*: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene* **153**: 25-31.

Hutchings, M.I., Palmer, T., Harrington, D.J., and Sutcliffe, I.C. (2009) Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold them, knowing when to fold them. *Trends Microbiol* **17**: 13-21.

Ibrahim, M., Guillot, A., Wessner, F., Algaron, F., Besset, C., Courtin, P., Gardan, R., and Monnet, V. (2007a) Control of the transcription of a short gene encoding a cyclic peptide in *Streptococcus thermophilus*: a new quorum-sensing system? *J Bacteriol* **189**: 8844-8854.

Ibrahim, M., Nicolas, P., Bessières, P., Bolotin, A., Monnet, V., and Gardan, R. (2007b) A genome-wide survey of short coding sequences in *streptococci*. *Microbiology* **153**: 3631-3644.

Idigoras, P., Valiente, A., Iglesias, L., Trieu-Cout, P., and Poyart, C. (2001) Meningitis due to *Streptococcus salivarius*. *J Clin Microbiol* **39**: 3017.

Ikebe, T., Ato, M., Matsumura, T., Hasegawa, H., Sata, T., Kobayashi, K., and Watanabe, H. (2010) Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. *PLoS Pathog* **6**: e1000832.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., Chu, L., Mazur, M., Goltsman, E., Larsen, N., D'Souza, M., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Dusko Ehrlich, S., Overbeek, R., and Kyrpides, N. (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423**: 87-91.

Ji, G., Beavis, R.C., and Novick, R.P. (1995) Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc. Natl. Acad. Sci. U S A* **92**: 12055-12059.

Ji, G., Beavis, R., and Novick, R.P. (1997) Bacterial interference caused by autoinducing peptide variants. *Science* **276**: 2027-2030.

Johnsborg, O., Eldholm, V., and Håvarstein, L.S. (2007) Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**: 767-778.

Johnsborg, O., and Håvarstein, L.S. (2009) Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiol Rev* **33**: 627-642.

Juille, O., Le Bars, D., and Juillard, V. (2005) The specificity of oligopeptide transport by *Streptococcus thermophilus* resembles that of *Lactococcus lactis* and not that of pathogenic *streptococci*. *Microbiology* **151**: 1987-1994.

Kappeler, K.V., Anbalagan, S., Dmitriev, A.V., McDowell, E.J., Neely, M.N., and Chaussee, M.S. (2009) A naturally occurring Rgg variant in serotype M3 *Streptococcus pyogenes* does not activate *speB* expression due to altered specificity of DNA binding. *Infect Immun* **77**: 5411-5417

Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E., and Håvarstein, L.S. (2005) Choline-Binding Protein D (CbpD) in *Streptococcus pneumoniae* Is Essential for Competence-Induced Cell Lysis. *J Bacterio.* **187**: 4338-4345.

Kawamura, Y., Hou, X.G., Sultana, F., Liu, S., Yamamoto, H., and Ezaki, T. (1995a) Transfer of *Streptococcus adjacens* and *Streptococcus defectivus* to *Abiotrophia* gen. nov. as *Abiotrophia adiacens* comb. nov. and *Abiotrophia defectiva* comb. nov., respectively. *Int J Syst Bacteriol* **45**: 798-803.

Kawamura, Y., Hou, X.G., Sultana, F., Miura, H., and Ezaki, T. (1995b) 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.

Keller, L., and Surette, M.G. (2006) Communication in bacteria: an ecological and evolutionary perspective. *Nat Rev Microbiol* **4**: 249-258.

Keyes, P.H. (1960) The infectious and transmissible nature of experimental dental caries. Findings and implications. *Arch Oral Biol* **1**: 304-320.

Koenig, R.L., Ray, J.L., Maleki, S.J., Smeltzer, M.S., and Hurlburt, B.K. (2004) *Staphylococcus aureus* AgrA Binding to the RNAIII-agr Regulatory Region. *J Bacteriol* **186**: 7549-7555.

Koonin, E.V., Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709-742.

Kozlowicz, B.K., Dworkin, M., and Dunny, G.M. (2006a) Pheromone-inducible conjugation in *Enterococcus faecalis*: a model for the evolution of biological complexity? *Int J Med Microbiol* **296**: 141-147.

Kozlowicz, B.K., Shi, K., Gu, Z.Y., Ohlendorf, D.H., Earhart, C.A., and Dunny, G.M. (2006b) Molecular basis for control of conjugation by bacterial pheromone and inhibitor peptides. *Mol Microbiol* **62**: 958-969.

Kreikemeyer, B., McIver, K.S., and Podbielski, A. (2003) Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* **11**: 224-232.

Kreikemeyer, B., Gamez, G., Margarit, I., Giard, J.C., Hammerschmidt, S., Hartke, A., and Podbielski, A. (2011) Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic *Streptococci* and *Enterococci*. *Int J Med Microbiol* **301**: 240-251.

Kreth, J., Merritt, J., Bordador, C., Shi, W., and Qi, F. (2004) Transcriptional analysis of mutacin I (*mutA*) gene expression in planktonic and biofilm cells of *Streptococcus mutans* using fluorescent protein and glucuronidase reporters. *Oral Microbiol Immunol* **19**: 252-256.

Kreth, J., Merritt, J., Shi, W., and Qi, F. (2005) Co-ordinated bacteriocin production and competence development: a possible mechanism for taking up DNA from neighbouring species. *Mol Microbiol* **57**: 392-404.

Lancefield, R.C. (1933) A Serological Differentiation of Human and Other Groups of Hemolytic *Streptococci*. *J Exp Med* **57**: 571-595.

Lanigan-Gerdes, S., Dooley, A.N., Faull, K.F., and Lazazzera, B.A. (2007) Identification of subtilisin, Epr and Vpr as enzymes that produce CSF, an extracellular signalling peptide of *Bacillus subtilis*. *Mol Microbiol* **65**: 1321-1333.

Lauer, P., Rinaudo, C.D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A.R., Mora, M., Rappuoli, R., Grandi, G., and Telford, J.L. (2005) Genome analysis reveals pili in Group B *Streptococcus*. *Science* **309**: 105.

Lazazzera, B.A. (2001) The intracellular function of extracellular signaling peptides. *Peptides* **22**: 1519-1527.

Ledenberg, J., and Tatum, E.L. (1946) Gene recombination in *Escherichia coli*. *Nature* **19**: 458-558

Lee, J.-H., and Lee, J. (2010) Indole as an intercellular signal in microbial communities. *FEMS Microbiol Rev* **34**: 426-444.

Lee, M.S., and Morrison, D.A. (1999) Identification of a New Regulator in *Streptococcus pneumoniae* Linking Quorum Sensing to Competence for Genetic Transformation. *J Bacteriol* **181**: 5004-5016.

Lefébure, 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.

Leonard, B.A., Podbielski, A., Hedberg, P.J., and Dunny, G.M. (1996) *Enterococcus faecalis* pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. *Proc Natl Acad Sci U S A* **93**: 260-264.

Lereclus, D., Agaisse, H., Gominet, M., Salamitou, S., and Sanchis, V. (1996) Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. *J Bacteriol* **178**: 2749-2756.

Lereclus, D., Agaisse, H., Grandvalet, C., Salamitou, S., and Gominet, M. (2000) Regulation of toxin and virulence gene transcription in *Bacillus thuringiensis*. *Int J Med Microbiol* **290**: 295-299.

Letort, C., and Juillard, V. (2001) Development of a minimal chemically-defined medium for the exponential growth of *Streptococcus thermophilus*. *J Appl Microbiol* **91**: 1023-1029.

Li, J., Wang, W., Xu, S.X., Magarvey, N.A., and McCormick, J.K. (2011) *Lactobacillus reuteri*-produced cyclic dipeptides quench agr-mediated expression of toxic shock syndrome toxin-1 in staphylococci. *Proc Natl Acad Sci USA* **108**: 3360-3365.

Li, Y.H., Lau, P.C.Y., Lee, J.H., Ellen, R.P., and Cvitkovitch, D.G. (2001) Natural Genetic Transformation of *Streptococcus mutans* Growing in Biofilms. *J Bacteriol* **183**: 897-908.

Li, Y.H., Tang, N., Aspiras, M.B., Lau, P.C.Y., Lee, J.H., Ellen, R.P., and Cvitkovitch, D.G. (2002) A Quorum-Sensing Signaling System Essential for Genetic Competence in *Streptococcus mutans* Is Involved in Biofilm Formation. *J Bacteriol* **184**: 2699-2708.

Lick, S., Keller, M., Bockelmann, W., and Heller, K.J. (1996) Rapid identification of *Streptococcus thermophilus* by primer-specific PCR amplification based on its *lacZ* gene. *Syst Appl Microbiol* **19**: 74-77.

Linton, K.J., and Higgins, C.F. (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch* **453**: 555-567.

Liu, M., Siezen, R.J., and Nauta, A. (2009) *In silico* prediction of horizontal gene transfer events in *Lactobacillus bulgaricus* and *Streptococcus thermophilus* reveals protocooperation in yogurt manufacturing. *Appl Environ Microbiol* **75**: 4120-4129.

Liu, S.Q. (2003) Practical implications of lactate and pyruvate metabolism by lactic acid bacteria in food and beverage fermentations. *Int J Food Microbiol* **83**: 115-131.

Loughman, J.A., and Caparon, M.G. (2006) A novel adaptation of aldolase regulates virulence in *Streptococcus pyogenes*. *Embo J* **25**: 5414-5422.

Loughman, J.A., and Caparon, M.G. (2007a) Comparative functional analysis of the *lac* operons in *Streptococcus pyogenes*. *Mol Microbiol* **64**: 269-280.

Loughman, J.A., and Caparon, M.G. (2007b) Contribution of invariant residues to the function of Rgg family transcription regulators. *J Bacteriol* **189**: 650-655.

Luo, P., Li, H., and Morrison, D.A. (2004) Identification of ComW as a new component in the regulation of genetic transformation in *Streptococcus pneumoniae*. *Mol Microbiol* **54**: 172-183.

Lyon, W.R., Gibson, C.M., and Caparon, M.G. (1998) A role for trigger factor and an Rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of *Streptococcus pyogenes*. *Embo J* **17**: 6263-6275.

MacQuarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011) Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics* **27**: 141-148.

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 USA* **103**: 15611-15616.

Marciset, O., and Mollet, B. (1994) Multifactorial experimental design for optimizing transformation: Electroporation of *Streptococcus thermophilus*. *Biotechnol Bioeng* **43**: 490-496.

Martin, B., Quentin, Y., Fichant, G., and Claverys, J.P. (2006) Independent evolution of competence regulatory cascades in *streptococci*? *Trends Microbiol* **14**: 339-345.

Martin, B., Granadel, C., Campo, N., Hénard, V., Prudhomme, M., and Claverys, J.P. (2010) Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol Microbiol* **75**: 1513-1528.

Mashburn-Warren, L., Morrison, D.A., and Federle, M.J. (2010) A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol Microbiol* **78**: 589-606.

McDonnell, G.E., Wood, H., Devine, K.M., and McConnell, D.J. (1994) Genetic control of bacterial suicide: regulation of the induction of PBSX in *Bacillus subtilis*. *J Bacteriol* **176**: 5820-5830.

McIver, K.S. (2009) Stand-alone response regulators controlling global virulence networks in *streptococcus pyogenes*. *Contrib Microbiol* **16**: 103-119.

Menard, S., Candalh, C., Bambou, J.C., Terpend, K., Cerf-Bensussan, N., and Heyman, M. (2004) Lactic acid bacteria secrete metabolites retaining anti-inflammatory properties after intestinal transport. *Gut* **53**: 821-828.

Mercenier, A., Slos, P., Faelen, M., and Lecocq, J.P. (1988) Plasmid transduction in *Streptococcus thermophilus*. *Mol Gen Genet* **212**: 386-389.

Merritt, J., Kreth, J., Shi, W., and Qi, F. (2005) LuxS controls bacteriocin production in *Streptococcus mutans* through a novel regulatory component. *Mol Microbiol* **57**: 960-969.

Michiels, J., Dirix, G., Vanderleyden, J., and Xi, C. (2001) Processing and export of peptide pheromones and bacteriocins in Gram-negative bacteria. *Trends Microbiol* **9**: 164-168.

Miller, M.B., and Bassler, B.L. (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* **55**: 165-199.

Mitchell, A.M., and Mitchell, T.J. (2010) *Streptococcus pneumoniae*: virulence factors and variation. *Clin Microbiol Infect* **16**: 411-418.

Monnet, V. (2003) Bacterial oligopeptide-binding proteins. *Cell Mol Life Sci* **60**: 2100-2114.

Mora, D., Fortina, M. G., Parini, C., Ricci, G., Gatti, M., Giraffa, G., and Manachini, P.L. (2002) Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J Appl Microbiol* **93**: 278-287.

Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G., Maggi, T., Taddei, A.R., Grandi, G., and Telford, J.L. (2005) Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A* **102**: 15641-15646.

Moschetti, G., Blaiotta, G., Aponte, M., Catzeddu, P., Villani, F., Deiana, P., and Coppola, S. (1998) Random amplified polymorphic DNA and amplified ribosomal DNA spacer polymorphism: powerful methods to differentiate *Streptococcus thermophilus* strains. *J Appl Microbiol* **85**: 25-36.

Mullaney, D.M. (2001) Group B streptococcal infections in newborns. *J Obstet Gynecol Neonatal Nurs* **30**: 649-658.

Nealson, K. H., Platt, T., and Hastings, J.W. (1970) Cellular control of the synthesis and activity of the bacterial biolumionescent system. *J Bacteriol* **104**: 313-322.

Nealson, K.H., Hastings, J.W. (1979) Bacterial bioluminescence: its control and ecological significance. *Microbiol Rev* **43**:496-518.

Neely, M.N., Lyon, W.R., Runft, D.L., and Caparon, M. (2003) Role of RopB in growth phase expression of the SpeB cysteine protease of *Streptococcus pyogenes*. *J Bacteriol* **185**: 5166-5174.

Nepomuceno, R.S., Tavares, M.B., Lemos, J.A., Griswold, A.R., Ribeiro, J.L., Balan, A., Guimarães, K.S., Cai, S., Burne, R.A., Ferreira, L.C., and Ferreira, R.C. (2007) The oligopeptide (*opp*) gene cluster of *Streptococcus mutans*: identification, prevalence, and characterization. *Oral Microbiol Immunol* **22**: 277-284.

Ng, W.L., and Bassler, B.L. (2009) Bacterial quorum-sensing network architectures. *Annu Rev Genet* **43**: 197-222.

Ng, W.L., Perez, L.J., Wei, Y., Kraml, C., Semmelhack, M.F., and Bassler, B.L. (2011) Signal production and detection specificity in *Vibrio* CqsA/CqsS quorum-sensing systems. *Mol Microbiol* **79**: 1407-1417.

Nicolas, G.G., and Lavoie, M.C. (2011) *Streptococcus mutans* et les streptocoques buccaux dans la plaque dentaire. *Can J Microbiol* **57**: 1-20.

Nobbs, A.H., Lamont, R.J., and Jenkinson, H.F. (2009) *Streptococcus* adherence and colonization. *Microbiol Mol Biol Rev* **73**: 407-450.

Novick, R.P. (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol* **48**: 1429-1449.

Novick, R.P., and Geisinger, E. (2008) Quorum sensing in *staphylococci*. *Annu Rev Genet* **42**: 541-564.

Oehmcke, S., Shannon, O., Morgelin, M., and Herwald, H. (2010) Streptococcal M proteins and their role as virulence determinants. *Clin Chim Acta* **411**: 1172-1180.

Økstad, O.A., Gominet, M., Purnelle, B., Rose, M., Lereclus, D., and Kolsto, A.B. (1999) Sequence analysis of three *Bacillus cereus* loci carrying PIcR-regulated genes encoding degradative enzymes and enterotoxin. *Microbiology* **145**: 3129-3138.

O'Sullivan, T.F., and Fitzgerald, G.F. (1998) Comparison of *Streptococcus thermophillus* strains by pulse field gel electrophoresis of genomic DNA. *FEMS Microbiol Lett* **168**: 213-219.

Otto, M. (2009) *Staphylococcus epidermidis* - the 'accidental' pathogen. *Nat Rev Microbiol* **7**: 555-567.

Pacheco, A.R., and Sperandio, V. (2009) Inter-kingdom signaling: chemical language between bacteria and host. *Curr Opin Microbiol Cell Regulation* **12**: 192-198.

Palmer, T., Sargent, F., and Berks, B.C. (2005) Export of complex cofactor-containing proteins by the bacterial Tat pathway. *Trends Microbiol* **13**: 175-180.

Park, H.K., Lee, H.J., Jeong, E.G., Shin, H.S., and Kim, W. (2010) The *rgg* gene is a specific marker for *Streptococcus oralis*. *J Dent Res* **89**: 1299-1303.

Park, I.H., Pritchard, D.G., Cartee, R., Brandao, A., Brandileone, M.C., and Nahm, M.H. (2007) Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J Clin Microbiol* **45**: 1225-1233.

Partridge, S.M. (2000) Prosthetic valve endocarditis due to *Streptococcus vestibularis*. *J Infect* **41**: 284-285.

Pei, J., and Grishin, N.V. (2001) Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends in Biochemical Sciences* **26**: 275-277.

Pestova, E.V., Håvarstein, L.S., and Morrison, D.A. (1996) Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**: 853-862.

Petersen, F.C., Tao, L., and Scheie, A.A. (2005) DNA Binding-Uptake System: a Link between Cell-to-Cell Communication and Biofilm Formation. *J Bacteriol* **187**: 4392-4400.

Pfeiler, E.A., and Klaenhammer, T.R. (2007) The genomics of lactic acid bacteria. *Trends Microbiol* **15**: 546-553.

Picon, A., de Torres, B., Gaya, P., and Nuñez, M. (2005) Cheesemaking with a *Lactococcus lactis* strain expressing a mutant oligopeptide binding protein as starter results in a different peptide profile. *Int J Food Microbiol* **104**: 299-307.

Platt, T.G., and Fuqua, C. (2010) What's in a name? The semantics of quorum sensing. *Trends Microbiol* **18**: 383-387.

Pombert, J.F., Sistek, V., Boissinot, M., and Frenette, M. (2009) Evolutionary relationships among salivarius *streptococci* as inferred from multilocus phylogenies based on 16S rRNA-encoding, *recA*, *secA*, and *secY* gene sequences. *BMC Microbiol* **9**: 232.

Pomerantsev, A.P., Pomerantseva, O.M., Camp, A.S., Mukkamala, R., Goldman, S., and Leppla, S.H. (2009) PapR peptide maturation: role of the NprB protease in *Bacillus cereus* 569 PlcR/PapR global gene regulation. *FEMS Immunol Med Microbiol* **55**: 361-377.

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.

Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B., and Claverys, J.-P. (2006) Antibiotic Stress Induces Genetic Transformability in the Human Pathogen *Streptococcus pneumoniae*. *Science* **313**: 89-92.

Pulliainen, A.T., Hytönen, J., Haataja, S., and Finne, J. (2008) Deficiency of the Rgg regulator promotes H2O2 resistance, AhpCF-mediated H2O2 decomposition, and virulence in *Streptococcus pyogenes*. *J Bacteriol* **190**: 3225-3235.

Qi, F., Chen, P., and Caufield, P.W. (1999a) Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in *Streptococcus mutans*. *Appl Environ Microbiol* **65**: 652-658.

Qi, F., Chen, P., and Caufield, P.W. (1999b) Purification of mutacin III from group III *Streptococcus mutans* UA787 and genetic analyses of mutacin III biosynthesis genes. *Appl Environ Microbiol* **65**: 3880-3887.

Qi, F., Chen, P., and Caufield, P.W. (2000) Purification and biochemical characterization of mutacin I from the group I strain of *Streptococcus mutans*, CH43, and genetic analysis of mutacin I biosynthesis genes. *Appl Environ Microbiol* **66**: 3221-3229.

Rajagopal, L. (2009) Understanding the regulation of Group B Streptococcal virulence factors. *Future Microbiol* **4**: 201-221.

Rasmussen, T.B., Danielsen, M., Valina, O., Garrigues, C., Johansen, E., and Pedersen, M.B. (2008) *Streptococcus thermophilus* core genome: comparative genome hybridization study of 47 strains. *Appl Environ Microbiol* **74**: 4703-4710.

Rawlinson, E.L., Nes, I.F., and Skaugen, M. (2002) LasX, a transcriptional regulator of the lactocin S biosynthetic genes in *Lactobacillus sakei* L45, acts both as an activator and a repressor. *Biochimie* **84**: 559-567.

Rawlinson, E.L., Nes, I.F., and Skaugen, M. (2005) Identification of the DNA-binding site of the Rgg-like regulator LasX within the lactocin S promoter region. *Microbiology* **151**: 813-823.

Rea, M.C., Sit, C.S., Clayton, E., O'Connor, P.M., Whittal, R.M., Zheng, J., Vederas, J.C., Ross, R.P., and Hill, C. (2010) Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc Natl Acad Sci U S A* **107**: 9352-9357.

Redfield, R.J. (2002) Is quorum sensing a side effect of diffusion sensing? *Trends in Microbiology* **10**: 365-370.

Rezzonico, F., and Duffy, B. (2008) Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for luxS in most bacteria. *BMC Microbiol* **8**: 154.

Rickard, A.H., Palmer, R.J., Jr., Blehert, D.S., Campagna, S.R., Semmelhack, M.F., Egland, P.G., Bassler, B.L., and Kolenbrander, P.E. (2006) Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Mol Microbiol* **60**: 1446-1456.

Rocha-Estrada, J., Aceves-Diez, A.E., Guarneros, G., and de la Torre, M. (2010) The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Appl Microbiol Biotechnol* **87**: 913-923.

Rosini, R., Rinaudo, C.D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzo, C., Brettoni, C., Buccato, S., Margarit, I., Grandi, G., and Telford, J.L. (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. *Mol Microbiol* **61**: 126-141.

Ryan, R.P., and Dow, J.M. (2008) Diffusible signals and interspecies communication in bacteria. *Microbiology* **154**: 1845-1858.

Salamitou, S., Ramisse, F.o., Brehélin, M., Bourguet, D., Gilois, N., Gominet, M., Hernandez, E., and Lereclus, D. (2000) The *plcR* regulon is involved in the opportunistic properties of *Bacillus thuringiensis* and *Bacillus cereus* in mice and insects. *Microbiology* **146**: 2825-2832.

Samen, U., Gottschalk, B., Eikmanns, B.J., and Reinscheid, D.J. (2004) Relevance of peptide uptake systems to the physiology and virulence of *Streptococcus agalactiae*. *J Bacteriol* **186**: 1398-1408.

Samen, U.M., Eikmanns, B.J., and Reinscheid, D.J. (2006) The transcriptional regulator RovS controls the attachment of *Streptococcus agalactiae* to human epithelial cells and the expression of virulence genes. *Infect Immun* **74**: 5625-5635.

Sanders, J.W., Leenhouts, K., Burghoorn, J., Brands, J.R., Venema, G., and Kok, J. (1998) A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol Microbiol* **27**: 299-310.

Saurin, W., Hofnung, M., and Dassa, E. (1999) Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol* **48**: 22-41.

