



**HAL**  
open science

# Mécanismes moléculaires de la biogenèse du pilus chez *Streptococcus pneumoniae*

Lamya El Mortaji

► **To cite this version:**

Lamya El Mortaji. Mécanismes moléculaires de la biogenèse du pilus chez *Streptococcus pneumoniae*. Biochimie [q-bio.BM]. Université de Grenoble, 2010. Français. NNT : . tel-00567141

**HAL Id: tel-00567141**

**<https://theses.hal.science/tel-00567141>**

Submitted on 18 Feb 2011

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

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



UNIVERSITÉ DE GRENOBLE

## THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE**  
Spécialité **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

**EL MORTAJI Lamya**

le **10 décembre 2010**

---

**MECANISMES MOLECULAIRES DE LA BIOGENESE DU PILUS CHEZ  
*STREPTOCOCCUS PNEUMONIAE***

---

Thèse dirigée par **Anne Marie Di Guilmi**

## JURY

**Mr Jamin Marc**  
**Mme Dramsi Shaynoor**  
**Mr O'Callaghan David**  
**Mr Terradot Laurent**  
**Mr Grangeasse Christophe**

**Professeur à l'UJF**  