Schauder, S., and Bassler, B.L. (2001) The languages of bacteria. *Genes Dev* **15**: 1468-1480.
Schleifer, K.H. (2009) Classification of Bacteria and Archaea: past, present and future. *Syst Appl Microbiol* **32**: 533-542.
Schmoll, M., Seibel, C., Tisch, D., Dorrer, M., and Kubicek, C.P. (2010) A novel class of peptide pheromone precursors in ascomycetous fungi. *Mol Microbiol* **77**: 1483-1501.
Sherman, J.M. (1937) The *Streptococci*. *Bacteriol Rev* **1**: 3-97.
Shi, K., Brown, C.K., Gu, Z.Y., Kozlowicz, B.K., Dunny, G.M., Ohlendorf, D.H., and Earhart, C.A. (2005) Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in *Enterococcus faecalis*. *Proc Natl Acad Sci U S A* **102**: 18596-18601.
Skaugen, M., Andersen, E.L., Christie, V.H., and Nes, I.F. (2002) Identification, characterization, and expression of a second, bicistronic, operon involved in the production of lactocin S in *Lactobacillus sakei* L45. *Appl Environ Microbiol* **68**: 720-727.
Slamti, L., and Lereclus, D. (2002) A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *Embo J* **21**: 4550-4559.
Slamti, L., and Lereclus, D. (2005) Specificity and polymorphism of the PlcR-PapR quorum-sensing system in the *Bacillus cereus* group. *J Bacteriol* **187**: 1182-1187.
Somkuti, G.A., and Steinberg, D.H. (1988) Genetic transformation of *Streptococcus thermophilus* by electroporation. *Biochimie* **70**: 579-585.
Sørensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* **3**: 700-710.
Stiles, M.E., and Holzapfel, W.H. (1997) Lactic acid bacteria of foods and their current taxonomy. *Int J Food Microbiol* **36**: 1-29.
Studer, S.V., Mandel, M.J., and Ruby, E.G. (2008) AinS Quorum Sensing Regulates the *Vibrio fischeri* Acetate Switch. *J Bacteriol* **190**: 5915-5923.
Sulavik, M.C., Tardif, G., and Clewell, D.B. (1992) Identification of a gene, *rgg*, which regulates expression of glucosyltransferase and influences the Spp phenotype of *Streptococcus gordonii* Challis. *J Bacteriol* **174**: 3577-3586.
Sulavik, M.C., and Clewell, D.B. (1996) Rgg is a positive transcriptional regulator of the *Streptococcus gordonii gtfG* gene. *J Bacteriol* **178**: 5826-5830.
Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Wu, L., Zhou, Z., Sun, T., Wang, L., Meng, H., Zhang, H., and Chen, W. (2011) Complete genome sequence of *Streptococcus thermophilus* strain ND03. *J Bacteriol* **193**: 793-794.
Sung, C.K., and Morrison, D.A. (2005) Two Distinct Functions of ComW in Stabilization and Activation of the Alternative Sigma Factor ComX in *Streptococcus pneumoniae*. *J Bacteriol* **187**: 3052-3061.
Teuber, M., Meile, L., and Schwarz, F. (1999) Acquired antibiotic resistance in lactic acid bacteria from food. *Antonie van Leeuwenhoek* **76**: 115-137.
Thibessard, A., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2002) Transposition of pGh9:ISS1 is random and efficient in *Streptococcus thermophilus* CNRZ368. *Can J Microbiol* **48**: 473-478.
Thibessard, A., Borges, F., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2004) Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. *Appl Environ Microbiol* **70**: 2220-2229.
Thoendel, M., and Horswill, A.R. (2009) Identification of *Staphylococcus aureus* AgrD residues required for autoinducing peptide biosynthesis. *J Biol Chem* **284**: 21828-21838.
Thoendel, M., and Horswill, A.R. (2010) Biosynthesis of peptide signals in gram-positive bacteria. *Adv Appl Microbiol* **71**: 91-112.
Thoendel, M., Kavanaugh, J.S., Flack, C.E., and Horswill, A.R. (2011) Peptide signaling in the *staphylococci*. *Chem Rev* **111**: 117-151.
Thomas, C.M., and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**: 711-721.
Tomasz, A. (1965) Control of the competent state in *pneumococcus* by a hormone-like cell product -an example of a new type of regulatory mechanism in bacteria. *Nature* **208**: 155-159.
Toomey, N., Monaghan, A., Fanning, S., and Bolton, D. (2009) Transfer of Antibiotic Resistance Marker Genes between Lactic Acid Bacteria in Model Rumen and Plant Environments. *Appl Environ Microbiol* **75**: 3146-3152.
Tortosa, P., and Dubnau, D. (1999) Competence for transformation: a matter of taste. *Curr Opin Microbiol* **2**: 588-592.

Van der Ploeg, J.R. (2005) Regulation of Bacteriocin Production in *Streptococcus mutans* by the Quorum-Sensing System Required for Development of Genetic Competence. *J Bacteriol* **187**: 3980-3989.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**: 407-438.

Venturi, V. (2006) Regulation of quorum sensing in *Pseudomonas*. *FEMS Microbiol Rev* **30**: 274-291.

Vickerman, M.M., Sulavik, M.C., and Clewell, D.B. (1995) Oral streptococci with genetic determinants similar to the glucosyltransferase regulatory gene, *rgg*. *Infect Immun* **63**: 4524-4527.

Vickerman, M.M., Jones, G.W., and Clewell, D.B. (1997) Molecular analysis of representative *Streptococcus gordonii* Spp phase variants reveals no differences in the glucosyltransferase structural gene, gtfG. *Oral Microbiol Immunol* **12**: 82-90.

Vickerman, M.M., Minick, P.E., and Mather, N.M. (2001) Characterization of the *Streptococcus gordonii* chromosomal region immediately downstream of the glucosyltransferase gene. *Microbiology* **147**: 3061-3070.

Vickerman, M.M., and Minick, P.E. (2002) Genetic analysis of the *rgg-gtfG* junctional region and its role in *Streptococcus gordonii* glucosyltransferase activity. *Infect Immun* **70**: 1703-1714.

Vickerman, M.M., Wang, M., and Baker, L.J. (2003) An amino acid change near the carboxyl terminus of the *Streptococcus gordonii* regulatory protein Rgg affects its abilities to bind DNA and influence expression of the glucosyltransferase gene *gtfG*. *Microbiology* **149**: 399-406.

Vickerman, M.M., Flannagan, S.E., Jesionowski, A.M., Brossard, K.A., Clewell, D.B., and Sedgley, C.M. (2010) A genetic determinant in *Streptococcus gordonii* Challis encodes a peptide with activity similar to that of enterococcal sex pheromone cAM373, which facilitates intergeneric DNA transfer. *J Bacteriol* **192**: 2535-2545.

Villani, F., Pepe, O., Mauriello, G., Salzano, G., Moschetti, G., and Coppola, S. (1995) Antilisterial activity of thermophilin 347, a bacteriocin produced by *Streptococcus thermophilus*. *Int J Food Microbiol* **25**: 179-190.

Walters, M., and Sperandio, V. (2006) Quorum sensing in *Escherichia coli* and *Salmonella*. *Int J Med Microbiol* **296**: 125-131.

Waters, C.M., and Bassler, B.L. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **21**: 319-346.

Ween, O., Gaustad, P., and Håvarstein, L.S. (1999) Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol Microbiol* **33**: 817-827.

Winans, S.C. (2006) Bacterial evolution by intelligent design. *ACS Chem Biol* **1**: 429-431.

Winzer, K., Hardie, K.R., and Williams, P. (2002) Bacterial cell-to-cell communication: sorry, can't talk now - gone to lunch! *Curr Opin Microbiol* **5**: 216-222.

Woodruff, W.A., Novak, J., and Caufield, P.W. (1998) Sequence analysis of *mutA* and *mutM* genes involved in the biosynthesis of the lantibiotic mutacin II in *Streptococcus mutans*. *Gene* **206**: 37-43.

Wozniak, R.A., and Waldor, M.K. (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **8**: 552-563.

Wright, J.S., III, Traber, K.E., Corrigan, R., Benson, S.A., Musser, J.M., and Novick, R.P. (2005) The *agr* Radiation: an Early Event in the Evolution of Staphylococci. *J Bacteriol* **187**: 5585-5594.

Wuster, A., and Babu, M.M. (2008) Conservation and Evolutionary Dynamics of the *agr* Cell-to-Cell Communication System across Firmicutes. *J Bacteriol* **190**: 743-746.

Yang, J., Evans, B.A., and Rozen, D.E. (2010) Signal diffusion and the mitigation of social exploitation in pneumococcal competence signalling. *P R S Biological Sciences* **277**: 2991-2999.

Yang, Z., and Nielsen, R. (2000) Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol Biol Evol* **17**: 32-43.

Yarwood, J.M., and Schlievert, P.M. (2003) Quorum sensing in *Staphylococcus* infections. *J Clin Invest* **112**: 1620-1625.

Zhang, L., Lin, J., and Ji, G. (2004) Membrane anchoring of the AgrD N-terminal amphipathic region is required for its processing to produce a quorum-sensing pheromone in *Staphylococcus aureus*. *J. Biol. Chem.* **279**: 19448-19456.

Zhang, Z.G., Ye, Z.Q., Yu, L., and Shi, P. (2011) Phylogenomic reconstruction of lactic acid bacteria: an update. *BMC Evol Biol* **11**: 1.

Zheng, F., Ji, H., Cao, M., Wang, C., Feng, Y., Li, M., Pan, X., Wang, J., Qin, Y., Hu, F., and Tang, J. (2011) Contribution of the Rgg transcription regulator to metabolism and virulence of *Streptococcus suis* serotype 2. *Infect Immun* **79**: 1319-1328

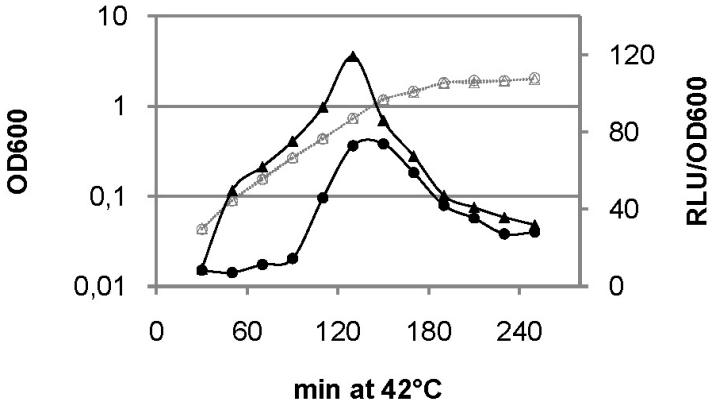
Zuber, P. (2001) A peptide profile of the *Bacillus subtilis* genome. *Peptides* **22**: 1555-1577.

Annexes

Annexe 1

Fig. S1

Fig. S1 – The triggering of the expression of the *shp* gene is a competence of the presence of the SHP1358(15-23) pheromone in the culture medium and not of the transition from the exponential growth phase to the stationary phase



Growth and luciferase activity of the strain LMD-9 containing a P*shp*$_{1358}$-*luxAB* fusion after adding (▲) or not adding (●) the synthetic SHP1358$_{(15-23)}$ pheromone at 2500 ng/ml 30 min following the inoculation of the culture at OD_{600} 0.025. Growth curves (OD_{600}) are presented in gray and luciferase activities (RLU / OD_{600}) in black. Data shown are representative of three independent experiments.

Annexe 2

Table S1

Supplementary Table S1. Rgg-like proteins associated or not with SHP or XIP hydrophobic peptides in Gram positive bacteria.

Notes:

a. Location of the gene on the sequence as given in the GenBank file, "complement" indicates that the ORF is encoded on the reverse strand.

b. This number indicates the position of the locus in the tree presented in Fig. 1 by providing the angle in degree computed counter-clockwise. The * symbol in Fig.1 marks the first locus in the table (0°).

c. Type (SHP/Rgg or XIP/ComR) indicates if a SHP or XIP peptide encoding gene has been identified upstream or downstream from the *rgg*-like gene according to the criteria described in note d. Three groups of SHP/Rgg loci have been delineated (see main text and Fig. 1) and the group number is also reported in this column in brackets.

d. The following criteria have been used for entry of a peptide in these columns: 1) a high hydrophobicity, 2) the presence of basic residues (lys and/or arg) at the N-terminal part 3) the presence of a glu or an asp residue at their C-terminal part (SHP type) or a double tryp motif (XIP type). The XIP peptides that do not follow strictly these criteria but that were described by Mashburn-Warren et al (2010) were added to the list.

e. This number indicates whether the gene encoding the hydrophobic peptide is transcribed in the same orientation as the *rgg* or *comR* gene (1) or in the opposite orientation (-1).

Organism		Rgg-like proteins		SHP/Rgg and XIP/ComR proteins		Upstream peptide[d]			Downstream peptide[d]		
Strains	GenBank id	Location[a]	Position in tree[b]	Type[c]	Genbank id	Location[a]	Strand[e]	Sequence	Location[a]	Strand[e]	Sequence
Lactobacillus plantarum JDM1	CP001617	2506695..2507606	0								
Lactobacillus plantarum WCFS1	AL935263	2595720..2596631	0,767590618								
Lactobacillus brevis ATCC 367	CP000416	complement(460040..460924)	1,535181237								
Leuconostoc kimchii IMSNU 11154	CP001758	complement(1160108..1161007)	2,302771855								
Enterococcus faecalis V583	AE016830	complement(2599576..2600499)	3,070362473								
Leuconostoc citreum KM20	DQ489736	complement(1699269..1700183)	3,837953092								
Enterococcus faecalis V583	AE016833	complement(3980..4939)	4,60554371								
Enterococcus faecalis V583	AE016830	74846..75763	5,373134328								
Enterococcus faecalis V583	AE016831	complement(4101..5042)	6,140724947								
Lactobacillus helveticus DPC 4571	CP000517	809259..810068	6,908315565								
Lactobacillus crispatus ST1	FN692037	790550..791374	7,675906183								
Lactobacillus johnsonii NCC 533	AE017198	complement(1225918..1226733)	8,443496802								
Lactobacillus crispatus ST1	FN692037	complement(1923286..1924077)	9,21108742								
Lactobacillus johnsonii NCC 533	AE017198	504477..505115	9,978678038								
Lactobacillus helveticus DPC 4571	CP000517	812746..813546	10,74626866								
Lactobacillus crispatus ST1	FN692037	786878..787687	11,51385928								
Lactobacillus helveticus DPC 4571	CP000517	810313..811098	12,28144989								
Lactobacillus rhamnosus GG	AP011548	complement(523075..523911)	13,04904051								
Lactobacillus rhamnosus GG	FM179322	complement(524051..524887)	13,04904051								
Lactobacillus johnsonii FI9785	FN298497	complement(35300..36109)	13,81663113								
Lactobacillus johnsonii NCC 533	AE017198	complement(32837..33646)	14,58422175								
Lactobacillus crispatus ST1	FN692037	complement(42184..42996)	15,35181237								
Lactobacillus gasseri ATCC 33323	CP000413	complement(1064547..1065365)	16,11940299								
Lactobacillus johnsonii NCC 533	AE017198	493315..494133	16,8869936								
Lactobacillus gasseri ATCC 33323	CP000413	418311..419129	17,65458422								
Lactobacillus crispatus ST1	FN692037	111137..111961	18,42217484								
Lactobacillus acidophilus NCFM	CP000033	complement(1870586..1871404)	19,18976546								
Lactobacillus acidophilus NCFM	CP000033	87843..88658	19,95735608								
Lactobacillus acidophilus NCFM	CP000033	119982..120791	20,7249467								
Lactobacillus crispatus ST1	FN692037	complement(1927799..1928632)	21,49253731								
Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365	CP000412	19634..20371	22,26012793								
Lactobacillus johnsonii FI9785	FN357112	4361..5203	23,02771855								
Lactobacillus crispatus ST1	FN692037	complement(46846..47664)	23,79530917								
Lactobacillus helveticus DPC 4571	CP000517	87686..88504	24,56289979								
Lactobacillus johnsonii NCC 533	AE017198	465795..466616	25,33049041								
Lactobacillus gasseri ATCC 33323	CP000413	1516490..1517296	26,09808102								
Lactobacillus johnsonii NCC 533	AE017198	1614627..1615433	26,86567164								
Lactobacillus crispatus ST1	FN692037	644646..645455	27,63326226								
Lactobacillus acidophilus NCFM	CP000033	849280..850083	28,40085288								
Lactobacillus johnsonii NCC 533	AE017198	complement(24563..25363)	29,1684435								
Lactobacillus johnsonii NCC 533	AE017198	39672..40538	29,93603412								
Lactobacillus gasseri ATCC 33323	CP000413	complement(1866875..1867741)	30,70362473								
Lactobacillus gasseri ATCC 33323	CP000413	complement(1881508..1882374)	31,47121535								
Lactobacillus helveticus DPC 4571	CP000517	1882618..1883487	32,23880597								
Lactobacillus crispatus ST1	FN692037	787684..788433	33,00639659								
Leuconostoc kimchii IMSNU 11154	CP001758	complement(1688246..1688890)	33,77398721								
Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	CP000414	complement(845830..846657)	34,54157783								
Leuconostoc kimchii IMSNU 11154	CP001758	complement(1961377..1962021)	35,30916844								
Leuconostoc citreum KM20	DQ489736	complement(797352..798179)	36,07675906								
Oenococcus oeni PSU-1	CP000411	525885..526682	36,84434968								
Lactobacillus casei BL23	FM177140	2991126..2992145	37,6119403								
Lactobacillus casei str. Zhang	CP001084	2783199..2784218	38,37953092								
Lactobacillus rhamnosus Lc 705	FM179323	2872528..2873547	39,14712154								
Lactobacillus rhamnosus GG	AP011548	2897502..2898521	39,91471215								
Lactobacillus rhamnosus GG	FM179322	2902532..2903551	39,91471215								
Leuconostoc kimchii IMSNU 11154	CP001758	complement(633573..634526)	40,68230277								
Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	CP000414	complement(1884668..1885615)	41,44989339								
Lactobacillus casei BL23	FM177140	complement(803088..804044)	42,21748401								
Lactobacillus casei str. Zhang	CP001084	complement(691468..692424)	42,98507463								
Lactobacillus rhamnosus Lc 705	FM179323	complement(733490..734464)	43,75266525								
Lactobacillus rhamnosus GG	AP011548	complement(737241..738215)	44,52025586								
Lactobacillus rhamnosus GG	FM179322	complement(739800..740774)	44,52025586								
Leuconostoc kimchii IMSNU 11154	CP001758	complement(631351..632283)	45,28784648								
Lactobacillus casei BL23	FM177140	complement(52309..53184)	46,0554371								
Lactobacillus casei str. Zhang	CP001084	complement(54029..54904)	46,82302772								
Lactobacillus rhamnosus Lc 705	FM179323	2745677..2746543	47,59061834								
Lactobacillus casei ATCC 334	CP000423	185807..186700	48,35820896								
Lactobacillus casei str. Zhang	CP001084	204324..205217	49,12579957								
Lactobacillus rhamnosus GG	AP011548	223456..224361	49,89339019								
Lactobacillus rhamnosus GG	FM179322	223557..224462	49,89339019								
Streptococcus pneumoniae CGSP14	CP001033	1246618..1247538	50,66098081								
Streptococcus gallolyticus UCN34	FN597254	1116156..1117019	51,42857143								

Streptococcus gallolyticus UCN34	FN597254	1110288..1111178	52,19616205								
Streptococcus gordonii str. Challis substr. CH1	CP000725	536544..537437	52,96375267								
Streptococcus sanguinis SK36	CP000387	601036..601917	53,73134328								
Lactococcus lactis subsp. cremoris MG1363	AM406671	complement(2389975..2390832)	54,4989339								
Lactococcus lactis subsp. cremoris NZ9000	CP002094	complement(2390791..2391648)	55,26652452								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(2459376..2460233)	56,03411514								
Lactococcus lactis subsp. lactis Il1403	AE005176	complement(2236512..2237369)	56,80170576								
Lactococcus lactis subsp. lactis Il1403	AE005176	1180044..1180904	57,56929638								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(358123..358950)	58,33688699								
Streptococcus pneumoniae JJA	CP000919	155069..155932	59,10447761	SHP/Rgg (group III)	SPJ_0171				complement(155922..155999)	-1	MKKQILTLLKIVAEIIIILPFLTNR
Streptococcus pneumoniae ATCC 700669	FM211187	145157..146020	59,87206823	SHP/Rgg (group III)	SPN23F_01510				complement(146010..146087)	-1	MKKQILTLLKIVAEIIIILPFLTNR
Streptococcus pneumoniae TCH8431/19A	CP001993	410520..411407	60,63965885								
Streptococcus pneumoniae TIGR4	AE005672	139839..140702	61,40724947	SHP/Rgg (group III)	SP_0141				complement(140692..140769)	-1	MKKQILTLLKIVAEIIIILPFLTNL
Streptococcus pneumoniae Taiwan19F-14	CP000921	166632..167495	62,17484009								
Streptococcus pneumoniae R6	AE007317	149499..150362	62,9424307								
Streptococcus pneumoniae D39	CP000410	149499..150362	63,71002132								
Streptococcus pneumoniae G54	CP001015	146138..147001	64,47761194								
Streptococcus pneumoniae 70585	CP000918	190183..191046	65,24520256								
Streptococcus pneumoniae Hungary19A-6	CP000936	222533..223396	66,01279318	SHP/Rgg (group III)	SPH_0253				complement(223386..223466)	-1	MKKRKIQLILLLISEWVIVIPFLTNL
Streptococcus mitis B6	FN568063	complement(2004676..2005539)	66,7803838								
Streptococcus pneumoniae CGSP14	CP001033	142954..143841	67,54797441	SHP/Rgg (group III)	SPCG_0143				complement(143831..143908)	-1	MKKQVLTLLTIVAEIIIFFPFLTNR
Streptococcus pneumoniae P1031	CP000920	175981..176844	68,31556503	SHP/Rgg (group III)	SPP_0211				complement(176834..176911)	-1	MKKQVLTLLTIVADIIIFFPFLTNR
Streptococcus thermophilus LMD-9	CP000419	complement(1207191..1208066)	69,08315565	SHP/Rgg (group III)	ster_1299				1207133..1207213	-1	MKKVIAIFLFIQTVVVIDIIIFPPFG
Streptococcus mutans UA159	AE014133	complement(110824..111684)	69,85074627								
Streptococcus mutans NN2025	AP010655	complement(117701..118492)	70,61833689								
Streptococcus gallolyticus UCN34	FN597254	complement(2250207..2251073)	71,38592751								
Streptococcus gallolyticus UCN34	FN597254	132907..133764	72,15351812								
Streptococcus agalactiae 2603V/R	AE009948	244411..245271	72,92110874								
Streptococcus agalactiae A909	CP000114	253798..254658	73,68869936								
Streptococcus agalactiae NEM316	AL732656	245821..246681	74,45628998								
Streptococcus gallolyticus UCN34	FN597254	909617..910474	75,2238806								
Streptococcus sanguinis SK36	CP000387	779661..780524	75,99147122								
Streptococcus gordonii str. Challis substr. CH1	CP000725	928723..929586	76,75906183								
Streptococcus pneumoniae 70585	CP000918	1082834..1083697	77,52665245								
Streptococcus pneumoniae P1031	CP000920	1023525..1024388	78,29424307								
Streptococcus pneumoniae JJA	CP000919	998108..998971	79,06183369								
Streptococcus pneumoniae ATCC 700669	FM211187	995330..996193	79,82942431								
Streptococcus pneumoniae TIGR4	AE005672	1044704..1045567	80,59701493								
Streptococcus pneumoniae D39	CP000410	1008803..1009666	81,36460554								
Streptococcus pneumoniae G54	CP001015	989826..990689	82,13219616								
Streptococcus pneumoniae R6	AE007317	1001242..1002135	82,89978678								
Streptococcus pneumoniae Taiwan19F-14	CP000921	1090910..1091773	83,6673774								
Streptococcus pneumoniae TCH8431/19A	CP001993	complement(1270071..1270934)	84,43496802								
Streptococcus pneumoniae Hungary19A-6	CP000936	1104555..1105418	85,20255864								
Streptococcus mitis B6	FN568063	complement(1165094..1165957)	85,97014925								
Streptococcus mitis B6	FN568063	446982..447839	86,73773987								
Streptococcus equi subsp. zooepidemicus	FM204884	1531796..1532689	87,50533049								
Streptococcus equi subsp. equi 4047	FM204883	complement(635147..636040)	88,27292111								
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	complement(652760..653653)	89,04051173								
Streptococcus dysgalactiae subsp. equisimilis GGS_124	AP010935	1429753..1430625	89,80810235								
Streptococcus suis 05ZYH33	CP000407	complement(1907807..1908670)	90,57569296	SHP/Rgg (group II)	SSU05_1997	1908778..1908873	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus suis 98HAH33	CP000408	complement(1907181..1908044)	91,34328358	SHP/Rgg (group II)	SSU98_2002	1908152..1908247	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus suis P1/7	AM946016	complement(1819013..1819876)	92,1108742	SHP/Rgg (group II)	SSU1789	1819984..1820079	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus suis SC84	FM252031	complement(1907422..1908285)	92,87846482	SHP/Rgg (group II)	SSUSC84_1811	1908393..1908488	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus suis BM407	FM252032	complement(1957673..1958536)	93,64605544	SHP/Rgg (group II)	SSUBM407_1859	1958644..1958739	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus suis GZ1	CP000837	complement(1852112..1852975)	94,41364606	SHP/Rgg (group II)	SSGZ1_1813	1853083..1853178	-1	MKQNYLIANITIVLILLISILKDIPPIIVIK			
Streptococcus thermophilus LMD-9	CP000419	865021..865878	95,18123667	SHP/Rgg (group II)	ster_0934	complement(864863..864934)	-1	MNKESFLAILLLIFESIIVIAVG			
Streptococcus thermophilus LMG 18311	CP000023	834855..835721	95,94882729	SHP/Rgg (group II)	stu0917	complement(834706..834777)	-1	MNKESFLAILLLIFESIIVIAVG			
Streptococcus thermophilus CNRZ1066	CP000024	840642..841508	96,71641791	SHP/Rgg (group II)	str0917	complement(840493..840564)	-1	MNKESFLAILLLIFESIIVIAVG			
Streptococcus thermophilus LMD-9	CP000419	1434695..1435558	97,48400853	SHP/Rgg (group II)	ster_1530	complement(1434537..1434608)	-1	MKKQKLLLLVVLVCEGIIVILVG			
Streptococcus thermophilus CNRZ1066	CP000024	1397038..1397901	98,25159915	SHP/Rgg (group II)	str1572	complement(1396880..1396951)	-1	MKKQKLLLLVVLVCEGIIVILVG			
Streptococcus thermophilus LMG 18311	CP000023	1389392..1390255	99,01918977	SHP/Rgg (group II)	stu1572	complement(1389234..1389305)	-1	MKKQKLLLLVVLVCEGIIVILVG			
Streptococcus thermophilus LMD-9	CP000419	complement(1261738..1262598)	99,78678038	SHP/Rgg (group II)	ster_1358	1262689..1262760	-1	MKKQILLTLLLVVFEGIIVIVVG			
Streptococcus dysgalactiae subsp. equisimilis GGS_124	AP010935	complement(994781..995668)	100,554371	SHP/Rgg (group II)	SDEG_1013	995759..995830	-1	MKKHGLLTLLLIILESIIVIGIG			
Streptococcus equi subsp. zooepidemicus	FM204884	complement(1381816..1382685)	101,3219616	SHP/Rgg (group II)	SZO_12400	1382767..1382844	-1	MNKRHFLMLLLMVLEEIIVIGVGYL			
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	744995..745801	102,0895522	SHP/Rgg (group II)	Sez_0716	complement(744773..744850)	-1	MNKRHFLMLLLMVLEEIIVIVVGYL			
Streptococcus equi subsp. zooepidemicus	FM204884	complement(1965108..1965971)	102,8571429	SHP/Rgg (group II)	SZO_17730	1966062..1966133	-1	MKKQKLLPILFLLFEGIIIIVVG			
Streptococcus mutans UA159	AE014133	complement(1434153..1435010)	103,6247335	SHP/Rgg (group II)	smu.1509	1435100..1435168	-1	MRNKIFMTLIVVLETIIIIGGG			
Streptococcus mutans NN2025	AP010655	627658..628515	104,3923241	SHP/Rgg (group II)	SmuNN2025_0595	complement(627500..627568)	-1	MRNKIFMTLIVVLETIIIIGGG			
Streptococcus gallolyticus UCN34	FN597254	767436..768308	105,1599147								
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	1861193..1862053	105,9275053								
Streptococcus equi subsp. equi 4047	FM204883	2105808..2106668	106,6950959								
Streptococcus sanguinis SK36	CP000387	complement(608416..609285)	107,4626866								
Streptococcus gordonii str. Challis substr. CH1	CP000725	complement(543960..544829)	108,2302772								
Streptococcus pneumoniae 70585	CP000918	complement(2057839..2058600)	108,9978678								
Streptococcus pneumoniae CGSP14	CP001033	complement(2081974..2082825)	109,7654584								
Streptococcus pneumoniae JJA	CP000919	complement(1996659..1997420)	110,533049								
Streptococcus pneumoniae TCH8431/19A	CP001993	complement(119210..120061)	111,3006397								
Streptococcus pneumoniae Taiwan19F-14	CP000921	complement(1985337..1986098)	112,0682303								
Streptococcus pneumoniae Hungary19A-6	CP000936	complement(2120328..2121179)	112,8358209								
Streptococcus pneumoniae R6	AE007317	complement(1914475..1915326)	113,6034115								
Streptococcus pneumoniae D39	CP000410	complement(1922007..1922858)	114,3710021								
Streptococcus pneumoniae P1031	CP000920	complement(1992008..1992859)	115,1385928								
Streptococcus pneumoniae G54	CP001015	complement(1949193..1950044)	115,9061834								
Streptococcus pneumoniae ATCC 700669	FM211187	complement(2094478..2095329)	116,673774								
Streptococcus mitis B6	FN568063	122124..122975	117,4413646								
Streptococcus gallolyticus UCN34	FN597254	1253444..1254307	118,2089552								
Streptococcus equi subsp. zooepidemicus	FM204884	930106..930972	118,9765458								
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	complement(1181509..1182285)	119,7441365								
Streptococcus equi subsp. equi 4047	FM204883	complement(1329107..1329973)	120,5117271								
Streptococcus pneumoniae CGSP14	CP001033	158833..159606	121,2793177								
Streptococcus pneumoniae D39	CP000410	complement(948318..949172)	122,0469083	SHP/Rgg (group I)	SPD_0939	949253..949324	-1	MKKISKFLPILFLFLVMDIIIIIVGG			

Streptococcus pneumoniae R6	AE007317	complement(940787..941641)	122,8144989	SHP/Rgg (group I)	spr0960	941722..941793	-1	MKKISKFLPILFLVMDIIIIVGG			
Streptococcus pneumoniae JJA	CP000919	complement(946934..947788)	123,5820896	SHP/Rgg (group I)	SPJ_0996	947868..947930	-1	MKKISKFLPILVLVMDIIII			
Streptococcus pneumoniae ATCC 700669	FM211187	complement(944159..945010)	124,3496802	SHP/Rgg (group I)	SPN23F_09790	945090..945152	-1	MKKISKFLPILVLVMDIIII			
Streptococcus thermophilus LMG 18311	CP000023	933089..933943	125,1172708	SHP/Rgg (group I)	stu1044	complement(932936..933007)	-1	MEKVSKILPILILVMDIIIIVGG			
Streptococcus thermophilus CNRZ1066	CP000024	939089..939943	125,8848614	SHP/Rgg (group I)	str1044	complement(938936..939007)	-1	MEKVSKILPILILVMDIIIIVGG			
Streptococcus pneumoniae G54	CP001015	complement(937570..938421)	126,652452	SHP/Rgg (group I)	SPG_0976	938503..938574	-1	MKKISKFFPILMLVMDIIIIVGG			
Streptococcus pyogenes MGAS2096	CP000261	complement(441098..441970)	127,4200426	SHP/Rgg (group I)	MGAS2096_Spy0459	442029..442100	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS9429	CP000259	complement(439188..440060)	128,1876333	SHP/Rgg (group I)	MGAS9429_Spy0439	440119..440190	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS6180	CP000056	complement(438698..439570)	128,9552239	SHP/Rgg (group I)	M28_Spy0428	439629..439700	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS10750	CP000262	complement(450882..451754)	129,7228145	SHP/Rgg (group I)	MGAS10750_Spy0460	451813..451884	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes NZ131	CP000829	complement(448963..449814)	130,4904051	SHP/Rgg (group I)	Spy49_0449c	449894..449965	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes M1 GAS	AE004092	complement(434181..435032)	131,2579957	SHP/Rgg (group I)	Spy0533	435112..435192	-1	MKKISKFLPILILAMDIIIIVGGEVT			
Streptococcus pyogenes MGAS5005	CP000017	complement(436109..436960)	131,2579957	SHP/Rgg (group I)	M5005_Spy_0440	437040..437111	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes SSI-1	BA000034	1474664..1475515	132,0255864	SHP/Rgg (group I)	Sps1476	complement(1474511..1474582)	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS315	AE014074	complement(421228..422079)	132,0255864	SHP/Rgg (group I)	SpyM3_0377	422160..422231	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS8232	AE009949	complement(482726..483577)	132,793177	SHP/Rgg (group I)	SpyM18_0599	483658..483729	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes str. Manfredo	AM295007	1404577..1405428	133,5607676	SHP/Rgg (group I)	SpyM51424	complement(1404425..1404496)	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS10394	CP000003	complement(468709..469695)	134,3283582	SHP/Rgg (group I)	M6_Spy0474	469640..469711	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus pyogenes MGAS10270	CP000260	complement(438202..439131)	135,0959488	SHP/Rgg (group I)	MGAS10270_Spy0441	439134..439205	-1	MKKISKFLPILILAMDIIIIVGG			
Streptococcus agalactiae 2603V/R	AE009948	complement(1510527..1511375)	135,8635394	SHP/Rgg (group I)	SAG1490	1511463..1511531	-1	MKKINKALLFTLIMDILIIVGG			
Streptococcus agalactiae NEM316	AL732656	complement(1621728..1622576)	136,6311301	SHP/Rgg (group I)	gbs1555	1622664..1622732	-1	MKKINKALLFTLIMDILIIVGG			
Streptococcus agalactiae A909	CP000114	complement(1516673..1517521)	137,3987207	SHP/Rgg (group I)	SAK_1520	1517609..1517677	-1	MKKINKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS10750	CP000262	410175..411056	138,1663113	SHP/Rgg (group I)	MGAS10750_Spy0420	complement(410033..410101)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS10394	CP000003	428012..428893	138,9339019	SHP/Rgg (group I)	M6_Spy0433	complement(427870..427938)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS10270	CP000260	401924..402805	139,7014925	SHP/Rgg (group I)	MGAS10270_Spy0408	complement(401782..401850)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes str. Manfredo	AM295007	complement(1445255..1446121)	140,4690832	SHP/Rgg (group I)	SpyM51462	1446210..1446278	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes NZ131	CP000829	412673..413539	141,2366738	SHP/Rgg (group I)	Spy49_0415	complement(412516..412584)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS2096	CP000261	404919..405800	142,0042644	SHP/Rgg (group I)	MGAS2096_Spy0426	complement(404777..404845)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS9429	CP000259	403009..403890	142,771855	SHP/Rgg (group I)	MGAS9429_Spy0406	complement(402867..402935)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS8232	AE009949	439869..440735	143,5394456	SHP/Rgg (group I)	SpyM18_0554	complement(439712..439780)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS6180	CP000056	402517..403398	144,3070362	SHP/Rgg (group I)	M28_Spy0395	complement(402375..402443)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes SSI-1	BA000034	complement(1510945..1511811)	145,0746269	SHP/Rgg (group I)	Sps1508	1511900..1511968	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS315	AE014074	384935..385801	145,0746269	SHP/Rgg (group I)	SpyM3_0346	complement(384778..384846)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes M1 GAS	AE004092	397890..398756	145,8422175	SHP/Rgg (group I)	Spy0496	complement(397733..397801)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus pyogenes MGAS5005	CP000017	399818..400684	145,8422175	SHP/Rgg (group I)	MutR	complement(399661..399729)	-1	MKKVNKALLFTLIMDILIIVGG			
Streptococcus dysgalactiae subsp. equisimilis GGS_124	AP010935	498944..499798	146,6098081	SHP/Rgg (group I)	SDEG_0529	complement(498786..498854)	-1	MKKINKALLLTLIMDILIIVGG			
Streptococcus equi subsp. equi 4047	FM204883	complement(63659..64522)	147,3773987	SHP/Rgg (group I)	SEQ_0051	64565..64636	-1	MMRKSYKLLKLILDIIIIIGLCQ			
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	complement(63585..64448)	148,1449893	SHP/Rgg (group I)	*Sez_0050*	64491..64562	-1	MMRKSYKLLKLILDIIIIIGLCQ			
Streptococcus pneumoniae G54	CP001015	1228560..1229423	148,91258	SHP/Rgg (group I)	SPG_1268	complement(1228444..1228512)	-1	MKKYYQIFLLLFDIIIIIGLYQ			
Streptococcus thermophilus CNRZ1066	CP000024	188916..189779	149,6801706	SHP/Rgg (group I)	str0182	complement(188782..188844)	-1	MKLLKIIVLLTCIYTIVGGV			
Streptococcus thermophilus LMG 18311	CP000023	189057..189773	150,4477612	SHP/Rgg (group I)	stu0182	complement(188776..188838)	-1	MKLLKIIVLLTCIYIIVGGV			
Lactococcus lactis subsp. cremoris NZ9000	CP002094	complement(1786218..1787066)	151,2153518								
Lactococcus lactis subsp. cremoris MG1363	AM406671	complement(1785382..1786230)	151,9829424								
Lactococcus lactis subsp. lactis Il1403	AE005176	762819..763631	152,750533								
Lactococcus lactis subsp. lactis KF147	CP001834	794284..795132	153,5181237								
Lactococcus lactis subsp. lactis Il1403	AE005176	complement(2287402..2288286)	154,2857143								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(2526386..2527270)	155,0533049								
Lactococcus lactis subsp. cremoris MG1363	AM406671	complement(2182296..2183456)	155,8208955								
Lactococcus lactis subsp. lactis Il1403	AE005176	complement(2044002..2045162)	156,5884861								
Lactococcus lactis subsp. cremoris NZ9000	CP002094	complement(2183113..2184273)	157,3560768								
Lactococcus lactis subsp. cremoris SK11	CP000425	complement(2089188..2090249)	158,1236674								
Lactococcus lactis subsp. lactis Il1403	AE005176	1077643..1078500	158,891258								
Lactococcus lactis subsp. lactis KF147	CP001834	1104556..1105413	159,6588486								
Lactococcus lactis subsp. cremoris SK11	CP000425	complement(1326392..1327282)	160,4264392								
Lactococcus lactis subsp. cremoris MG1363	AM406671	1144571..1145401	161,1940299								
Lactococcus lactis subsp. cremoris NZ9000	CP002094	1145405..1146235	161,9616205								
Lactococcus lactis subsp. lactis Il1403	AE005176	complement(1326911..1327741)	162,7292111								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(1416772..1417602)	163,4968017								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(58461..59294)	164,2643923								
Lactococcus lactis subsp. cremoris MG1363	AM406671	743605..744438	165,0319829								
Lactococcus lactis subsp. cremoris NZ9000	CP002094	744437..745270	165,7995736								
Lactococcus lactis subsp. cremoris SK11	CP000425	complement(1727728..1728561)	166,5671642								
Lactococcus lactis subsp. lactis KF147	CP001834	complement(1886133..1886966)	167,3347548								
Lactococcus lactis subsp. lactis Il1403	AE005176	complement(1717007..1717840)	168,1023454								
Streptococcus thermophilus LMG 18311	CP000023	1720872..1721765	168,869936								
Streptococcus thermophilus CNRZ1066	CP000024	1722744..1723637	169,6375267								
Streptococcus thermophilus LMD-9	CP000419	1780410..1781282	170,4051173								
Lactococcus lactis subsp. cremoris MG1363	AM406671	63015..63896	171,1727079								
Lactococcus lactis subsp. cremoris NZ9000	CP002094	63015..63896	171,9402985								
Lactococcus lactis subsp. cremoris SK11	CP000425	complement(2384356..2385195)	172,7078891								
Lactobacillus rhamnosus Lc 705	FM179323	complement(712413..714233)	173,4754797								
Streptococcus mitis B6	FN568063	1785424..1786353	174,2430704								
Streptococcus mitis B6	FN568063	1478943..1479848	175,010661								
Lactobacillus casei ATCC 334	CP000423	complement(2816639..2817487)	175,7782516								
Lactobacillus casei str. Zhang	CP001084	complement(2784385..2785233)	176,5458422								
Lactobacillus casei BL23	FM177140	complement(2992312..2993160)	177,3134328								
Lactobacillus rhamnosus Lc 705	FM179323	complement(2873724..2874572)	178,0810235								
Lactobacillus rhamnosus GG	AP011548	complement(2898699..2899547)	178,8486141								
Lactobacillus rhamnosus GG	FM179322	complement(2903729..2904577)	178,8486141								
Lactobacillus rhamnosus Lc 705	FM179323	2373135..2374031	179,6162047								
Lactobacillus casei BL23	FM177140	2465322..2466218	180,3837953								
Lactobacillus rhamnosus GG	AP011548	complement(2395164..2396039)	181,1513859								
Lactobacillus rhamnosus GG	FM179322	complement(2400098..2400973)	181,1513859								
Lactobacillus casei str. Zhang	CP001084	457025..457897	181,9189765								
Lactobacillus casei ATCC 334	CP000423	429742..430614	182,6865672								
Lactobacillus rhamnosus Lc 705	FM179323	510155..511027	183,4541578								
Lactobacillus rhamnosus Lc 705	FM179323	107054..107935	184,2217484								
Lactobacillus rhamnosus Lc 705	FM179323	2350076..2350960	184,989339								
Lactobacillus rhamnosus GG	AP011548	2384759..2385643	185,7569296								
Lactobacillus rhamnosus GG	FM179322	2389693..2390577	185,7569296								
Streptococcus equi subsp. equi 4047	FM204883	1604986..1605837	186,5245203								
Streptococcus equi subsp. zooepidemicus	FM204884	complement(633177..634028)	187,2921109								

Streptococcus thermophilus CNRZ1066	CP000024	complement(1338715..1339578)	188,0597015								
Streptococcus thermophilus LMG 18311	CP000023	complement(1331070..1331933)	188,8272921								
Streptococcus gordonii str. Challis substr. CH1	CP000725	1401386..1402249	189,5948827								
Lactobacillus crispatus ST1	FN692037	1932361..1933239	190,3624733								
Lactobacillus crispatus ST1	FN692037	414260..415228	191,130064								
Lactobacillus helveticus DPC 4571	CP000517	complement(279635..280501)	191,8976546								
Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842	CR954253	complement(131339..132178)	192,6652452								
Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365	CP000412	complement(126461..127303)	193,4328358								
Lactobacillus crispatus ST1	FN692037	complement(469378..470232)	194,2004264								
Lactobacillus gasseri ATCC 33323	CP000413	1816867..1817742	194,9680171								
Lactobacillus johnsonii FI9785	FN298497	1686110..1686970	195,7356077								
Lactobacillus johnsonii NCC 533	AE017198	1913791..1914654	196,5031983								
Lactobacillus crispatus ST1	FN692037	1991330..1992190	197,2707889								
Lactobacillus helveticus DPC 4571	CP000517	2039727..2040584	198,0383795								
Lactobacillus acidophilus NCFM	CP000033	1946226..1947074	198,8059701								
Lactobacillus casei str. Zhang	CP001084	complement(2785230..2786105)	199,5735608								
Lactobacillus casei BL23	FM177140	complement(2993157..2994032)	200,3411514								
Lactobacillus casei ATCC 334	CP000423	complement(2817484..2818359)	201,108742								
Lactobacillus rhamnosus Lc 705	FM179323	complement(2874569..2875444)	201,8763326								
Lactobacillus rhamnosus GG	AP011548	complement(2899544..2900419)	202,6439232								
Lactobacillus rhamnosus GG	FM179322	complement(2904574..2905449)	202,6439232								
Oenococcus oeni PSU-1	CP000411	184442..185311	203,4115139								
Streptococcus pyogenes MGAS2096	CP000261	1723590..1724432	204,1791045								
Streptococcus pyogenes MGAS9429	CP000259	1699711..1700553	204,9466951								
Streptococcus pyogenes MGAS10394	CP000003	1746325..1747167	205,7142857								
Streptococcus pyogenes MGAS10750	CP000262	1790319..1791161	206,4818763								
Streptococcus pyogenes MGAS6180	CP000056	1734584..1735426	207,249467								
Streptococcus pyogenes SSI-1	BA000034	1757321..1758163	208,0170576								
Streptococcus pyogenes str. Manfredo	AM295007	1689939..1690781	208,7846482								
Streptococcus pyogenes MGAS8232	AE009949	1756248..1757090	209,5522388								
Streptococcus pyogenes NZ131	CP000829	1677441..1678283	210,3198294								
Streptococcus pyogenes MGAS315	AE014074	1763535..1764377	211,08742								
Streptococcus pyogenes MGAS5005	CP000017	1699085..1699927	211,8550107								
Streptococcus pyogenes M1 GAS	AE004092	1700798..1701640	212,6226013								
Streptococcus pyogenes MGAS10270	CP000260	1761756..1762598	213,3901919								
Lactobacillus fermentum IFO 3956	AP008937	1217236..1218177	214,1577825								
Lactobacillus fermentum CECT 5716	CP002033	1218180..1219121	214,9253731								
Pediococcus pentosaceus ATCC 25745	CP000422	complement(1045033..1045878)	215,6929638								
Lactobacillus plantarum JDM1	CP001617	complement(405879..406769)	216,4605544								
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	complement(702455..703351)	217,228145								
Listeria monocytogenes serotype 4b str. F2365	AE017262	complement(708463..709359)	217,9957356								
Listeria monocytogenes HCC23	CP001175	1950347..1951243	218,7633262								
Listeria innocua Clip11262	AL592022	complement(692972..693868)	219,5309168								
Listeria monocytogenes EGD-e	AL591824	complement(699410..700306)	220,2985075								
Listeria monocytogenes 08-5923	CP001604	complement(703636..704532)	221,0660981								
Listeria monocytogenes 08-5578	CP001602	complement(736932..737828)	221,8336887								
Listeria monocytogenes 08-5578	CP001602	367490..368392	222,6012793								
Listeria monocytogenes 08-5923	CP001604	367511..368413	223,3688699								
Listeria monocytogenes EGD-e	AL591824	349326..350228	224,1364606								
Listeria monocytogenes serotype 4b str. F2365	AE017262	362491..363393	224,9040512								
Listeria monocytogenes HCC23	CP001175	complement(2283356..2284258)	225,6716418								
Listeria innocua Clip11262	AL592022	363055..363957	226,4392324								
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	347725..348627	227,206823								
Listeria monocytogenes EGD-e	AL591824	863798..864688	227,9744136								
Listeria monocytogenes 08-5923	CP001604	866480..867370	228,7420043								
Listeria monocytogenes 08-5578	CP001602	899776..900666	229,5095949								
Listeria monocytogenes serotype 4b str. F2365	AE017262	863820..864710	230,2771855								
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	864232..865122	231,0447761								
Listeria monocytogenes HCC23	CP001175	complement(1788495..1789385)	231,8123667								
Listeria monocytogenes HCC23	CP001175	complement(2282321..2283193)	232,5799574								
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	348791..349663	233,347548								
Listeria monocytogenes 08-5578	CP001602	368556..369428	234,1151386								
Listeria monocytogenes 08-5923	CP001604	368577..369449	234,8827292								
Listeria monocytogenes serotype 4b str. F2365	AE017262	363557..364429	235,6503198								
Listeria monocytogenes EGD-e	AL591824	350392..351264	236,4179104								
Listeria innocua Clip11262	AL592022	364120..364959	237,1855011								
Lactobacillus plantarum WCFS1	AL935263	complement(430926..431795)	237,9530917								
Lactobacillus rhamnosus Lc 705	FM179323	696264..697835	238,7206823								
Streptococcus agalactiae 2603V/R	AE009948	complement(2011647..2012723)	239,4882729								
Streptococcus agalactiae NEM316	AL732656	complement(2058333..2059409)	240,2558635								
Streptococcus agalactiae A909	CP000114	complement(1964400..1965476)	241,0234542								
Listeria monocytogenes serotype 4b str. F2365	AE017262	2851210..2852040	241,7910448								
Streptococcus pneumoniae JJA	CP000919	47219..48112	242,5586354								
Streptococcus pneumoniae 70585	CP000918	78419..79312	243,326226								
Streptococcus pneumoniae Hungary19A-6	CP000936	115263..116156	244,0938166								
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	540941..541834	244,8614072								
Streptococcus equi subsp. zooepidemicus	FM204884	complement(1651177..1652070)	245,6289979								
Streptococcus gordonii str. Challis substr. CH1	CP000725	complement(2144235..2145155)	246,3965885	XIP/ComR	SGO_2080				complement(2143991..2144041)	1	MLFATTLLSAFYWWFY
Streptococcus pneumoniae JJA	CP000919	complement(1854728..1855627)	247,1641791								
Streptococcus pneumoniae R6	AE007317	complement(1774530..1775429)	247,9317697								
Streptococcus pneumoniae P1031	CP000920	complement(1849810..1850709)	248,6993603								
Streptococcus pneumoniae TCH8431/19A	CP001993	complement(2066591..2067490)	249,466951								
Streptococcus pneumoniae TIGR4	AE005672	complement(1892864..1893763)	250,2345416								
Streptococcus pneumoniae ATCC 700669	FM211187	complement(1950922..1951821)	251,0021322								
Streptococcus pneumoniae D39	CP000410	complement(1782062..1782961)	251,7697228								
Streptococcus pneumoniae Hungary19A-6	CP000936	complement(1965932..1966831)	252,5373134								
Streptococcus pneumoniae 70585	CP000918	complement(1898132..1899031)	253,3049041								
Streptococcus pneumoniae G54	CP001015	complement(1803693..1804592)	254,0724947								
Streptococcus pneumoniae CGSP14	CP001033	complement(1936660..1937559)	254,8400853								
Streptococcus pneumoniae Taiwan19F-14	CP000921	complement(1843903..1844802)	255,6076759								
Streptococcus gallolyticus UCN34	FN597254	49789..50688	256,3752665	XIP/ComR	GALLO_0038				50765..50812	1	MLNIFSIVITGWWGL
Streptococcus agalactiae NEM316	AL732656	64256..65158	257,1428571	XIP/ComR	gbs0048				65230..65289	1	MTLVIKLVGTLLTMGWWGL
Streptococcus mutans UA159	AE014133	61631..62545	257,9104478	XIP/ComR	smu.61				62613..62666	1	MFSILTSILMGLDWWSL

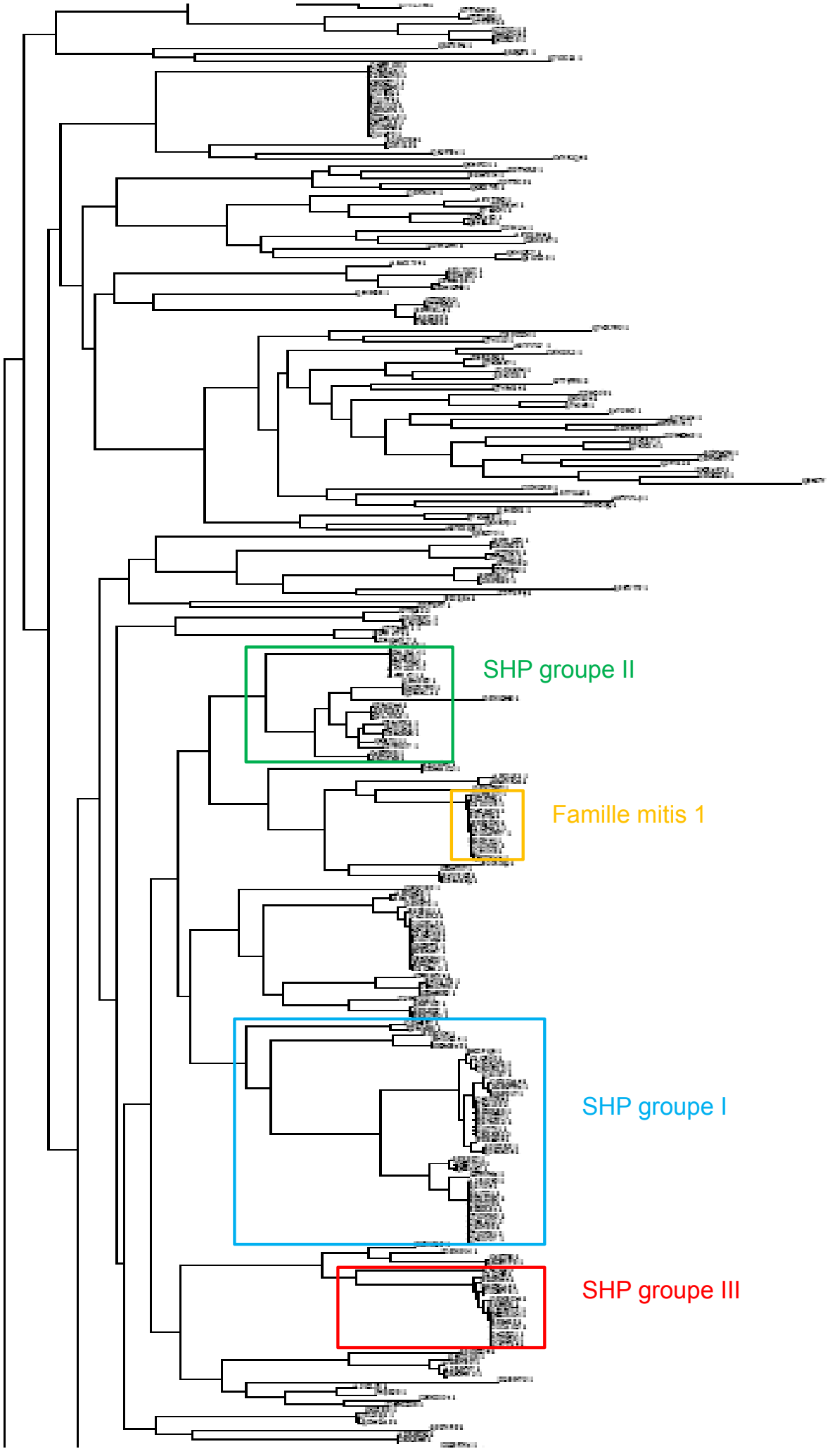
Streptococcus mutans NN2025	AP010655	59969..60883	258,6780384	XIP/ComR	SmuNN2025_0050				60952..61005	1	MFSILTSILMGLDWWSL
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	47940..48854	259,445629	XIP/ComR	Sez_0036				48937..49035	1	MFKKYQYYLFLAALFLLHSAQLLSDIDWWRVG
Streptococcus equi subsp. zooepidemicus	FM204884	49236..50150	260,2132196	XIP/ComR	SZO_00390				50233..50331	1	MFKKYQYYLFLAALFLLHSAQLLSDIDWWRVG
Streptococcus pyogenes M1 GAS	AE004092	52706..53617	260,9808102	XIP/ComR	Spy0037				53701..53796	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS5005	CP000017	53294..54205	260,9808102	XIP/ComR	M5005_Spy_0034				54289..54384	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS10394	CP000003	94131..95042	261,7484009	XIP/ComR	M6_Spy0083				95126..95221	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS9429	CP000259	53074..53985	262,5159915	XIP/ComR	MGAS9429_Spy0035				54069..54164	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS10750	CP000262	54550..55461	263,2835821	XIP/ComR	MGAS10750_Spy0036				55545..55640	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS6180	CP000056	52754..53665	264,0511727	XIP/ComR	M28_Spy0034				53749..53844	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS2096	CP000261	53048..53959	264,8187633	XIP/ComR	MGAS2096_Spy0036				54043..54138	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes NZ131	CP000829	53471..54391	265,5863539	XIP/ComR	Spy49_0032				54475..54570	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus pyogenes MGAS8232	AE009949	53359..54270	266,3539446	XIP/ComR	spyM18_0038				54355..54450	1	MLKKYKYYFIFAALLSFKVVQELSAVDWWRL
Streptococcus dysgalactiae subsp. equisimilis GGS_124	AP010935	49951..50862	267,1215352	XIP/ComR	SDEG_0056				50949..51044	1	MFKRYHYYFILTAMLAFKAAQMISQVDWWRL
Streptococcus pyogenes MGAS10270	CP000260	53452..54363	267,8891258	XIP/ComR	MGAS10270_Spy0035				54454..54552	1	MLKKVKPFLLLAAVVAFKVARVMHEFDWWNLG
Streptococcus pyogenes str. Manfredo	AM295007	52930..53841	268,6567164	XIP/ComR	SpyM50034				53932..54030	1	MLKKVKPFLLLAAVVAFKVARVMHEFDWWNLG
Streptococcus pyogenes SSI-1	BA000034	47473..48384	269,424307	XIP/ComR	SPs0032				48475..48573	1	MLKKVKPFLLLAAVVAFKVARVMHEFDWWNLG
Streptococcus pyogenes MGAS315	AE014074	53474..54385	269,424307	XIP/ComR	SpyM3_0031				54476..54574	1	MLKKVKPFLLLAAVVAFKVARVMHEFDWWNLG
Streptococcus uberis 0140J	AM946015	62014..62949	270,1918977	XIP/ComR	SUB0057					1	MFKKIHFYVTTFSFLAVALITFLSEKDWWHIG
Streptococcus agalactiae 2603V/R	AE009948	64023..64934	270,9594883	XIP/ComR	SAG0048				65003..65101	1	MIKKCFLIVLLVLLSNLGSFDLFLGMGWWNMG
Streptococcus agalactiae A909	CP000114	63833..64741	271,7270789	XIP/ComR	SAK_0081				64823..64873	1	MFKVFFTVMTGVFWWG
Streptococcus suis GZ1	CP000837	54736..55650	272,4946695								
Streptococcus suis SC84	FM252031	55414..56313	273,2622601								
Streptococcus suis BM407	FM252032	55418..56323	274,0298507								
Streptococcus suis 98HAH33	CP000408	55395..56207	274,7974414								
Streptococcus suis 05ZYH33	CP000407	55400..56314	275,565032								
Streptococcus suis P1/7	AM946016	55420..56319	276,3326226								
Streptococcus gordonii str. Challis substr. CH1	CP000725	complement(1871493..1872410)	277,1002132								
Streptococcus gordonii str. Challis substr. CH1	CP000725	complement(1825700..1826623)	277,8678038								
Streptococcus thermophilus CNRZ1066	CP000024	270119..271018	278,6353945	XIP/ComR	str0270				271107..271181	1	MKTLKIFVLFSLLIAILPYFAGCL
Streptococcus thermophilus LMD-9	CP000419	272013..272912	279,4029851	XIP/ComR	ster_0316				273000..273074	1	MKTLKIFVLFSLLIAILPYFAGCL
Streptococcus thermophilus LMG 18311	CP000023	270090..270989	280,1705757	XIP/ComR	stu0270				271077..271151	1	MKTLKIFVLFSLLIAILPYFAGCL
Streptococcus mutans UA159	AE014133	complement(359781..360683)	280,9381663								
Streptococcus mutans NN2025	AP010655	complement(1759003..1759899)	281,7057569								
Lactobacillus casei ATCC 334	CP000423	complement(1529774..1530604)	282,4733475								
Lactobacillus casei BL23	FM177140	complement(1692436..1693266)	283,2409382								
Pediococcus pentosaceus ATCC 25745	CP000422	complement(309719..310543)	284,0085288								
Leuconostoc kimchii IMSNU 11154	CP001758	857108..857947	284,7761194								
Enterococcus faecalis V583	AE016830	1283601..1284428	285,54371								
Lactobacillus plantarum WCFS1	AL935263	complement(2117517..2118413)	286,3113006								
Lactobacillus plantarum JDM1	CP001617	complement(2084003..2084899)	287,0788913								
Lactobacillus casei str. Zhang	CP001084	1143696..1144583	287,8464819								
Lactobacillus casei BL23	FM177140	1345900..1346787	288,6140725								
Lactobacillus casei ATCC 334	CP000423	1171686..1172573	289,3816631								
Lactobacillus rhamnosus Lc 705	FM179323	1209875..1210762	290,1492537								
Lactobacillus rhamnosus GG	AP011548	1195853..1196740	290,9168443								
Lactobacillus rhamnosus GG	FM179322	1199142..1200029	290,9168443								
Lactobacillus sakei subsp. sakei 23K	CR936503	822516..823370	291,684435								
Lactobacillus reuteri F275	CP000705	531685..532530	292,4520256								
Lactobacillus reuteri F275	AP007281	539987..540832	293,2196162								
Lactobacillus fermentum CECT 5716	CP002033	532440..533285	293,9872068								
Lactobacillus fermentum IFO 3956	AP008937	531868..532758	294,7547974								
Lactobacillus brevis ATCC 367	CP000416	complement(1240172..1241056)	295,5223881								
Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365	CP000412	1601662..1602522	296,2899787								
Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842	CR954253	1584349..1585209	297,0575693								
Lactobacillus crispatus ST1	FN692037	1686716..1687576	297,8251599								
Lactobacillus acidophilus NCFM	CP000033	1638552..1639409	298,5927505								
Lactobacillus johnsonii NCC 533	AE017198	1689920..1690762	299,3603412								
Lactobacillus johnsonii FI9785	FN298497	1499760..1500602	300,1279318								
Lactobacillus plantarum JDM1	CP001617	1337370..1338248	300,8955224								
Lactobacillus plantarum WCFS1	AL935263	1357506..1358384	301,663113								
Lactobacillus fermentum IFO 3956	AP008937	complement(1248002..1248889)	302,4307036								
Lactobacillus fermentum CECT 5716	CP002033	complement(1248970..1249857)	303,1982942								
Lactobacillus reuteri F275	AP007281	961694..962584	303,9658849								
Lactobacillus reuteri F275	CP000705	952334..953224	304,7334755								
Lactobacillus plantarum JDM1	CP001617	complement(199144..199986)	305,5010661								
Lactobacillus plantarum WCFS1	AL935263	complement(182176..183018)	306,2686567								
Lactobacillus brevis ATCC 367	CP000416	complement(1883575..1884429)	307,0362473								
Enterococcus faecalis V583	AE016830	1191552..1192436	307,803838								
Leuconostoc kimchii IMSNU 11154	CP001758	1369495..1370355	308,5714286								
Leuconostoc kimchii IMSNU 11154	CP001758	912862..913728	309,3390192								
Leuconostoc citreum KM20	DQ489736	221056..221922	310,1066098								
Enterococcus faecalis V583	AE016830	complement(1552223..1553068)	310,8742004								
Streptococcus equi subsp. zooepidemicus MGCS10565	CP001129	complement(1959248..1960105)	311,641791								
Streptococcus gallolyticus UCN34	FN597254	complement(2244229..2245089)	312,4093817								
Streptococcus suis SC84	FM252031	230211..231068	313,1769723								
Streptococcus suis BM407	FM252032	232477..233334	313,9445629								
Streptococcus suis GZ1	CP000837	231770..232648	314,7121535								
Streptococcus suis P1/7	AM946016	230403..231260	315,4797441								
Streptococcus pneumoniae G54	CP001015	1209724..1210587	316,2473348								
Streptococcus pneumoniae Hungary19A-6	CP000936	complement(1929197..1930060)	317,0149254								
Streptococcus pneumoniae G54	CP001015	complement(1766952..1767815)	317,782516								
Streptococcus mitis B6	FN568063	1352101..1352964	318,5501066								
Streptococcus pneumoniae Taiwan19F-14	CP000921	complement(1783870..1784733)	319,3176972								
Streptococcus pneumoniae 70585	CP000918	complement(1861929..1862792)	320,0852878								
Streptococcus pneumoniae P1031	CP000920	complement(1811904..1812767)	320,8528785								
Streptococcus pneumoniae ATCC 700669	FM211187	complement(1906673..1907536)	321,6204691								
Streptococcus pneumoniae D39	CP000410	complement(1738236..1739099)	322,3880597								
Streptococcus pneumoniae TCH8431/19A	CP001993	complement(2003708..2004571)	323,1556503								
Streptococcus pneumoniae CGSP14	CP001033	complement(1900165..1901028)	323,9232409								
Streptococcus pneumoniae JJA	CP000919	complement(1810390..1811253)	324,6908316								
Streptococcus pneumoniae TIGR4	AE005672	complement(1849039..1849902)	325,4584222								
Streptococcus pneumoniae R6	AE007317	complement(1730704..1731567)	326,2260128								
Streptococcus dysgalactiae subsp. equisimilis GGS_124	AP010935	1287230..1288102	326,9936034								

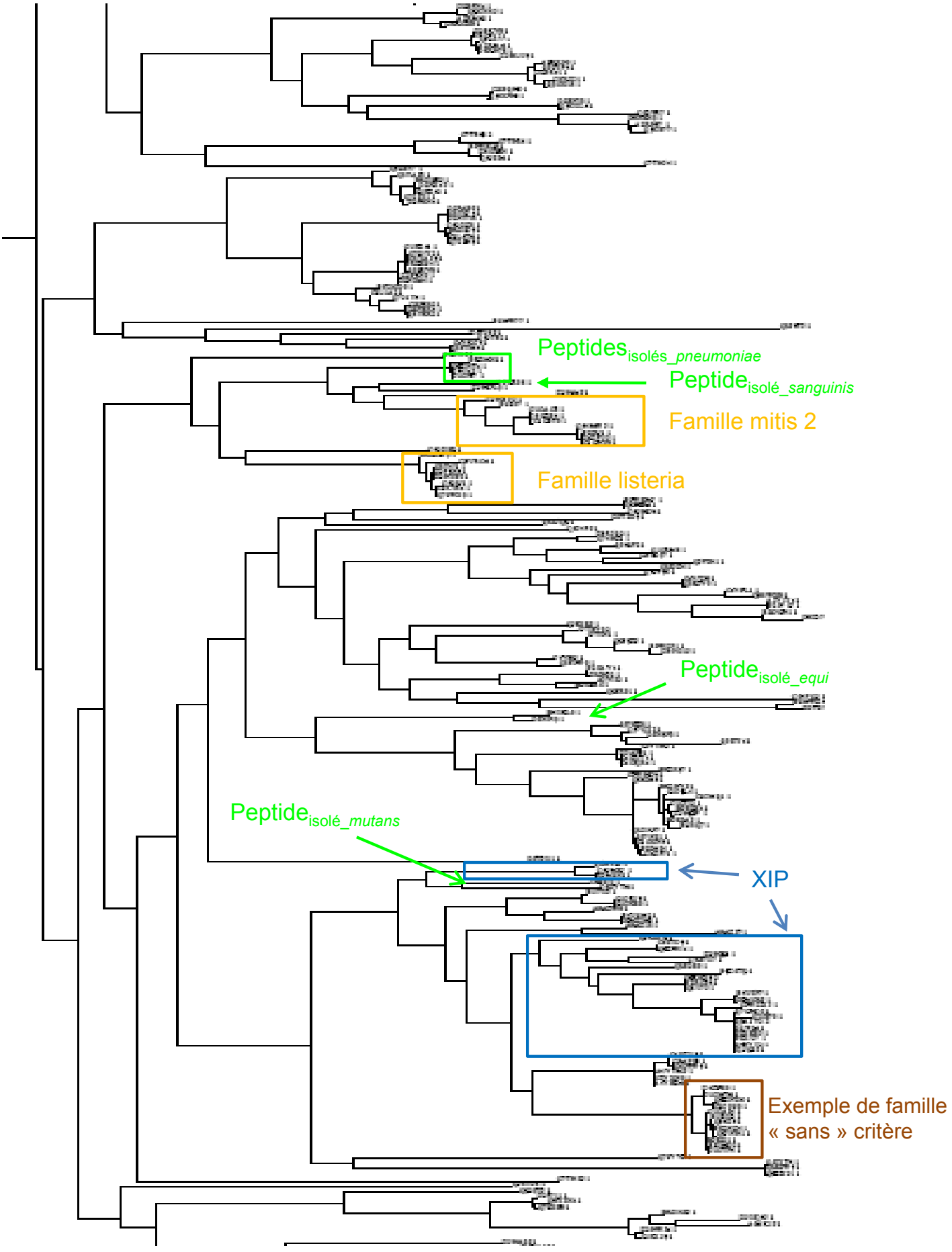
Streptococcus pneumoniae ATCC 700669	FM211187	1243602..1244462	327,761194									
Streptococcus mitis B6	FN568063	complement(1263076..1263942)	328,5287846									
Streptococcus thermophilus LMD-9	CP000419	1582839..1583708	329,2963753									
Streptococcus pneumoniae TIGR4	AE005672	993954..994820	330,0639659									
Streptococcus pneumoniae TIGR4	AE005672	159192..160058	330,8315565									
Oenococcus oeni PSU-1	CP000411	743290..744138	331,5991471									
Listeria seeligeri serovar 1/2b str. SLCC3954	FN557490	complement(2127778..2129070)	332,3667377									
Listeria welshimeri serovar 6b str. SLCC5334	AM263198	complement(2137380..2138660)	333,1343284									
Listeria monocytogenes HCC23	CP001175	474380..475651	333,901919									
Listeria monocytogenes 08-5578	CP001602	complement(2281292..2282563)	334,6695096									
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	complement(2163741..2165012)	335,4371002									
Listeria monocytogenes serotype 4b str. F2365	AE017262	complement(2152623..2153894)	336,2046908									
Listeria monocytogenes EGD-e	AL591824	complement(2164106..2165377)	336,9722814									
Listeria monocytogenes 08-5923	CP001604	complement(2248023..2249294)	337,7398721									
Listeria innocua Clip11262	AL592022	complement(2208964..2210244)	338,5074627									
Lactobacillus sakei subsp. sakei 23K	CR936503	complement(597576..598355)	339,2750533									
Listeria monocytogenes serotype 4b str. CLIP 80459	FM242711	2491502..2492410	340,0426439									
Listeria monocytogenes serotype 4b str. F2365	AE017262	2473702..2474610	340,8102345									
Listeria innocua Clip11262	AL592022	2566232..2567140	341,5778252									
Listeria monocytogenes 08-5578	CP001602	2631327..2632235	342,3454158									
Listeria monocytogenes 08-5923	CP001604	2598375..2599283	343,1130064									
Listeria monocytogenes EGD-e	AL591824	2518971..2519879	343,880597									
Listeria monocytogenes HCC23	CP001175	complement(150296..151204)	344,6481876									
Listeria seeligeri serovar 1/2b str. SLCC3954	FN557490	2412855..2413766	345,4157783									
Listeria welshimeri serovar 6b str. SLCC5334	AM263198	1251844..1252740	346,1833689									
Listeria innocua Clip11262	AL592022	complement(899647..900486)	346,9509595									
Streptococcus pneumoniae JJA	CP000919	1525715..1526566	347,7185501									
Streptococcus pneumoniae R6	AE007317	1532170..1533021	348,4861407									
Streptococcus pneumoniae D39	CP000410	1539701..1540552	349,2537313									
Streptococcus pneumoniae ATCC 700669	FM211187	1650504..1651355	350,021322									
Streptococcus pneumoniae R6	AE007317	complement(112225..113124)	350,7889126									
Streptococcus pneumoniae JJA	CP000919	complement(123880..124743)	351,5565032									
Streptococcus pneumoniae 70585	CP000918	complement(154216..155079)	352,3240938									
Streptococcus pneumoniae TCH8431/19A	CP001993	complement(378205..379068)	353,0916844									
Streptococcus pneumoniae Taiwan19F-14	CP000921	complement(134293..135156)	353,8592751									
Streptococcus pneumoniae CGSP14	CP001033	complement(110309..111208)	354,6268657									
Streptococcus pneumoniae D39	CP000410	complement(112225..113088)	355,3944563									
Streptococcus pneumoniae Hungary19A-6	CP000936	complement(182199..183062)	356,1620469									
Streptococcus mitis B6	FN568063	2029986..2030849	356,9296375									
Streptococcus mitis B6	FN568063	complement(55570..56445)	357,6972281									
Streptococcus gallolyticus UCN34	FN597254	2280842..2281696	358,4648188									
Streptococcus sanguinis SK36	CP000387	2247209..2248057	359,2324094									

Annexe 3

Arbre phylogénétique des protéines Rgg et Rgg-like.

Localisation des familles ou peptides « souche spécifique » présentés dans le chapitre 1 des résultats.


SHP groupe II
Famille mitis 1
SHP groupe I
SHP groupe III


Peptides isolés_pneumoniae
Peptide isolé_sanguinis
Famille mitis 2
Famille listeria
Peptide isolé_equi
Peptide isolé_mutans
XIP
Exemple de famille « sans » critère

Annexe 4

Tableau non exhaustif de spCDS bornant des protéines régulatrices Rgg et Rgg-like.

Tableau non exhaustif de spCDS bornant des protéines régulatrices Rgg et Rgg-like.

Chaque ligne au niveau des souches est représentative d'une protéine Rgg / Rgg-like. Aussi, dans le cas de *Streptococcus agalactiae* NEM316, la même protéine Rgg-like est bornée par 7 peptides amont et 1 peptide aval.

	Peptides SHOW amont		**Peptides SHOW aval**	
Streptococcus thermophilus				
LMD-9	1	MIIQILNSLPSDSFEFKDRLRITIYS		
	1	MTTQNLNKLPGTIAVPMDKIRITIYCPSF		
LMG 18311	1	MIIQILNSLPSDSFEFKDRLRITIYS		
	1	MTTQNLNKLPGTIAVPMDKIRITIYCPSF		
CNRZ1066	1	MIIQILNSLPSDSFEFKDRLRITIYS		
	1	MTTQNLNKLPGTIAVPMDKIRITIYCPSF		
Streptococcus agalactiae				
NEM316	1	MILQIKENDYISAHILVKVGKYKLLKLPHKFTNEITVRKEY	1	MGFFNISPFGLL
	1	MVSTQYMKVTLREKFPDKLTYVDLEELVI		
	-1	MTLGQFLNITQKTPHGVNHEVLPIADI		
	1	MKDNKVYYLHFGKAIDKNEAKNDKYLNQIINN		
	1	MTKLDEMILF		
	1	MLRKLTILIVIIDMKIIKLD		
	1	MNYKRAFLKGSFIIHYFSMTNISEVNFYLTIVIINAVLPLINQRKSFF		
2603V/R	1	MILQIKENDYISAHILVKVGKYKLLKLPHKFTNEITVRKEY	1	MGFFNISPFGLL
	1	MEEFCMVSTQYMKVTLREKFPDKLTYVDLEELVI		
	-1	MTLGQFLNITQKTPHGVNHEVLPIADI		
	1	MKDNKVYYLHFGKAIDKNEAKNDKYLNQIINN		
	1	MLRKLTILIVIIDMKIIKLD		
	1	MNYKRAFLKGSFIIHYFSMTNISEVNFYLTIVIINAVLPLINQRKSFF		
A909	1	MILQIKEKDYISAHILVKVGKYKLLKLPHKFTNEITVRKEY	1	MGFFNISPFGLL
	1	MEEFCMVSTQYMKVTLREKFPDKLTYVDLEELVI		
	-1	MTLGQFLNITQKTPHGVNHEVLPIADI		
	1	MKDNKVYYLHFGKAIDKNEAKNDKYLNQIINN		
	1	MTKLDEMILF		
	1	MLRKLTILIVIIDMKIIKLD		
	1	MNYKKAFLKGSFIIHYFSMTNISEVNFYLTIVIINAVLPLINQRKSFF		

Annexe 5

Tableau exhaustif des protéines régulatrices Rgg et Rgg-like associées à des peptides présentant un motif GG.

Tableau exhaustif des protéines régulatrices Rgg et Rgg-like associées à des peptides présentant un motif GG (page 1/2).

Ce tableau est complémentaire des résultats présentés dans le chapitre 1, le peptide de *Streptococcus mutans* n'est donc pas représenté.

	Peptides SHOW amont		Peptides SHOW aval	
Streptococcus pneumoniae				
R6			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
D39			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
TIGR4			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
			1	MRSLFRKIVALLVIGLILLGTAGGTQVHKMARGIDPGPANGIYR
JJA			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
G54			-1	MSRTSRSETDSESHNDKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
CGSP14			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLMFK
			-1	MKIRQTGGCFVSHLTERDGLRVT
ATCC 700669			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
70585			1	MGGTLFVLLSIGIWFYL
			-1	MSRTSRSETDSESHNYKADSINLTISLTGYLMFK
			-1	MKIRQTGGCFVSHLTERDGLRVT
Hungary19A-6			1	MGGTLFVLLSIGIWFYL
			-1	MSRTSRSETDSESHNDKADSINLTISLTGYLTFQKNYNVKIN
			-1	MKIRQTGGCFVSHLTERDGLRVT
Streptococcus equi				
subsp. zooepidemicus	-1	MVVQEIREHDFDQELDLTFKNLVCGGNDWCMRMFRLYFPRERYI		
subsp. equi 4047	-1	MVVQEIREHDFDQELDLTFKNLVCGGNDWCMRMFRLYFPRERYI		
Streptococcus suis				
05ZYH33	1	MKTGGLYTSYKLFLSSDV	1	MFNYLKFFGRLGGNWGTWVEE
			1	MKNNIIKKGAGKTSMMAPPSPTE
98HAH33	1	MKTGGLYTSYKLFLSSDV	1	MFNYLKFFGRLGGNWGTWVEE
			1	MKNNIIKKGAGKTSMMAPPSPTE
P1/7	1	MKTGGLYTSYKLFLSSDV	1	MFNYLKFFGRLGGNWGTWVEE
			1	MKNNIIKKGAGKTSMMAPPSPTE
GZ1	1	MKTGGLYTSYKLFLSSDV	1	MFNYLKFFGRLGGNWGTWVEE
			1	MKNNIIKKGAGKTSMMAPPSPTE

Tableau exhaustif des protéines régulatrices Rgg et Rgg-like associées à des peptides présentant un motif GG (page 2/2).

	Peptides SHOW amont	Peptides SHOW aval
Streptococcus suis (suite)		
SC84	1 MKTGGLYTSYKLFLSSDV	1 MFNYLKFFGRLGGNWGTWVEE 1 MKNNIIKKGAGKTSMMAPPSPTE
BM407	1 MKTGGLYTSYKLFLSSDV	1 MFNYLKFFGRLGGNWGTWVEE 1 MKNNIIKKGAGKTSMMAPPSPTE
Streptococcus sanguinis		
SK36	-1 MKKIVYNLILLAVTSIVTTSVFPWWWLWW -1 MYFHGGGYGGKVNRIGKGRVQWGR	
Listeria		
welshimeri serovar 6b str. SLCC5334	-1 MISMISLFSGLLGVSVRVGG	
Lactobacillus		
rhamnosus Lc 705	-1 MMMTVLLSILVPLGGRVIAGG	

Annexe 6

Tableaux des gènes cibles putatifs des systèmes SHP/Rgg, identifiés par analyse *in silico.*

Tableaux des gènes cibles putatifs des systèmes SHP/Rgg, identifiés par analyse *in silico*

- **Tableau A :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe I.
- **Tableau B :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe II.
- **Tableau C :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe III.

La colonne « nM » indique nombre de mismatches détectés entre le motif recherché et celui de la cible.

Dans la colonne brin, « c » indique que le motif est situé sur le brin complémentaire et « w » sur le brin codant.

La colonne « Dir » indique la direction du motif, c'est-à-dire soit dans la même direction que le motif recherché « S » (same) ou soit dans la direction opposée « O » (opposite).

La colonne « dATG » indique la distance en nucléotides entre le codon start du gène cible « locus_tag » (ou de la spCDS quand détectée) et le motif. Lorsque cette dernière est mise en gras, ceci indique que le gène cible est considéré comme proximal.

Les colonnes « Gène » et « Fonction » indiquent respectivement l'annonation GenBank du gène et sa fonction.

Dans la colonne localisation, « C » signifie « complément » et indique que la spCDS est codée sur le brin inverse.

La séquence du motif recherché est indiquée en couleur, pour chaque locus, en haut de la colonne. La distance « motif / codon start du gène shp » est indiquée dans la colonne dATG. Les couleurs bleu, vert et rouge font respectivement référence au groupe I, II et III.

NA indique des erreurs lors de la production de données n'ayant pas pu être vérifiées pour l'écriture de ce manuscrit.

Légende :

Biosynthèse d'ADN et d'ARN	Régulateurs transcriptionnels
Biosynthèse du mevalonate	Réplication / Recombinaison / Réparation de l'ADN / Maintenance cellulaire
Biosynthèse du peptidoglycane / paroi cellulaire	Traduction
Division cellulaire / ségrégation du chromosome	Transporteurs / Permeases
Métabolisme des acides aminés	Transport des métaux, vitamines et ions
Métabolisme glucidique	Transport des sucres
Métabolisme lipidique	Transposition
Methyltransférase	Hyp. protein
Protéines secrétées et surface / facteurs de virulence / pathogénicité	Autres

Annexe 6 - Tableau A : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe I.

Strains	Gène *rgg*	Cibles nM	Brin	Motif cible	Dir	dATG	Locus_tag	Gène	Fonction	spCDS identifiées dans bactgeneSHOW Localisation	Séquence
S. thermophilus											
				CCCATTTCCCCAAC		**75**					
LMG18311	*stu1044*	2	c	CCCATTTtCtCAAC	O	35	*stu0091*	-	Hyp. Protein		
		3	c	CCCATTTttCaAAC	O	187	*stu0416*	-	Hyp. Protein		
		3	w	CCCATcTtCaCAAC	S	58	*stu0537*	*rplL*	50S ribosomal protein		
CNRZ1066	*str1044*	2	c	CCCATTTtCtCAAC	O	9	*str0091*	-	Hyp. Protein	99651..99725	MGKVIPMLKCSLKLVNTLESQLMT
		3	w	CCCATcTtCaCAAC	S	58	*str0537*	*rplL*	50S ribosomal protein		
		3	w	CCCATaTtCCCAAg	O	93	*str0652*	*mip*	Macrophage infectivity potentiator-related protein	C(625421..625795)	MYSYLNLESTKTSFSGGKDSGVLYSYLNLESTKTIFSKTFSIPKLYSYLNLESTK TYVEDNQLDVVLYSYLNLESTKTYIVGCVPYNLLYSYLNLESTKTKVERYRKEIK LYSYLNLESTKTGS
				CCCATATATGCAAC		**76**					
LMG18311	*stu0182*	2	w	CCCgTtTATGCAAC	S	148	*stu0183*	-	Hyp. Protein	189923..189964	MLFTHDQLNALKS
		3	w	CCCAcATAaGaAAC	S	282	*stu0656*	-	Hyp. Protein	622957..623043	MVPPYDGVHLYANAYLKVYEALEPYLLD
		3	w	CCaATATgTtCAAC	O	127	*stu1251*	*glmM*	Phosphoglucosamine mutase		
		1	w	CCCATATATGCAAa	O	**74**	*stu1570*	*shp*		C(1389234..1389305)	MKKQKLLLLVVLVCEGIIVILVG
CNRZ1066	*str0182*	2	w	CCCgTtTATGCAAC	S	148	*str0183*	-	Hyp. Protein	189874..189963	MKNFQALLKVTKSSGVKSDAFYARSIKCP
		3	w	CCaATATgTtCAAC	O	126	*str1251*	*glmM*	Phosphoglucosamine mutase		
		1	w	CCCATATATGCAAa	O	**74**	*str1570*	*shp*		C(1396880..1396951)	MKKQKLLLLVVLVCEGIIVILVG
S. agalactiae											
				CCCACTTTCACAAC		**85**					
A909	*sak1520*	3	w	CCaACTTcCtCAAC	O	263	*sak_0510*	-	Cupin domain protein	C(454032..454613)	MATGYHAAVTAEVKEGDTVVVMGDGAVGLCGVIAAKMLGANCIIAMSRHKDRQEL ALTFGATDIVEERGDEAVKRVLDLTNQAGADAILECVGTEQSVDTATQIARPGAV IGRVGIPQNPDMNTNNLFWKNIGLRGGIASVTTFDKSVLLDAVLTHKINPGLVFT KSFVLDDIQKAYEAMDKRDAIKSLVIVD
S. dysgalactiae											
				CCCACTTTCACAAC		**85**					
subsp. equisimilis GGS_124	*sdeg_0529*	3	c	CCatCTTTtACAAC	O	224	*sdeg_0303*	-	Hyp. Protein	272678..273544	MLGRDNAKVSWIRNFEYPDKWDIELLQSENNLPKLMLTEILGSSKPLLFCEGDDK SSLDYHIYKSLFGEKYTIIPSAGHKQVVNNVRAVKQIENIQDAYGIIDLDSLTED EKNKFENERVKVLEFNEIEMLFFEEHVMIEVMRNIYSVEYEEKIKIFKLKLHDII RNKSEKIALAYVKKRVENYLMNEKIQDYSNLEQITDGLKGISNYDISTQYEEKKK QIKTIIDDNAYPELLKVCNLKTEISKGLANRYLDNNYIEKAKQKILTDIELREYL VKKYFSFLGEVQV
		3	w	CCCctTTTCACAAg	O	21	*sdeg_1173*	-	Hyp. Protein		
		2	c	CCCttTTTCACAAC	S	76	*sdeg_2024*	-	Hyp. Protein	C(1959149..1959214)	MQFRETEVCAATEVNICIVSK
S. pyogenes											
				CCCACTTTCACAAC		**84**					
M1 GAS	*spy_0496*	1	w	CCCACTTTCcCAAC	S	**74**	*spy_0534*	*shp*		435112..435192	MKKISKFLPILILAMDIIIIVGGEVT
		3	c	CCCtCTTTCtCAAg	O	95	*spy_1053*	*lep*	Elongation factor (GTP-binding protein)	861369..861566	MEQTFFMIKPYGVKQGEVVGEVLRWIERLRFTFKRFELRQASSKYLAKHDEALVI NLLILNLKLT
		1	w	CCCACTTTCcCAAC	O	19	*NA*				
str. Manfredo	*spyM51462*	3	c	CCCAtTTTaACAcC	S	161	*spym50103*	*rofA*	RALP-type regulators		
		3	c	CCCAtTTTaACAcC	O	226	*spym50104*	-	Putative surface-anchored protein	119308..119382	MIDQKKTLSYFSLEGGRGYIAKEG
		3	w	CCCctTTTCACAAg	O	21	*spym51096*	-	Hyp. Protein		
		3	w	CCaACTTTCAaAAa	O	138	*spym51370*	-	Hyp. Protein		
		1	c	CCCACTTTCcCAAC	S	**74**	*spym51423*	*shp*		C(1404425..1404496)	MKKISKFLPILILAMDIIIIVGG
		3	w	CCCtCTTTCtCAAg	O	92	*NA*				
MGAS8232	*spyM18_0554*	3	c	CCCAtTTTaACAcC	S	161	*spym18_0125*	*nra*	RALP-type regulators		
		3	c	CCCAtTTTaACAcC	O	282	*spym18_0126*	*cbp*	Putative collagen binding protein		
		3	c	CCCcaTTTaACAAC	O	8	*spym18_0370*	-	Hyp. phage protein		
		1	w	CCCACTTTCcCAAC	S	**74**	*spym18_0600*	*shp*		483658..483729	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCAaAAa	O	138	*spym18_0661*	-	Hyp. Protein		
SSI-1	*sps1508*	3	c	CCCAtTTTaACAcC	S	161	*sps0099*	*nra*	RALP-type regulators		
		3	c	CCCAtTTTaACAcC	O	283	*sps0100*	*cbp*	Putative collagen binding protein		
		3	c	CCCtCTTTCtCAAg	O	288	*sps0938*	*lep*	Elongation factor (GTP-binding protein)	919861..919905	MTSGPVLIGTLLGD
		1	c	CCCACTTTCcCAAC	S	**76**	*sps1475*	*shp*		C(1474511..1474582)	MKKISKFLPILILAMDIIIIVGG
MGAS315	*spyM3_0346*	3	c	CCCAtTTTaACAcC	S	161	*spyM3_0097*	*nra*	RALP-type regulators		
		3	c	CCCAtTTTaACAcC	O	282	*spyM3_0098*	*cbp*	Putative collagen binding protein		
		1	w	CCCACTTTCcCAAC	S	**75**	*spyM3_0378*	*shp*		422160..422231	MKKISKFLPILILAMDIIIIVGG

		3	c	CCCcaTTTaACAAC	O	8	*spyM3_0711*	-	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	288	*spyM3_0737*	*lep*	GTP-binding protein	788973..789017	MTSGPVLIGTLLGD
		3	w	CCaACaTTaACAAC	O	27	*spyM3_0865*	-	Putative sugar ABC transporter		
MGAS10394	*M6_spy0433*	3	c	CCCtCTTTCtCAAg	O	95	*M6_spy0794*	-	Nucleoside diphosphate kinase	799344..799397	MEQTFFMIKPDGVKRGG
MGAS5005	*mutR*	1	w	CCCACTTTCcCAAC	S	**74**	*M5005_spy0441*	*shp*		437040..437111	MKKISKFLPILILAMDIIIIVGG
		3	c	CCCtCTTTCtCAAg	S	33	*M5005_spy0773*	-	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	113	*M5005_spy0774*	-	Nucleoside diphosphate kinase		
MGAS6180	*M28_spy0395*	3	c	CCCtCTTTCtCAAg	O	122	*M28_spy0752*	-	Transposase		
MGAS9429	*MGAS9429_spy0406*	3	c	CCCtCTTTCtCAAg	O	113	*MGAS9429_spy0889*	-	Nucleoside diphosphate kinase		
MGAS10270	*MGAS10270_spy0408*	3	c	CCCctTTTCACAAg	O	3	*MGAS10270_spy0770*	-	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	160	*MGAS10270_spy0890*	-	Nucleoside diphosphate kinase		
MGAS2096	*MGAS2096_spy0426*	3	c	CCCtCTTTagCAAC	S	31	*MGAS2096_spy0151*	-	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	113	*MGAS2096_spy0847*	-	Nucleoside diphosphate kinase		
MGAS10750	*MGAS10750_spy0420*	3	c	CCCtCTTTCtCAAg	O	161	*MGAS10750_spy0925*	-	Nucleoside diphosphate kinase		
NZ131	*spy49_0415*	3	c	CCCAtTTTaACAcC	S	161	*spy49_0111*	*nra*	RALP-type regulators		
		3	c	CCCAtTTTaACAcC	O	9	*spy49_0112*	*cbp*	Putative collagen binding protein	118812..118859	MGKKLNEFSIRIAIQ
		1	w	CCCACTTTCcCAAC	S	74	*spy49_0450*	*shp*		449894..449965	MKKISKFLPILILAMDIIIIVGG
		3	c	CCCAtTTTCAaAAa	O	61	*spy49_0501*	-	Hyp. Protein	492085..492162	MSHKKSEKKTIFSTLISNSIDEEET
				CCCACTTTCCCAAC		**74**					
M1 GAS	*spy_0533*	1	c	CCCACTTTCaCAAC	S	**84**	*spy_0492*	*shp*		C(397733..397801)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	95	*spy_1053*	*lep*	Elongation factor (GTP-binding protein)	861369..861566	MEQTFFMIKPYGVKQGEVVGEVLRWIERLRFTFKRFELRQASSKYLAKHDEALVI NLLILNLKLT
str. Manfredo	*spyM51424*	1	w	CCCACTTTCaCAAC	S	**84**	*spym51463*	*shp*		1446210..1446278	MKKVNKALLFTLIMDILIIVGG
		3	w	CCCtCTTTCtCAAg	O	92	*spym50990*	*lep*	Pseudogène		
MGAS8232	*spyM18_0599*	1	c	CCCACTTTCaCAAC	S	**84**	*spym18_0550*	*shp*		C(439712..439780)	MKKVNKALLFTLIMDILIIVGG
		3	w	CCCgCTTcCCCAAg	O	168	*spym18_1418*	-	AGP acyltransferase		
		3	w	CCCgCTTcCCCAAg	S	35	*spym18_1419*	-	Hyp. Protein		
SSI-1	*sps1476*	3	c	CCCtCTTTCtCAAg	O	288	*sps0938*	*lep*	Elongation factor (GTP-binding protein)	919861..919905	MTSGPVLIGTLLGD
		1	w	CCCACTTTCaCAAC	S	**84**	*sps1509*	*shp*		1511900..1511968	MKKVNKALLFTLIMDILIIVGG
MGAS315	*spyM3_0377*	1	c	CCCACTTTCaCAAC	S	**84**	*spyM3_0345*	*shp*		C(384778..384846)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	288	*spyM3_0737*	*lep*	Elongation factor (GTP-binding protein)	788973..789017	MTSGPVLIGTLLGD
MGAS10394	*M6_spy0474*	3	c	CCCACTTTCaCAAC	O	226	*M6_spy0701*	*mvaK2*	Phosphomevalote kise		
		3	c	CCCtCTTTCtCAAg	O	285	*M6_spy0992*	-	Phage infection protein		
MGAS5005	*M5005_spy_0440*	1	c	CCCACTTTCaCAAC	S	**84**	*M5005_spy0406*	*shp*		C(399661..399729)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	S	33	*M5005_spy0773*	-	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	113	*M5005_spy0774*	-	Nucleoside diphosphate kinase		
MGAS6180	*M28_spy0428*	1	c	CCCACTTTCaCAAC	S	**84**	*M28_spy0394*	*shp*		C(402375..402443)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	122	*M28_spy0752*	-	Transposase		
MGAS9429	*MGAS9429_spy0439*	1	c	CCCACTTTCaCAAC	S	**84**	*MGAS9429_spy0405*	*shp*		C(402867..402935)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	113	*MGAS9429_spy0889*	-	Nucleoside diphosphate kinase		
MGAS10270	*MGAS10270_spy0441*	1	c	CCCACTTTCaCAAC	S	**84**	*MGAS10270_spy0407*	*shp*		C(401782..401850)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	160	*MGAS10270_spy0890*	-	Nucleoside diphosphate kinase		
MGAS2096	*MGAS2096_spy0459*	3	c	CCCtCTTTagCAAC	S	31	*MGAS2096_spy0151*	-	Hyp. Protein		
		1	c	CCCACTTTCaCAAC	S	**84**	*MGAS2096_spy0425*	*shp*		C(404777..404845)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	113	*MGAS2096_spy0847*	-	Nucleoside diphosphate kinase		
MGAS10750	*MGAS10750_spy0460*	1	c	CCCACTTTCaCAAC	S	**84**	*MGAS10750_spy0419*	*shp*		C(410033..410101)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCCtCTTTCtCAAg	O	161	*MGAS10750_spy0925*	-	Nucleoside diphosphate kinase		
NZ131	*spy49_0449c*	1	c	CCCACTTTCaCAAC	S	**84**	*spy49_0414c*	*shp*		C(412516..412584)	MKKVNKALLFTLIMDILIIVGG
S. pneumoniae											
				CCCATTTTCCCAAC		**74**					
R6	*spr0960*	3	c	CCCATTTTtCaAAa	O	105	*spr0240*	*uppS*	Undecaprenyl pyrophosphate synthase		
D39	*spd_0939*	3	c	CCCATTTTCtaAAa	O	150	*spd_0114*	-	Hyp. Protein		
		3	c	CCCATTTcCtCAAg	O	730	*spd_0222*	-	Phosphoglycerate mutase family protein	211891..212223	MIAAVQNGSLNLEKLEAMTAICSVGLDMIAIPEDTPAETIAAMIADEAAIGVINM KTTAVRIIPKGREGDMIEFGGLLGTAPVMKVNGASSVDFISRGGQIPAPIHSFKN
		3	c	CCCATTTTtCaAAa	O	105	*spd_0243*	*uppS*	Undecaprenyl pyrophosphate synthase		
JJA	*spj_0996*	3	c	CCCATTTTatCAAa	S	31	*spj_0031*	-	Integrase		
		3	c	CCCATTTTtCaAAa	O	105	*spj_0271*	*uppS*	Undecaprenyl pyrophosphate synthase		
ATCC 700669	*spn23f__09790*	3	c	CCCATTTTatCAAa	S	31	*NA*				
		3	c	CCCATTTTtCaAAa	O	105	*spn23f_02490*	*uppS*	Undecaprenyl pyrophosphate synthase		
		3	c	CCCtTTTTaaCAAC	S	32	*spn23f_15380*	-	Hyp. phage protein		
		3	w	CCCATTTTatCAAa	S	31	*spn23f_15800*	-	Phage integrase		

Annexe 6 - Tableau B : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe II.

Strains	Gène *rgg*	Cibles nM	Brin	Target_motif	Dir	dATG	Locus_tag	Gène	Fonction	spCDS identifiées dans bactgeneSHOW Localisation	Séquence
S. thermophilus											
				GCATATATGGGA		**75**					
LMD-9	*ster_1358*	2	w	GaATATAgGGGA	O	237	*ster_0352*	*rsmG*	Ribosomal RNA small subunit methyltransferase G		
		2	w	GaATATAgGGGA	S	**71**	*ster_0353*	-	Hyp. Protein		
		2	c	GCATATAgGtGA	O	159	*ster_0708*	-	Hyp. Protein		
		2	c	GCATATAcGaGA	S	**70**	*shp_0933*	*shp*			
		1	c	GCAaATATGGGA	S	**77**	*ster_1357*	-	Pep1357C		
		2	c	GtATATATGaGA	O	222	*ster_1379*	*gshAB*	Glutamate-cysteine ligase		
		0	c	GCATATATGGGA	S	**71**	*shp1530*	*shp*			
		0	w	GCATATATGGGA	S	113	*ster_1531*	-	Major facilitator superfamily permease		
		2	c	GaATATATGcGA	S	**77**	*ster_1774*	-	Deoxyribonuclease, putative		
		2	c	GCtTATAaGGGA	O	31	*ster_1923*	-	Hyp. Protein	1785854..1786048	MKYGKIFKKFRESRGLSLKNVAKSGLSSSHLSRFENDEADLTISNFF SVRCNTYAYRRICLRCT
				GCATATATGGGA		**72**					
LMD-9	*ster_1530*	2	w	GaATATAgGGGA	O	237	*ster_0352*	*rsmG*	Ribosomal RNA small subunit methyltransferase G		
		2	w	GaATATAgGGGA	S	**71**	*ster_0353*	-	Hyp. Protein		
		2	c	GCATATAgGtGA	O	159	*ster_0708*	-	Hyp. Protein		
		2	c	GCATATAcGaGA	S	**70**	*shp_0933*	*shp*			
		1	c	GCAaATATGGGA	S	**77**	*ster_1357*	-	Pep1357C		
		0	w	GCATATATGGGA	S	**74**	*shp1358*	*shp*			
		0	w	GCATATATGGGA	S	113	*ster_1531*	-	Major facilitator superfamily permease		
		2	c	GaATATATGcGA	S	**77**	*ster_1774*	-	Deoxyribonuclease, putative		
		2	c	GtATATATGaGA	O	222	*ster_1379*	*gshAB*	Glutamate-cysteine ligase		
		2	w	GCgTATAaGGGA	S	**74**	*ster_1923*	-	Hyp. Protein		
		2	c	GCtTATAaGGGA	O	31	*ster_1923*	-	Hyp. Protein	1785854..1786048	MKYGKIFKKFRESRGLSLKNVAKSGLSSSHLSRFENDEADLTISNFF SVRCNTYAYRRICLRCT
LMG18311	*stu1572*	0	w	GCATATATGGGA	O	**77**	*shp0182*	*shp*			
		2	w	GCATATAgGtGA	S	139	*stu0234*	-	Hyp. Protein		
		2	w	GaATATAgGGGA	O	237	*stu0315*	*rsmG*	Ribosomal RNA small subunit methyltransferase G		
		2	w	GaATATAgGGGA	S	**65**	*stu0316*	-	Hyp. Protein		
		2	c	GCATATAgGtGA	S	125	*stu0652*	*mip*	Macrophage infectivity potentiator-related protein	C(622711..622866)	MLDKFIRIQTTDASIIDDIVDTSTPSKFTKGKKLHERGDNEEHLPLD DDLI
		2	c	GCATATAgGtGA	O	159	*stu0656*	-	Hyp. Protein		
		2	c	GCATATAcGaGA	S	70	*shp0917*	*shp*			
		2	c	GtATATATGaGA	O	223	*stu1413*	*gshAB*	Glutamate-cysteine ligase		
		0	w	GCATATATGGGA	S	113	*stu1573*	-	Major facilitator superfamily permease		
		2	w	GCgTATAaGGGA	S	**74**	*stu1951*	-	Hyp. Protein	1727277..1727384	MEKLQVIDLEELLEFDQGYVINASCGPSHSCGGGR
CNRZ1066	*str1572*	0	w	GCATATATGGGA	O	77	*shp0182*	*shp*			
		2	w	GCATATAgGtGA	S	139	*str0234*	-	Hyp. Protein		
		2	w	GaATATAgGGGA	O	237	*str0315*	*rsmG*	Ribosomal RNA small subunit methyltransferase G		
		2	w	GaATATAgGGGA	S	**65**	*str0316*	-	Hyp. Protein		
		2	c	GCATATAcGaGA	S	70	*shp0917*	*shp*			
		0	w	GCATATATGGGA	S	113	*str1573*	-	Major facilitator superfamily permease		
		2	c	GaATATATGcGA	S	**77**	*str1801*	-	Deoxyribonuclease, putative		
		2	w	GCgTATAaGGGA	S	**74**	*str1951*	-	Hyp. Protein	1729146..1729253	MEKLQVIDLEELLEFDQGYVINASCGPSHSCGGGR
				GCATATACGAGA		**70**					
LMD-9	*ster_0934*	2	c	GCATATAgGtGA	O	159	*ster_0708*	-	Hyp. Protein		
		2	w	GaATATAgGAGA	O	248	*ster_0883*	-	Transposase	C(826216..826311)	MPWASGLLISLNFFFLITLYQILSKLDEENK
		2	w	GaAaATACGAGA	O	**77**	*ster_1048*	-	Hyp. Protein	C(973737..974555)	MAKYNIEPLVTLSHYEMPLNLAREYNGWTNRKIIGFYENYVRTVFKR YKDKVKYWLTFNEVNSVLHAPFMLGGIAAPMEGLSKQELYQAVHHEL VGSASITKIGHEIKPEFKIGCMVLAMPAYGMTATPLDQLAVPEFENQ NYLFSDIHARGKYPNYIKYYFKDNGIEIQFSPGDDEILKNTVNFISF SFAMSEATAHNPEDYKVVKGNILGGVENPYLEKSEWGWAIDPIGLRL VLDDFYDRYQLPLFIVENGPDGPTVEDDYRMITFKNT
		2	w	GCATATAtGgGA	S	**74**	*shp1358*	*shp*			
		2	c	GtATATAtGAGA	O	222	*ster_1379*	*gshAB*	Glutamate-cysteine ligase		

		2	c	GCATATAtGgGA	S	**71**	*shp1530*	*shp*			
		2	w	GCATATAtGgGA	S	113	*ster_1531*	-	Major facilitator superfamily permease		
		2	w	GgcTATACGAGA	O	97	*ster_1390*	-	Hyp. Protein		
LMG18311	*stu0917*	2	w	GCATATAtGgGA	O	**77**	*shp0182*	*shp*			
		2	w	GCATATAgGtGA	S	139	*stu0234*	-	Hyp. Protein		
		2	c	GCATATAgGtGA	S	125	*stu0652*	*mip*	Macrophage infectivity potentiator-related protein	C(622711..622866)	MLDKFIRIQTTDASIIDDIVDTSTPSKFTKGKKLHERGDNEEHLPLD DDLI
		2	c	GCATATAgGtGA	O	159	*stu0656*	-	Hyp. Protein		
		2	c	GtATATAtGAGA	O	223	*stu1413*	*gshAB*	Glutamate-cysteine ligase		
		2	w	GgcTATACGAGA	O	97	*stu1422*	-	Predicted membrane protein		
		2	c	GCATATAtGgGA	S	**71**	*shp1572*	*shp*			
		2	w	GCATATAtGgGA	S	113	*stu1573*	-	Major facilitator superfamily permease		
CNRZ1066	*str0917*	2	w	GCATATAtGgGA	O	**77**	*shp0182*	*shp*			
		2	w	GCATATAgGtGA	S	139	*str0234*	-	Hyp. Protein		
		2	w	GgcTATACGAGA	O	97	*str1422*	-	Hyp. Protein		
		2	c	GCATATAtGgGA	S	**71**	*shp1572*	*shp*			
		2	w	GCATATAtGgGA	S	113	*str1573*	-	Major facilitator superfamily permease		
S. mutans											
				GTCTATATGGGA		**74**					
UA159	*smu.1509*	2	w	GTaTATATGaGA	S	176	*smu_320*	-	Putative 5-formyltetrahydrofolate cyclo-ligase		
		2	w	GatTATATGGGA	S	6	*smu_952*	*mmuM*	Putative methyltransferase		
		2	w	GgCTATATGaGA	S	97	*smu_1031*	-	Transposase		
		1	c	GTgTATATGGGA	S	**74**	*smu_1508c*	-	Putative coenzyme PQQ synthesis protein	C(1434004..1434069)	MLTKKEFSVPKTTKVNCWGKH
NN2025	*smunn2025_0595*	1	w	GTgTATATGGGA	S	**74**	*smunn2025_0596*	-	Putative coenzyme PQQ synthesis protein	628599..628664	MLTKKEFSVPKTTKVNCWGKH
		2	w	GTCaATATaGGA	S	41	*smunn2025_0804*	-	Hyp. Protein	839780..840004	MKLKTLKASFTQKEINEDDLQILRQAIVEDDGTRISHREVFKKFGTK VSNSFNLCNIKCLTRWFIGRPYLLLGN
		2	c	GTaTATATGaGA	S	179	*smunn2025_1629*	-	Putative 5-formyltetrahydrofolate cyclo-ligase		
		2	c	GTaTATATGaGA	O	106	*smunn2025_1630*	-	N-acetyl-gamma-glutamyl-phosphate reductase		
		2	w	GTCTATAaaGGA	O	17	*NA*				
S. dysgalactiae											
				GCATATATGGGA		**75**					
subsp. equisimilis	*sdeg_1013*	2	c	GCAaATcTGGGA	O	59	*sdeg_0255*	-	PTS system, mannitol		
GGS_124		1	c	GCATATAcGGGA	S	**80**	*sdeg_1012*	-	transporter, major facilitator family protein		
S. equi											
				GTACATACGGGA		**75**					
MGCS10565	*sez_0716*	2	w	GagCATACGGGA	S	**83**	*sez_0517*	-	Transcriptional regulator PlcR	540523..540561	MIKKKEMDMRIK
		1	c	GcACATACGGGA	S	**64**	*sez_0631*	-	Positive transcriptional regulator MutR family		
		3		GcAgATAtGGGA	S	**75**	*sez_0717*	-	Radical SAM protein		
subsp. zooepidemicus	*szo_12400*	3		GcAgATAtGGGA	S	**75**	*szo_12390*	-	Radical SAM protein	C(1381625..1381735)	MQNDSQKESESLDLSINKIDLSSVILLYESSKSHGH
		2	w	GcAtATACGGGA	S	**64**	*szo_13660*	-	Positive transcriptional regulator MutR family		
		2	c	GagCATACGGGA	S	**83**	*szo_14750*	-	Transcriptional regulator PlcR	C(1652450..1652488)	MIKKKEMDMRIK
				GCATATATGGGA		**75**					
subsp. zooepidemicus	*szo_17730*	2	w	GCtTATATGaGA	S	6	*szo_05150*	-	Hyp. Protein	596052..596156	MRTFKRCSRFLFSGGKAKYRRTIITTHRGVLSYG
		1	c	GCAgATATGGGA	S	**75**	*szo_12390*	-	Radical SAM protein	C(1381625..1381735)	MQNDSQKESESLDLSINKIDLSSVILLYESSKSHGH
		1	w	GCATATAcGGGA	S	64	*szo_13660*	-	Positive transcriptional regulator MutR family		
		2	c	GCAgATATaGGA	S	**74**	*szo_17710*	-	Radical SAM protein	C(1964965..1965030)	MSPKKEFNAPKTTKVNSWGKH
S. suis											
				GTATATAAGGGA		**92**					
SC84	*ssusc84_1811*	2	c	GTAaATAAGcGA	S	199	*ssusc84_0805*	*trmFO*	Methyltransferase trmFO		
		1	w	GTATATAAGaGA	O	124	*ssusc84_0936*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GgAaATAAGGGA	S	122	*ssusc84_1213*	-	Cation transport ATPase		
		2	w	GgAaATAAGGGA	O	126	*ssusc84_1212*	*prfC*	Peptide chain release factor 3		
		2	c	GTAaATAAGaGA	S	115	*ssusc84_1271*	*ftsE*	Cell division protein		
		2	c	GcgTATAAGGGA	S	**91**	*ssusc84_1810*	-	Radical SAM superfamily protein	C(1907212..1907325)	MSVIEFKKLSSKLLKVSKIETLSGCGTKRTQQTKGWY
GZ1	*ssgz1_1813*	2	c	GTtTATAAGtGA	O	68	*ssgz1_0583*	-	Transposase		
		2	c	GTAaATAAGcGA	S	199	*ssgz1_0868*	*trmFO*	Methyltransferase trmFO		
		1	w	GTATATAAGaGA	O	103	*ssgz1_0915*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GgAaATAAGGGA	S	122	*ssgz1_1198*	-	Cation transport ATPase		
		2	w	GgAaATAAGGGA	O	126	*ssgz1_1197*	*prfC*	Peptide chain release factor 3		
		2	c	GTAaATAAGaGA	S	61	*ssgz1_1254*	*ftsE*	Cell division protein		
		2	c	GTggATAAGGGA	S	178	*ssgz1_1264*	-	Transposase		
		2	c	GcgTATAAGGGA	S	**91**	*ssgz1_1812*	-	Radical SAM superfamily protein	C(1851902..1852015)	MSVIEFKKLSSKLLKVSKIETLSGCGTKRTQQTKGWY

P1/7	*ssu1789*	2	c	GTAaATAAGcGA	S	199	*ssu0843*	*trmFO*	Methyltransferase trmFO		
		1	w	GTATATAAGaGA	O	124	*ssu0891*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GgAaATAAGGGA	S	122	*ssu1180*	-	Cation transport ATPase		
		2	w	GgAaATAAGGGA	O	126	*ssu1179*	*prfC*	Peptide chain release factor 3		
		2	c	GTAaATAAGaGA	S	115	*ssu1239*	*ftsE*	Cell division protein		
		2	c	GcgTATAAGGGA	S	**91**	*ssu1788*	-	Radical SAM superfamily protein	C(1818803..1818916)	MSVIEFKKLSSKLLKVSKIETLSGCGTKRTQQTKGWY
98HAH33	*ssu98_2002*	2	w	GcATATAAGcGA	S	34	*ssu98_0873*	-	Hyp. Protein		
		2	c	GTAaATAAGcGA	S	199	*ssu98_0906*	*trmFO*	Methyltransferase trmFO		
		1	w	GTATATAAGaGA	O	103	*ssu98_1048*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GgAaATAAGGGA	S	122	*ssu98_1363*	-	Cation transport ATPase		
		2	w	GgAaATAAGGGA	O	126	*ssu98_1362*	*prfC*	Peptide chain release factor 3		
		2	c	GTggATAAGGGA	O	40	*ssu98_1436*	-	Transposase		
		2	c	GTggATAAGGGA	S	202	*NA*			C(1369730..1370032)	MMPDHVHMLVLISPKLSISDFMGYLKSKSALMIFDKHANLKYKYGNR KFWARGYYVSTVGLNEKTVAKYIREQEKNDIALDKLSVKEYEDSFSD SSFRTR
05ZYH33	*ssu05_1997*	2	c	GTAaATAAGcGA	S	199	*ssu05_0901*	*trmFO*	Methyltransferase trmFO		
		1	w	GTATATAAGaGA	O	103	*ssu05_1036*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GgAaATAAGGGA	S	122	*ssu05_1348*	-	Cation transport ATPase		
		2	w	GgAaATAAGGGA	O	126	*ssu05_1347*	*prfC*	Peptide chain release factor 3		
		2	c	GTggATAAGGGA	S	202	*ssu05_1420*	-	Hyp. Protein	C(1370301..1370603)	MMPDHVHMLVLISPKLSISDFMGYLKSKSALMIFDKHANLKYKYGNR KFWARGYYVSTVGLNEKTVAKYIREQEKNDIALDKLSVKEYEDSFSD SSFRTR
		2	c	GTggATAAGGGA	O	40	*ssu05_1423*	-	Transposase		
		2	c	GcgTATAAGGGA	S	**91**	*ssu05_1996*	-	Radical SAM superfamily protein	C(1907597..1907710)	MSVIEFKKLSSKLLKVSKIETLSGCGTKRTQQTKGWY
BM407	*ssubm407_1859*	2	w	GTAaATAAGaGA	S	115	*ssubM407_0551*	*ftsE*	Cell division protein		
		2	c	GgAaATAAGGGA	S	122	*ssubM407_0609*	-	Cation transport ATPase		
		2	c	GgAaATAAGGGA	O	126	*ssubM407_0610*	*prfC*	Peptide chain release factor 3		
		1	c	GTATATAAGaGA	O	124	*ssubM407_0888*	*lacG*	6-phospho-beta-galactosidase		
		2	w	GTATATcAaGGA	S	262	*ssubM407_0979*	-	Transposase		
		2	c	GcgTATAAGGGA	S	**91**	*ssubm407_1858*	-	Radical SAM superfamily protein	C(1957463..1957576)	MSVIEFKKLSSKLLKVSKIETLSGCGTKRTQQTKGWY

Annexe 6 - Tableau C : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe III.

		Cibles								spCDS identifiées dans bactgeneSHOW	
Strains	**Gène *rgg***	**nM**	**Brin**	**Target_motif**	**Dir**	**dATG**	**Locus_tag**	**Gène**	**Fonction**	**Localisation**	**Séquence**
S. thermophilus											
				ATnTCnnATnTnCnATTTT		**75**					
LMD-9	*ster_1299*	0	w	ATTTCATATCTTCAATTTT	S	**78**	*ster_1297*	*pepF*	Oligoendopeptidase F	C(1206895..1206960)	MFIEDIEVEEQVCNGGCFIAC
S. pneumoniae											
				ATnTCnnATnTnCnATTTT		**77**					
ATCC 700669	*spn23f_01510*	0	c	ATATCAAATATGCGATTTT	S	127	*spn23f01520*		Hyp. Protein		
JJA	*spj_0171*	0	c	ATATCAAATATGCGATTTT	S	127	*spj_0172*		Hyp. Protein		
Hungary19A-6	*sph_0253*	0	c	ATATCAAATATGCGATTTT	S	127	*sph_0254*		Hyp. Protein		
TIGR4	*sp_0141*	0	c	ATATCAAATATGCGATTTT	S	127	*sp_0142*		Hyp. Protein		
P1031	*spp_0211*	0	c	ATATCAAATATGCGATTTT	S	127	*spp_0212*		Hyp. Protein		
CGSP14	*spcg_0143*	0	c	ATATCAAATATGCGATTTT	S	110	*spcg_0144*		Hyp. Protein		

Annexe 7

Tableaux des gènes cibles putatifs contenant un motif ADN dans leur séquence codante.

Tableaux des gènes cibles putatifs contenant un motif ADN dans leur séquence codante

- **Tableau A :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe I.
- **Tableau B :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe II.
- **Tableau C :** Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe III.

Se référer à l'Annexe 6 sauf pour les deux points ci-dessous :

La colonne « dATG » indique la distance en nucléotides motif / codon start du gène (ou de la spCDS quand détectée) en amont duquel le motif a été identifié. Ce gène est appelé « locus_tag ».

La colonne « Gène » et « Fonction » indiquent respectivement le gène dans lequel le motif a été identifié et la fonction de ce dernier.

NA indique des erreurs lors de la production de données n'ayant pas pu être vérifiées pour l'écriture de ce manuscrit.

Légende :

Biosynthèse d'ADN et d'ARN	Synthèse ATP
Biosynthèse du mevalonate	Traduction
Biosynthèse du peptidoglycane / paroi cellulaire	Transcription
Division cellulaire / ségrégation du chromosome	Transfert de groupe phosphoryle
Machinerie d'internalisation de l'ADN	Transporteur d'oligopeptides
Métabolisme des acides aminés	Transporteurs / Permeases
Métabolisme glucidique	Transport des acides aminés
Métabolisme lipidique	Transport des métaux, vitamines et ions
Methyltransferase	Transport des sucres
Protéines de stress	Transposition
Protéines secrétées et surface / facteurs de virulence / pathogénicité	Hyp. protein
Régulateurs transcriptionnels	Autres
Réplication / Recombinaison / DNA repair / cell maintenance	

Annexe 7 - Tableau A : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe I.

Strains	Gène *rgg*	Cibles nM	Brin	Target_motif	Dir	dATG	Locus_tag	Gène_tag	Fonction	spCDS identifiées dans bactgeneSHOW Localisation	Séquence
S. thermophilus											
				CCCATTTCCCCAAC		**75**					
LMG18311	*stu1044*	2	c	CCCAcTTCCaCAAC	S	246	*stu0080*	*pnpA*	Polynucleotide phosphorylase/polyadenylase		
		3	w	CCCATcTCatCAAC	O	294	*stu0159*	*uppP*	Undecaprenyl pyrophosphate phosphatase		
		3	w	CCCtTTTCgtCAAC	O	43	*stu0789*	*dfp2*	Phosphopantothenate--cysteine ligase		
		3	w	CCCgTaTCCCCAAa	O	39	*NA*				
CNRZ1066	*str1044*	2	c	CCCAcTTCCaCAAC	S	246	*str0080*	*pnpA*	Polynucleotide phosphorylase/polyadenylase		
		3	w	CCCATcTCatCAAC	O	294	*str0159*	*uppP*	Undecaprenyl pyrophosphate phosphatase		
		3	c	CCCATTTttCaAAC	O	187	*str0416*	*str04165*	Hyp. Protein		
		3	w	CCCtTTTCgtCAAC	O	43	*str0789*	*dfp2*	Phosphopantothenate--cysteine ligase		
				CCCATATATGCAAC		**76**					
LMG18311	*stu0182*	3	c	CCCATgTtTGCAgC	O	278	*stu0665*	*stu0664*	Metallopeptidase, SprT family		
CNRZ1066	*str0182*	3	c	CCCATgTtTGCAgC	O	278	*str0665*	*str0664*	Metallopeptidase, SprT family		
S. agalactiae											
				CCCACTTTCACAAC		**85**					
NEM316	*gbs.1555*	3	c	CCCAaTTTCAaAAa	S	108	*gbs1124*	*gbs1125*	Hyp. Protein		
2603V/R	*sag1490*	3	w	CCCttTTTCcCAAC	O	40	*sag0545*	*NA*	*PepSHOW*	C(559910..560053)	MEERLRQNKKAPRSNFVQGEREANSSIVKTCFLGGVFTIPILTENEV
S. dysgalactiae											
				CCCACTTTCACAAC		**85**					
subsp.	*sdeg_0529*	3	c	CCaAtTTTCAaAAC	O	158	*sdeg_0370*	*sdeg_0369*	RNA-binding protein		
equisimilis		3	w	CCCAaTTTgACAAg	O	252	*sdeg_1008*	*smf*	DNA processing protein		
GGS_124		3	w	CCatCTTTgACAAC	O	188	*sdeg_1308*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		2	w	CCCAgTTTCgCAAC	S	209	*sdeg_1635*	*sdeg_1634*	Phage protein		
		3	w	CCaACTTctACAAC	O	47	*sdeg_1940*	*purR*	Purine operon repressor		
S. pyogenes											
				CCCACTTTCACAAC		**84**					
M1 GAS	*spy_0496*	3	w	CCCAtTTTCAaAgC	O	126	*spy_0131*	*spy_0133*	Transposase	c(123789..123827)	MISIELMAFLQV
		3	c	CCaACTTctACAAC	O	47	*spy_0269*	*purR*	Purine operon repressor	NA	
		3	c	CCaAtTTTCAaAAC	O	158	*spy_0308*	*spy_0307*	RNA-binding protein, YhbY family	NA	
		3	w	CCaACTTagACAAC	S	275	*spy_0696*	*spy_0695*	Phage protein		
		3	c	CCCAaTTTgACAAg	O	252	*spy_1164*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*spy_1310*	*dltB*	Putative integral membrane protein	NA	
		3	w	CCatCTTTgACAAC	O	188	*spy_1312*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1	NA	
		3	c	CCCtCTTTaACAgC	O	175	*spy_1509*	*clpE*	Putative ATP-dependent protease	NA	
		3	c	CCCcCcTcCACAAC	S	202	*spy_2018*	*mga*	M protein trans-acting positive regulator	NA	
str. Manfredo	*spyM51462*	3	c	CCaACTTctACAAC	O	47	*spyM50208*	*purR*	Purine operon repressor		
		3	w	CCaACTTagACAAC	S	275	*spyM50519*	*spyM50518*	Phage protein		
		3	w	CCCtCTTTaACAgC	O	175	*spyM50611*	*clpE*	Putative ATP-dependent protease		
		3	c	CCatCTTTgACAAC	O	188	*spyM50793*	*dltB*	Putative integral membrane protein		
		3	w	CCCtCTTTCAaAgC	S	142	*spyM50794*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	w	CCCAaTTTgACAAg	O	251	*spyM50905*	*smf*	DNA processing protein		
		3	w	CCaACTTTCtCAAa	O	274	*spyM51124*	*mvaD*	Mevalonate diphosphate decarboxylase		
		2	w	CCCAtTTcCACAAC	O	253	*spyM51296*	*spyM51297*	Hyp. Protein		
		3	w	CCaAtTTTCAaAAC	O	158	*spyM51591*	*spyM51592*	RNA-binding protein, YhbY family		
MGAS8232	*spyM18_0554*	3	c	CCaACTTctACAAC	O	47	*spyM18_0256*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*spyM18_0304*	*spyM18_03043*	RNA-binding protein, YhbY family		

		2	w	CCCACgTTCACAAa	S	274	*spyM18_0361*	*spyM18_0360*	Hyp. phage protein		
		3	w	CCCACcTTCAaAAa	O	257	*spyM18_0577*	*spyM18_0579*	ABC transporter ATP-binding protein		
		3	c	CCaACTTTCtCAAa	O	274	*spyM18_0939*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*spyM18_1124*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*spyM18_1322*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*spyM18_1323*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*spyM18_1528*	*clpE*	Putative ATP-dependent protease		
		3	c	CCaACTTagACAAC	S	275	*spyM18_1764*	*spyM18_1765*	Phage protein		
SSI-1	*sps1508*	3	c	CCaACTTctACAAC	O	47	*sps0202*	*purR*	Purine operon repressor		
		3	w	CCaACTTagACAAC	S	275	*sps0643*	*sps0642*	Phage protein		
		3	w	CCCtCTTTaACAgC	O	175	*sps0698*	*clpE*	Putative ATP-dependent protease		
		3	c	CCatCTTTgACAAC	O	188	*sps0865*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		2	c	CCCAtTTcCACAAC	O	253	*sps0905*	*sps0904*	Hyp. Protein		
		3	c	CCCAaTTTgACAAg	O	251	*sps1021*	*smf*	DNA processing protein		
		3	w	CCCcaTTTaACAAC	O	8	*sps1141*	*sps1141*	Phage protein		
		2	c	CCCACaTTCACAAa	S	237	*sps1150*	*sps1151*	Putative immunity repressor protein - phage associated		
		3	w	CCaACTTTCtCAAa	O	274	*sps1256*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	w	CCaAtTTTCAaAAC	O	158	*sps1634*	*sps16345*	RNA-binding protein, YhbY family		
MGAS315	*spyM3_0346*	3	c	CCaACTTctACAAC	O	47	*spyM3_0197*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*spyM3_0226*	*spyM3_0225*	RNA-binding protein, YhbY family		
		3	c	CCaACTTTCtCAAa	O	274	*spyM3_0597*	*mvaD*	Mevalonate diphosphate decarboxylase		
		2	w	CCCACaTTCACAAa	S	273	*spyM3_0702*	*spyM3_0701*	Putative immunity repressor protein - phage associated		
		3	c	CCCAaTTTgACAAg	O	251	*spyM3_0820*	*smf*	DNA processing protein		
		3	w	CCatCTTTgACAAC	O	188	*spyM3_0993*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		2	w	CCCAtTTcCACAAC	O	253	*spyM3_0951*	*spyM3_0952*	Hyp. Protein		
		3	c	CCCtCTTTaACAgC	O	175	*spyM3_1164*	*clpE*	Putative ATP-dependent protease		
		3	c	CCaACTTagACAAC	S	275	*spyM3_1218*	*spyM3_1219*	Phage protein		
MGAS10394	*M6_spy0433*	3	c	CCaACTTctACAAC	O	29	*M6_spy0261*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*M6_spy0291*	*M6_spy0290*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	73	*M6_spy0475*	*shp*		469640..469711	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCAaAAa	O	138	*M6_spy0515*	*M6_spy05154*	Hyp. Protein		
		3	c	CCaACTTTCtCAAa	O	226	*M6_spy0701*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*M6_spy0881*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*M6_spy1041*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*M6_spy1042*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		2	c	CCCACTTTCAaAAg	S	86	*M6_spy1139*	*M6_spy1140*	Phage protein		
		3	c	CCCtCTTTaACAgC	O	175	*M6_spy1262*	*clpE*	Putative ATP-dependent protease		
		3	c	CCaACTTagACAAC	S	275	*M6_spy1352*	*M6_spy1353*	Phage protein		
		3	c	CCCcCcTcCACAAC	S	201	*M6_spy1719*	*mga*	M protein trans-acting positive regulator		
MGAS5005	*mutR*	3	w	CCCAtTTTCAaAgC	O	126	*M5005_spy_0112*	*M5005_spy_0113*	Transposase	C(124345..124383)	MISIELMAFLQV
		3	c	CCaACTTctACAAC	O	47	*M5005_spy_0228*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*M5005_spy_0263*	*M5005_spy_02632*	RNA-binding protein, YhbY family		
		3	c	CCaACTTTCtCAAa	O	274	*M5005_spy_0228*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*M5005_spy_0884*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*M5005_spy_1071*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*M5005_spy_1073*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*M5005_spy_1240*	*clpE*	Putative ATP-dependent protease		
		3	c	CCCcCcTcCACAAC	S	202	*M5005_spy_1719*	*mga*	M protein trans-acting positive regulator		
MGAS6180	*M28_spy0395*	3	c	CCaACTTctACAAC	O	29	*M28_spy0223*	*purR*	Purine operon repressor		

		3	c	CCaAtTTTCAaAAC	O	158	*M28_spy0255*	*M28_spy0254*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	74	*M28_spy0429*	*shp*		439629..439700	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCtCAAa	O	226	*M28_spy0664*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*M28_spy0859*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*M28_spy1052*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*M28_spy1053*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*M28_spy1180*	*clpE*	Putative ATP-dependent protease		
		3	c	CCaACTTagACAAC	S	275	*M28_spy1237*	*M28_spy1238*	Phage protein		
		3	w	CCCAtTTTCAaAAa	O	196	*M28_spy1307*	*M28_spy1308*	Hyp. Protein		
MGAS9429	*MGAS9429_spy0406*	3	c	CCaACTTctACAAC	O	29	*MGAS9429_spy0231*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*MGAS9429_spy0262*	*MGAS9429_spy0261*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	74	*MGAS9429_spy0440*	*shp*		440119..440190	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCtCAAa	O	226	*MGAS9429_spy0739*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	199	*MGAS9429_spy1003*	*smf*	DNA processing protein	959057..959179	MGYYAISFKQTVIGSVSNWEQKLQLSQKRVLKSPPLKRNQ
		3	c	CCCtCTTTCAaAgC	S	142	*MGAS9429_spy1113*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*MGAS9429_spy1114*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*MGAS9429_spy1236*	*clpE*	Putative ATP-dependent protease		
		3	c	CCCACgTcaACAAC	O	82	*MGAS9429_spy1341*	*atoD1*	Acetate CoA-transferase alpha subunit		
MGAS10270	*MGAS10270_spy0408*	3	c	CCaACTTctACAAC	O	29	*MGAS10270_spy0229*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*MGAS10270_spy0260*	*MGAS10270_spy0259*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	74	*MGAS10270_spy0442*	*shp*		439134..439205	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCtCAAa	O	226	*MGAS10270_spy0742*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	253	*MGAS10270_spy0999*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*MGAS10270_spy1127*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*MGAS10270_spy1128*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*MGAS10270_spy1257*	*clpE*	Putative ATP-dependent protease		
		3	c	CCaACTTagACAAC	S	275	*MGAS10270_spy1314*	*MGAS10270_spy1315*	Phage protein		
		3	w	CCCAtTTTCAaAAa	O	196	*MGAS10270_spy1382*	*MGAS10270_spy1383*	Hyp. Protein		
MGAS2096	*MGAS2096_spy0426*	3	c	CCaACTTctACAAC	O	29	*MGAS2096_spy0247*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*MGAS2096_spy0281*	*MGAS2096_spy0280*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	74	*MGAS2096_spy0460*	*shp*		442029..442100	MKKISKFLPILILAMDIIIIVGG
		3	w	CCaACTTagACAAC	S	275	*MGAS2096_spy0589*	*MGAS2096_spy0588*	Phage protein		
		3	c	CCaACTTTCtCAAa	O	226	*MGAS2096_spy0755*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*MGAS2096_spy0959*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*MGAS2096_spy1070*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*MGAS2096_spy1071*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*MGAS2096_spy1259*	*clpE*	Putative ATP-dependent protease		
		3	c	CCCACgTcaACAAC	O	82	*MGAS2096_spy1367*	*atoD1*	Acetate CoA-transferase alpha subunit		
MGAS10750	*MGAS10750_spy0420*	3	c	CCaACTTctACAAC	O	29	*MGAS10750_spy0224*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*MGAS10750_spy0257*	*MGAS10750_spy0256*	RNA-binding protein, YhbY family		
		1	w	CCCACTTTCcCAAC	S	74	*MGAS10750_spy0461*	*shp*		451813..451884	MKKISKFLPILILAMDIIIIVGG
		3	c	CCaACTTTCtCAAa	O	226	*MGAS10750_spy0776*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*MGAS10750_spy1034*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*MGAS10750_spy1164*	*dltB*	Putative integral membrane protein		
		3	w	CCatCTTTgACAAC	O	188	*MGAS10750_spy1165*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCtCTTTaACAgC	O	175	*MGAS10750_spy1348*	*clpE*	Putative ATP-dependent protease		
NZ131	*spy49_0415*	3	c	CCaACTTctACAAC	O	47	*spy49_0229*	*purR*	Purine operon repressor		
		3	c	CCaAtTTTCAaAAC	O	158	*spy49_0260*	*spy49_0259*	RNA-binding protein, YhbY family		
		3	c	CCaACTTTCtCAAa	O	274	*spy49_0692*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTgACAAg	O	252	*spy49_0915*	*smf*	DNA processing protein		
		3	c	CCCtCTTTCAaAgC	S	142	*spy49_1035c*	*dltB*	Putative integral membrane protein		

		3	w	CCatCTTTgACAAC	O	188	*spy49_1036c*	*dltA*	D-alanine--poly(phosphoribitol) ligase subunit 1		
		3	c	CCCACgTcaACAAC	O	82	*spy49_1270*	*atoD1*	Acetate CoA-transferase alpha subunit		
				CCCACTTTCCCAAC		**74**					
M1 GAS	*spy_0533*	3	c	CCaACTTTCtCAAa	O	274	*spy_0877*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	199	*spy_0894*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*spy_1410*	*spy_1411*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*spy_1519*	*ftsZ*	Cell division protein		
		3	w	CCCAtTTTtCCAAa	O	201	*spy_1704*	*lacC1*	Tagatose-6-phosphate kinase	C(1414969..1415232)	MATGFSGGKLGDFIIHQLQEQGISNQFFKIKGETRNCIAVLH EGMQTKILEAGPYSDVDEAEVSLSHMSTIAKPFDVLTSQAVC LGD
str. Manfredo	*spyM51424*	3	c	CCCAaTTTCCaAAa	O	147	*spyM50604*	*ftsZ*	Cell division protein		
		3	c	CCCgCTTcCCCAAg	O	246	*spyM50710*	*spyM50709*	Methyltransferase		
		3	c	CCCtCTTTCCaAgC	S	131	*spyM50902*	*spyM50903*	Oxaloacetate decarboxylase		
		3	w	CCCAaTTTaCCAAg	O	199	*spyM51108*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCaACTTTCtCAAa	O	274	*spyM51124*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	w	CCCAtTTcCaCAAC	O	253	*spyM51296*	*spyM51297*	Hyp. Protein		
MGAS8232	*spyM18_0599*	3	w	CCCACgTTCaCAAa	S	274	*spyM18_0361*	*spyM18_0360*	Hyp. phage protein		
		3	c	CCCcgTTTtCCAAC	O	285	*spyM18_0386*	*hylP*	phage hyaluronidase		
		3	c	CCaACTTTCtCAAa	O	274	*spyM18_0939*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*spyM18_0956*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCAaTTTCCaAAa	O	147	*spyM18_1536*	*ftsZ*	Cell division protein		
		3	w	CCCAgTTTaCCAcC	O	269	*spyM18_1714*	*lacC1*	Tagatose-6-phosphate kinase	C(1427898..1427975)	MTMATFKQLYQQLQVREVSHDNHSK
		3	w	CCCcgTTTtCCAAC	O	287	*spyM18_1756*	*hylP*	Phage hyaluronidase		
SSI-1	*sps1476*	3	c	CCCAaTTTCCaAAa	O	147	*sps0691*	*ftsZ*	Cell division protein		
		3	c	CCCgCTTcCCCAAg	O	168	*sps0788*	*sps0787*	Methyltransferase		
		3	c	CCCAtTTcCaCAAC	O	253	*sps0905*	*sps0904*	Hyp. Protein		
		3	c	CCCtCTTTCtCAAg	O	288	*sps0938*	*NA*	*pepSHOW*	919861..919905	MTSGPVLIGTLLGD
		3	c	CCCACaTTCaCAAa	S	237	*sps1150*	*sps1151*	Putative immunity repressor protein - phage associated		
		3	w	CCCAaTTTaCCAAg	O	199	*sps1240*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCaACTTTCtCAAa	O	274	*sps1256*	*mvaD*	Mevalonate diphosphate decarboxylase		
MGAS315	*spyM3_0377*	3	c	CCaACTTTCtCAAa	O	274	*spyM3_0597*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	w	CCCACaTTCaCAAa	S	273	*spyM3_0702*	*spyM3_0701*	Putative immunity repressor protein - phage associated		
		3	w	CCCAtTTcCaCAAC	O	253	*spyM3_0951*	*spyM3_0952*	Hyp. Protein		
		3	w	CCCgCTTcCCCAAg	O	246	*spyM3_1075*	*spyM3_1076*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*spyM3_1171*	*ftsZ*	Cell division protein		
MGAS10394	*M6_spy0474*	1	w	CCCACTTTCcCAAC	S	84	*M6_spy0432*	*shp*		C(427870..427938)	MKKVNKALLFTLIMDILIIVGG
		3	c	CCaACTTTCtCAAa	O	193	*M6_spy0717*	*deoD*	Purine nucleoside phosphorylase		
		3	c	CCCAaTTTaCCAAg	O	95	*M6_spy0794*	*NA*	*pepSHOW*	799344..799397	MEQTFFMIKPDGVKRGG
		3	w	CCCcgTTTtCCAAC	S	86	*M6_spy1139*	*M6_spy1140*	Phage protein		
		3	c	CCCACTTTCaaAAg	O	168	*M6_spy1177*	*M6_spy1176*	Methyltransferase		
		3	w	CCCgCTTcCCCAAg	O	147	*M6_spy1269*	*ftsZ*	Cell division protein		
		3	w	CCCAaTTTCCaAAa	O	287	*M6_spy1347*	*M6_spy13478*	Phage hyaluronidase		
		3	w	CCCcgTTTtCCAAC	O	269	*M6_spy1444*	*lacC1*	Tagatose-6-phosphate kinase	C(1441948..1442025)	MTMATFKQLYQQLQVREVSHDNHSK
MGAS5005	*M5005_spy_0440*	3	c	CCaACTTTCtCAAa	O	274	*M5005_spy_0684*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	199	*M5005_spy_0700*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	246	*M5005_spy_1149*	*M5005_spy_1150*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*M5005_spy_1248*	*ftsZ*	Cell division protein		
		3	w	CCCAtTTTtCCAAa	O	46	*M5005_spy_1396*	*lacC1*	Tagatose-6-phosphate kinase		
MGAS6180	*M28_spy0428*	3	c	CCaACTTTCtCAAa	O	226	*M28_spy0664*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*M28_spy0680*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*M28_spy1145*	*M28_spy1144*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*M28_spy1187*	*ftsZ*	Cell division protein		
		3	w	CCCAgTTTaCCAcC	O	269	*M28_spy1438*	*lacC1*	Tagatose-6-phosphate kinase	C(1439485..1439562)	MTMATFKQLYQQLQVREVSHDNHSK

		3	w	CCCAtTTTtCCAAa	O	201	*M28_spy1439*	*lacB1*	Galactose-6-phosphate isomerase		
MGAS9429	*MGAS9429_spy0439*	3	c	CCaACTTTCtCAAa	O	226	*MGAS9429_spy0739*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*MGAS9429_spy0756*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*MGAS9429_spy1197*	*MGAS9429_spy1196*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*MGAS9429_spy1243*	*ftsZ*	Cell division protein		
		3	w	CCCAgTTTaCCAcC	O	269	*MGAS9429_spy1397*	*lacC1*	Tagatose-6-phosphate kinase	C(1364378..1364455)	MTMATFKQLYQQLQVREVSHDNHSK
MGAS10270	*MGAS10270_spy0441*	3	c	CCaACTTTCtCAAa	O	226	*MGAS10270_spy0742*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*MGAS10270_spy0758*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*MGAS10270_spy1221*	*MGAS10270_spy1220*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*MGAS10270_spy1264*	*ftsZ*	Cell division protein		
		3	w	CCCAgTTTaCCAcC	O	269	*MGAS10270_spy1516*	*lacC1*	Tagatose-6-phosphate kinase	C(1463403..1463480)	MTMATFKQLYQQLQVREVSHDNHSK
		3	w	CCCAtTTTtCCAAa	O	201	*MGAS10270_spy1517*	*lacB1*	Galactose-6-phosphate isomerase		
MGAS2096	*MGAS2096_spy0459*	3	c	CCCAtTTTtCaAAC	S	172	*MGAS2096_spy0266*	*MGAS2096_spy0267*	D-alanyl-D-alanine carboxypeptidase		
		3	c	CCaACTTTCtCAAa	O	226	*MGAS2096_spy0755*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*MGAS2096_spy0772*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*MGAS2096_spy1217*	*MGAS2096_spy1216*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*MGAS2096_spy1266*	*ftsZ*	Cell division protein		
MGAS10750	*MGAS10750_spy0460*	3	c	CCaACTTTCtCAAa	O	226	*MGAS10750_spy0776*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	193	*MGAS10750_spy0792*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	168	*MGAS10750_spy1257*	*MGAS10750_spy1256*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*MGAS10750_spy1355*	*ftsZ*	Cell division protein		
NZ131	*spy49_0449c*	3	c	CCaACTTTCtCAAa	O	274	*spy49_0692*	*mvaD*	Mevalonate diphosphate decarboxylase		
		3	c	CCCAaTTTaCCAAg	O	199	*spy49_0709*	*deoD*	Purine nucleoside phosphorylase		
		3	w	CCCgCTTcCCCAAg	O	246	*spy49_1127c*	*spy49_1128*	Methyltransferase		
		3	w	CCCAaTTTCCaAAa	O	147	*spy49_1172c*	*ftsZ*	Cell division protein		
		3	w	CCCAtTTTtCCAAa	O	46	*spy49_1324c*	*lacC1*	Tagatose-6-phosphate kinase		
S. pneumoniae											
				CCCATTTTCCCAAC		**74**					
R6	*spr0960*	3	c	CCCATTTTCtaAAa	O	150	*spr0107*	*spr0106*	Transporter-truncation		
		3	c	CCCtTTTaCCCAAa	O	97	*spr0425*	*lacG*	6-phospho-beta-galactosidase		
		3	c	CCCtTTTTaCCAAg	S	165	*spr0498*	*NA*	*pepSHOW*	C(505454..505546)	MPKTPLGALTRGTIRFHLTSQTFISKSMRK
		3	c	CCCtaTTTCCaAAC	S	49	*spr0622*	*glnP*	Glutamine ABC transporter permease		
		3	c	CCCcaTTTCaCAAC	O	116	*spr0834*	*proA*	Gamma-glutamyl phosphate reductase		
		3	w	CCCATTTcaCaAAC	S	258	*spr1045*	*rnhB*	Ribonuclease HII		
		3	w	CCCtgTTTCtCAAC	O	23	*spr1083*	*recN*	RecN		
		3	c	CCCAcTTgCCaAAC	S	76	*spr1107*	*mutY*	A/G-specific adenine glycosylase		
		3	w	CCCATTTTTtCAgC	O	31	*spr1359*	*atpD*	F0F1 ATP synthase subunit beta		
		3	w	CCatTTTTCtCAAC	O	10	*spr1369*	*greA*	Elongation factor		
		3	c	CCCAaTTTCaCAAg	S	62	*spr1509*	*ftsZ*	Cell division protein		
		3	w	CCaAaTTTCtCAAC	O	293	*spr1594*	*spr1595*	HAD phosphatase, family IIIA		
		2	c	CCCATTTgaCCAAC	S	229	*spr1881*	*pgi*	Glucose-6-phosphate isomerase		
		2	w	CCCATTTTtCCAAg	O	158	*spr1965*	*spr1966*	Alpha-L-fucosidase (F-type lectins)		
D39	*spd_0939*	3	c	CCCtTTTaCCCAAa	O	97	*spd_0428*	*lacG*	6-phospho-beta-galactosidase		
		3	c	CCCtTTTTaCCAAg	S	8082	*spd_0499*	*NA*	*pepSHOW*	C(512984..513076)	MPKTPLGALTRGTIRFHLTSQTFISKSMRK
		3	c	CCCtaTTTCCaAAC	S	49	*spd_0616*	*glnP*	Glutamine ABC transporter permease		
		3	c	CCCcaTTTCaCAAC	O	116	*spd_0824*	*proA*	Gamma-glutamyl phosphate reductase		
		3	w	CCCATTTcaCaAAC	S	273	*spd_1021*	*rnhB*	Ribonuclease HII		
		3	w	CCCtgTTTCtCAAC	O	23	*spd_1061*	*recN*	RecN		
		3	c	CCCAcTTgCCaAAC	S	127	*spd_1085*	*mutY*	A/G-specific adenine glycosylase		
		3	w	CCCATTTTTtCAgC	O	31	*spd_1334*	*atpD*	F0F1 ATP synthase subunit beta		
		3	w	CCatTTTTCtCAAC	O	10	*spd_1345*	*greA*	Elongation factor		
		3	c	CCCAaTTTCaCAAg	S	62	*spd_1478*	*ftsZ*	Cell division protein		
		3	w	CCaAaTTTCtCAAC	O	293	*spd_1559*	*spD_1560*	HAD phosphatase, family IIIA		
		2	c	CCCATTTgaCCAAC	S	229	*spd_1896*	*pgi*	Glucose-6-phosphate isomerase		
		2	w	CCCATTTTtCCAAg	O	158	*spd_1987*	*spD_1988*	Alpha-L-fucosidase (F-type lectins)		
JJA	*spj_0996*	3	c	CCCATTTTCtaAAa	O	150	*spj_0141*	*spj_0140*	LagD protein		

		3	c	CCCtTTTTaCCAAg	S	166	*spj_0520*	*NA*	*pepSHOW*	C(507176..507268)	MPKTPLGALTRGTIRFYLTSQTFISKSMRK
		3	c	CCCtaTTTCCaAAC	S	49	*spj_0649*	*glnP*	Glutamine ABC transporter permease		
		3	c	CCCcaTTTCaCAAC	O	116	*spj_0873*	*proA*	Gamma-glutamyl phosphate reductase		
		3	w	CCCATTTcaCaAAC	S	273	*spj_1077*	*rnhB*	Ribonuclease HII		
		3	w	CCCtgTTTCtCAAC	O	23	*spj_1119*	*recN*	RecN		
		3	c	CCCAcTTgCCaAAC	S	127	*spj_1141*	*mutY*	A/G-specific adenine glycosylase		
		3	w	CCCATTTTttCAgC	O	31	*spj_1409*	*atpD*	F0F1 ATP synthase subunit beta		
		3	w	CCatTTTTCtCAAC	O	10	*spj_1419*	*greA*	Elongation factor		
		3	c	CCCAaTTTCaCAAg	S	62	*spj_1560*	*ftsZ*	Cell division protein		
		3	w	CCaAaTTTCtCAAC	O	293	*spj_1645*	*spJ_1646*	HAD phosphatase, family IIIA		
		3	w	CCCATTTTggaAAC	O	89	*spj_1880*	*spJ_1881*	SSB single-strand binding protein family		
		2	c	CCCATTTgaCCAAC	S	229	*spj_2091*	*pgi*	Glucose-6-phosphate isomerase		
		2	w	CCCATTTTtCCAAg	O	158	*spj_2185*	*spJ_2186*	Alpha-L-fucosidase (F-type lectins)		
ATCC 700669	*spn23f__09790*	3	c	CCCtTTTTaCCAAg	S	165	*NA*	*NA*	*pepSHOW*	C(504144..504236)	MPKTPLGALTRGTIRFHLTSQTFISKSMRK
		3	c	CCCtaTTTCCaAAC	S	49	*spn23f_06340*	*glnP*	Glutamine ABC transporter permease		
		3	c	CCCcaTTTCaCAAC	O	116	*spn23f_08560*	*proA*	Gamma-glutamyl phosphate reductase		
		3	w	CCCATTTcaCaAAC	S	273	*spn23f_10620*	*rnhB*	Ribonuclease HII		
		3	w	CCCtgTTTCtCAAC	O	23	*spn23f_11010*	*recN*	RecN		
		3	c	CCCAcTTgCCaAAC	S	127	*spn23f_11230*	*mutY*	A/G-specific adenine glycosylase		
		3	w	CCCATTTTttCAgC	O	31	*spn23f_14710*	*atpD*	F0F1 ATP synthase subunit beta		
		3	w	CCatTTTTCtCAAC	O	10	*spn23f_14810*	*greA*	Elongation factor		
		3	c	CCCAaTTTCaCAAg	S	62	*spn23f_16660*	*ftsZ*	Cell division protein		
		3	w	CCaAaTTTCtCAAC	O	293	*spn23f_17560*	*spN23F_17570*	HAD phosphatase, family IIIA		
		3	c	CCaAaTTTtCCAAC	S	220	*spn23f_17900*	*spN23F_17910*	Hyp. Protein		
		2	c	CCCATTTgaCCAAC	S	229	*spn23f_20940*	*pgi*	Glucose-6-phosphate isomerase		
		2	w	CCCATTTTtCCAAg	O	158	*spn23f_21910*	*spN23F_21920*	Alpha-L-fucosidase (F-type lectins)		

Annexe 7 - Tableau B : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe II.

Strains	Gène *rgg*	Cibles nM	Brin	Target_motif	Dir	dATG	Locus_tag	Gène_tag	Fonction	spCDS identifiées dans bactgeneSHOW Localisation	Séquence
S. thermophilus											
				GCATATATGGGA		**75**					
LMD-9	*ster_1358*	2	c	GCcaATATGGGA	S	43	*ster_1405*	*amiE*			
		2	w	GgcTATATGGGA	O	40	*ster_1811*	*ster_1812*	5-formyltetrahydrofolate cyclo-ligase		
				GCATATATGGGA		**72**					
LMD-9	*ster_1530*	2	c	GCcaATATGGGA	S	43	*ster_1405*	*amiE*			
		2	w	GgcTATATGGGA	O	40	*ster_1811*	*ster_1812*	5-formyltetrahydrofolate cyclo-ligase		
LMG18311	*stu1572*	2	c	GCtgATATGGGA	O	232	*stu0048*	*argS*	Arginyl-tRNA synthetase		
		2	w	GCcTATAgGGGA	S	93	*stu0137*	*stu0136*	MutT/nudix family protein		
		2	c	GCcaATATGGGA	S	28	*stu1438*	*amiE*			
		2	c	GaATATATGcGA	S	77	*stu1801*	*stu1802*	CoA binding domain protein		
CNRZ1066	*str1572*	2	c	GCcaATATGGGA	S	28	*str1438*	*amiE*			
		2	c	GaATATATGcGA	S	77	*str1801*	*str1802*	CoA binding domain protein		
				GCATATACGAGA		**70**					
LMD-9	*ster_0934*	2	w	GCcaATACGAGA	S	138	*ster_0586*	*ster_0586*	Molybdopterin/thiamine biosynthesis family protein		
		2	w	GCAcATAaGAGA	O	165	*ster_1715*	*ster_1716*	Late competence protein		
LMG18311	*stu0917*	2	w	GCcaATACGAGA	S	103	*stu0546*	*stu0545*	Molybdopterin biosynthesis protein		
		2	w	GCAcATAaGAGA	O	165	*stu1740*	*comEB*	Late competence protein		
		2	w	GaATATAgGAGA	O	250	*NA*			C(786469..786564)	MPWASGLLISLNLFFLITLYQILSKLDEENK
CNRZ1066	*str0917*	2	w	GCcaATACGAGA	S	103	*str0546*	*str0545*	Molybdopterin biosynthesis protein		
		2	w	GCAcATAaGAGA	O	165	*str1740*	*comEB*	Late competence protein		
		2	w	GaATATAgGAGA	O	255	*NA*			C(792889..792978)	MGTGLLISLNLFFLITLYQILSKLDEENK
S. mutans											
				GTCTATATGGGA		**74**					
UA159	*smu.1509*	2	c	GTCTATAaGGGc	S	152	*smu_1085*	*kitH*	Thymidine kinase		
		2	w	GgCaATATGGGA	O	232	*smu_1210*	*smu_1211*	Putative glycerol-3-phosphate acyltransferase PlsY		
		2	c	GTCTATcTaGGA	O	123	*smu_1579*	*birA*	Biotin operon repressor		
		2	c	GatTATATGGGA	S	130	*smu_1711*	*smu_1712c*	Segregation and condensation protein B		
		2	c	GTgTATcTGGGA	O	175	*smu_1879*	*ptnC*	PTS system, mannose		
		2	c	GTCTATAaaGGA	O	16	*NA*				
nn2025	*smunn2025_0595*	2	w	GTgTATcTGGGA	O	175	*smunn2025_0263*	*ptnC*	PTS system, mannose		
		2	w	GatTATATGGGA	S	130	*smunn2025_0410*	*scpB*	Segregation and condensation protein B		
		2	w	GTCTATcTaGGA	O	123	*smunn2025_0520*	*NA*		C(552648..552848)	MAKLVFLLGSRQDGFGKLADNISLTQTAKEQNPPKTSPKELTFKN RIIKKNSFSHSRYQHDNEVLF
		2	c	GgCaATATGGGA	O	232	*smunn2025_0843*	*smunn2025_0842*	Putative glycerol-3-phosphate acyltransferase PlsY		
		2	w	GTCTATAaGGGc	S	152	*smunn2025_0951*	*kitH*	Thymidine kinase		
		2	c	GatTATATGGGA	S	6	*smunn2025_1077*	*NA*			
		2	c	GTCTATAaaGGA	O	16	*NA*				
S. dysgalactiae											
				GCATATATGGGA		**75**					
subsp. equisimilis GGS_124	*sdeg_1013*	2	w	GCATATAaaGGA	S	191	*sdeg_0357*	*oppA*			
		2	w	GatTATATGGGA	S	213	*sdeg_0420*	*scpB*	Segregation and condensation protein B		
		2	w	GCtgATATGGGA	O	294	*sdeg_1451*	*sdeg_1452*	Methyltransferase		
		2	c	GaAaATATGGGA	S	298	*sdeg_1462*	*ddl*	D-alanylalanine synthetase		
		2	w	GCATATAaGaGA	O	96	*sdeg_1491*	*feoB*	Ferrous ion transport protein B		
		2	c	GCgTATcTGGGA	S	251	*sdeg_1832*	*sdeg_1833*	Phosphoglycerate mutase		
		2	c	GaAaATATGGGA	S	6	*sdeg_1870*	*scrA*	PTS system, sucrose		
S. equi											
				GTACATACGGGA		**75**					
MGCS10565	*sez_0716*	2	c	GccCATACGGGA	S	265	*sez_0290*	*Sez_0291*	NAD(FAD)-utilizing dehydrogenase		
		2	w	GTACATAgGtGA	O	279	*sez_1630*	*ypaA*	Riboflavin transporter YpaA-like		
subsp. zooepidemicus	*szo_12400*	2	c	GTACATAgGtGA	O	255	*szo_03430*				
								szo_03420	Riboflavin transporter YpaA-like		
		2	w	GTAaATAgGGGA	O	48	*szo_19190*	*cbiO*	Cobalt transporter ATP-binding subunit		
				GCATATATGGGA		**75**					
subsp. zooepidemicus	*szo_17730*	2	w	GCAgATAaGGGA	S	193	*szo_00100*	*NA*			

		1	c	GCATATAaGGGA	O	176	*szo_03430*	*szo_03420*	Riboflavin transporter YpaA-like		
		2	w	GgATATATGaGA	S	262	*szo_08560*	*szo_08550*	Resolvase		
		2	c	GacTATATGGGA	S	131	*szo_11170*	*szo_11180*	Rhamnan synthesis protein F family protein		
		2	c	GCcaATATGGGA	S	297	*szo_12560*	*szo_12570*	Pseudogène		
S. suis											
				GTATATAAGGGA		**92**					
SC84	*ssusc84_1811*	2	c	GcAgATAAGGGA	S	263	*ssusc84_1491*	*ssuSC84_1492*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssusc84_1553*	*ssuSC84_1554*	Histone acetyltransferase		
GZ1	*ssgz1_1813*	2	c	GcAgATAAGGGA	S	254	*ssgz1_1482*	*ssgz1_1483*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssgz1_1547*	*ssgz1_1548*	Histone acetyltransferase		
P1/7	*ssu1789*	2	c	GcAgATAAGGGA	S	263	*ssu1462*	*ssu1463*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssu1527*	*ssu1528*	Histone acetyltransferase		
98HAH33	*ssu98_2002*	2	c	GTATATAcaGGA	O	263	*ssu98_0667*	*ssu98_0666*	Histone GNAT family acetyltransferase	647435..647512	MKTTVGYSEIINEIKKEKAAAIPPF
		2	c	GcAgATAAGGGA	S	254	*ssu98_1661*	*ssu98_1662*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssu98_1730*	*ssu98_1731*	Histone acetyltransferase		
05ZYH33	*ssu05_1997*	2	c	GTcaATAAGGGA	O	7	*ssu05_0386*	*ssu05_0385*	Hyp. Protein		
		2	w	GaATATAAGcGA	S	287	*ssu05_0871*	*ssu05_0870*	Hyp. Protein		
		2	c	GcAgATAAGGGA	S	254	*ssu05_1651*	*ssu05_1652*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssu05_1718*	*ssu05_1719*	Histone acetyltransferase		
BM407	*ssubm407_1859*	2	w	GTAaATAAGcGA	S	199	*ssubm407_1005*	*ssubm407_1004*	HAD superfamily hydrolase		
		2	c	GcAgATAAGGGA	S	263	*ssubm407_1538*	*ssubm407_1539*	Arginine repressor		
		2	c	GccTATAAGGGA	S	275	*ssubm407_1601*	*ssubm407_1602*	Histone acetyltransferase		

Annexe 7 - Tableau C : Gènes cibles putatifs des systèmes SHP/Rgg appartenant au groupe III.

		Cibles								spCDS identifiées dans bactgeneSHOW	
Strains	**Gène *rgg***	**nM**	**Brin**	**Target_motif**	**Dir**	**dATG**	***Locus_tag***	**Gène**	**Fonction**	**Localisation**	**Séquence**
S. thermophilus											
				ATnTCnnATnTnCnATTTT		**75**					
LMD-9											
S. pneumoniae											
				ATnTCnnATnTnCnATTTT		**77**					
ATCC 700669	*spn23f_01510*	0	c	ATGTCCAATATGCAATTTT	S	266	*spn23f22630*		ABC transporter		
JJA	*spj_0171*	0	c	ATGTCCAATATGCAATTTT	S	266	*spj_2256*		ABC transporter		
Hungary19A-6	*sph_0253*	0	c	ATGTCCAATATGCAATTTT	S	266	*sph_2425*		ABC transporter		
TIGR4	*sp_0141*	0	c	ATGTCCAATATGCAATTTT	S	266	*sp_2230*		ABC transporter		
P1031	*spp_0211*	0	c	ATGTCCAATATGCAATTTT	S	266	*spp_2282*		ABC transporter		
CGSP14	*spcg_0143*	0	c	ATGTCCAATATGCAATTTT	S	266	*spcg_2196*		ABC transporter		

Annexe 8

Valorisation du travail de thèse.

Publication scientifique

B. Fleuchot, C. Gitton, A. Guillot, J. Vidic, P. Nicolas, C. Besset, L. Fontaine, P. Hols, V. Monnet and R. Gardan (2011) Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol Microbiol* 80: 1102-1119.

Communications orales

6th International Conference on Gram-Positive Microorganisms, juin 2011

« Rgg-associated small hydrophobic peptides: a new family of pheromones involved in quorum-sensing in streptococci. » B. Fleuchot, C. Gitton, A. Guillot, J. Vidic, P. Nicolas, C. Besset, L. Fontaine, P. Hols, V. Monnet, **R. Gardan***.

17ème colloque du Club des Bactéries Lactiques, octobre 2010

« Molecular basis for a new quorum-sensing mechanism in streptococci. » **B. Fleuchot***, C. Gitton, A. Guillot, C. Besset, V. Monnet, R. Gardan.

Séminaires de présentation des travaux de recherche

INRA La Minière, équipe de D. Lereclus, février 2010

« Bases moléculaires d'un nouveau mécanisme de quorum-sensing chez les streptocoques. » **B. Fleuchot***, R. Gardan.

Posters

10th Symposium on Lactic acid bacteria, août-septembre 2011

« Rgg-associated small hydrophobic peptides: a new family of pheromones involved in quorum-sensing in streptococci. » B. Fleuchot, C. Gitton, A. Guillot, J. Vidic, P. Nicolas, C. Besset, L. Fontaine, P. Hols, **V. Monnet***, R. Gardan.

4th Congress of European Microbiologists, FEMS, juin 2011

« Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. » **B. Fleuchot***, C. Gitton, A. Guillot, J. Vidic, C. Besset, L. Fontaine, P. Hols, N. Leblond-Bourget, V. Monnet, R. Gardan.

« Do the multiple rgg/shp pairs found in streptococci genomes allow a crosstalk between different species? » **B. Fleuchot***, P. Nicolas, A. Guillot, V. Monnet, R. Gardan.

6th International Conference on Gram-Positive Microorganisms, juin 2011

« Do the SHP pheromones allow cell-cell communication between different streptococci species? » **B. Fleuchot***, P. Nicolas, A. Guillot, V. Monnet, R. Gardan.

17ème colloque du Club des Bactéries Lactiques, octobre 2010

« Global analysis of the *shp*/*rgg* quorum sensing mechanisms. » **B. Fleuchot***, V. Monnet, R. Gardan.

Jacques Monod Meeting, Institut Pasteur, mai 2010

« Molecular basis for a new quorum-sensing mechanism in streptococci. » **B. Fleuchot***, C. Gitton, A. Guillot, C. Besset, V. Monnet, R. Gardan.

Journées des Microbiologistes de l'INRA, mai 2010

« Bases moléculaires d'un nouveau mécanisme de quorum-sensing chez les streptocoques. » **B. Fleuchot***, C. Gitton, A. Guillot, C. Besset, V. Monnet, R. Gardan.

16ème colloque du Club des Bactéries Lactiques, mai 2009

« Un nouveau mécanisme de régulation par quorum sensing chez *Streptococcus thermophilus*. » **B. Fleuchot***, C. Gitton, C. Besset, L. Fontaine, P. Hols, V. Monnet, R. Gardan.

Annexe 9

Formations.

Formations de l'école doctorale ABIES

- Anglais scientifique « White Right », 21h, mai 2009.
- Préparation à l'insertion professionnelle : la thèse, au service de mon projet professionnel et personnel, 21h, novembre 2009.
- Communiquer sur sa thèse : l'exposé oral efficace, 14h, novembre 2009.
- Anglais scientifique « Poster power », 21h, avril 2010.
- Préparation à l'insertion professionnelle : objectif premier emploi, 21h, janvier 2011.

Formations INRA

- Maîtriser l'information scientifique et technique en recherche, 7h, mars 2009.
- Séminaire réflexive, 35h, mars 2009.
- Sauveteur secouriste du travail, 14h, juin 2009, 2010
- Savoir manipuler les extincteurs, 2h, juin 2009.

Abstract

The discovery of a genetic context – encoding a small hydrophobic peptide (SHP) and a transcriptional regulator belonging to the Rgg family (in nearly all streptococcal genomes) –, following by the study of one of this loci in *S. thermophilus* LMD-9, led to the hypothesis that the regulatory proteins Rgg in association with a putative pheromone SHP could define a novel quorum-sensing (QS) regulatory mechanism in Gram-positive bacteria.

The first part of my PhD consisted to validate this hypothesis. For this purpose, we analyzed the SHP/Rgg system in all the steps that are commonly involved in QS mechanisms: (i) secretion of the putative pheromone, (ii) maturation of the pheromone, (iii) capture of the pheromone from the external environment at a threshold concentration, (iv) importation of the pheromone inside the cell and (v) interaction of the transcriptional regulator to the promoter regions of targeted genes. Experimentally, we focused on the so-called *shp/rgg1358* locus of *S. thermophilus* LMD-9, which is the streptococcal species containing the largest number of *shp/rgg* pairs in its genome. By using genetic and biochemistry approaches, we uncovered a new QS mechanism that involves the pheromone SHP, the oligopeptide transporter AmiCDEF for the uptake of the pheromone and the transcriptional regulator Rgg for the control of target gene expression. Furthermore, we showed that the membrane protease Eep participates in the production of the mature pheromone, which has been identified by mass spectrometry. Once characterized, the second part of my PhD was to explore the functionality of this new QS system in other streptococcal strain or species, in order to determine if cross-reactivity phenomenon between streptococci can occur. By using heterologous expression in *S. thermophilus* LMD-9, we extended the functionality of the SHP/Rgg system to two pathogenic streptococcal species, *i.e. S. agalactiae* and *S. mutans*. The last part of my PhD consisted in identifying the regulon of all SHP/Rgg systems. Following the construction of a phylogenetic tree of the Rgg-like proteins in low GC Gram-positive bacteria, we identified 68 SHP/Rgg systems that we classified in three groups. Analyzing the promoter regions of all *shp* genes led to the identification of a putative Rgg DNA binding site specific to each SHP/Rgg group. An *in silico* approach was also used to scan all sequenced streptococcal genomes for the three identified patterns. Whereas proximal target genes were detected for groups II and III, distal target genes were found in groups I and II. In addition, we uncovered that putative Rgg DNA binding sites can be localized in coding or non-coding region. Currently, validations are in progress.

To sum-up, my PhD studies provided evidences that the Rgg proteins in association with small peptide pheromones define a new QS mechanism that seems to regulate the expression of distal and proximal genes in a species-dependent manner. Important insights should be obtained concerning a putative crosstalk among streptococci that involves the SHP/Rgg QS system. My studies may constitute a basis for the development of small peptides to optimize the use of *S. thermophilus* in dairy factories and reduce the virulence of pathogenic streptococci.

Key words : Quorum-sensing, streptococci, Short Hydrophobic Peptide, Rgg, Oligopeptides transporter Opp.

Résumé

La découverte d'un contexte génétique chez les streptocoques – codant un petit peptide hydrophobe (SHP) et un régulateur transcriptionnel appartenant à la famille Rgg –, suivi de l'étude d'un de ces loci chez *S. thermophilus* LMD-9, a conduit à l'hypothèse que les protéines régulatrices Rgg en association avec une phéromone putative SHP pourraient intervenir dans un mécanisme de type quorum-sensing (QS) chez les bactéries à Gram positif.

La première partie de ma thèse a consisté à confirmer cette hypothèse sur le locus *shp/rgg1358* de *S. thermophilus* LMD-9, espèce contenant le plus grand nombre de systèmes SHP/Rgg dans son génome. Pour ceci, les étapes impliquées dans un mécanisme de QS ont été étudiées : la sécrétion, la maturation et la détection à une concentration seuil de la phéromone, sa réimportation à l'intérieur de la cellule, son interaction avec un régulateur transcriptionnel et enfin l'interaction de la protéine régulatrice à l'ADN. Par l'utilisation d'approches génétiques et biochimiques, nous avons démontré l'existence d'un nouveau mécanisme de QS impliquant pour la première fois un régulateur transcriptionnel Rgg et une phéromone SHP, importée à l'intérieur de la cellule par le transporteur d'oligopeptides AmiCDEF. Le rôle de la protéase membranaire, Eep, a également été démontré dans la maturation de la phéromone, dont la forme mature a été déterminée par spectrométrie de masse et validée *in vivo*.

Dans un second temps, nous avons exploré la fonctionnalité de ce nouveau mécanisme sur d'autres loci *shp/rgg*, dans le but d'étudier l'existence d'éventuels phénomènes de cross-talk entre les bactéries. L'étude de nouveaux loci, en système hétérologue chez *S. thermophilus* LMD-9, a permis d'étendre la fonctionnalité du mécanisme à deux systèmes SHP/Rgg de streptocoques pathogènes, à savoir *S. agalactiae* et *S. mutans*.

En parallèle à ce travail de caractérisation, l'identification des régulons des systèmes SHP/Rgg a été entreprise. La construction d'un arbre phylogénétique des protéines Rgg-like a permis d'identifier 68 systèmes SHP/Rgg, que nous avons classés en trois groupes. L'analyse des régions promotrices des gènes *shp* a conduit à l'identification d'un site putatif de liaison des protéines Rgg à l'ADN spécifiques de chaque groupe SHP/Rgg. Une approche *in silico* a ensuite été menée afin de rechercher, dans les génomes séquencés de streptocoques, les gènes cibles putatifs. Alors que des cibles proximales ont été détectées pour les groupes II et III, des cibles distales ont été identifiées dans les groupes I et II. Actuellement, la validation de certaines cibles est en cours au laboratoire.

A l'avenir, ce travail pourrait permettre le développement de petits peptides permettant d'optimiser l'utilisation de *S. thermophilus* en industries laitières et de réduire la virulence des streptocoques pathogènes.

Discipline : Microbiologie.

Mots clés : Quorum-sensing, streptocoques, petit peptide hydrophobe SHP, Rgg, Transporteur d'oligopeptides Opp,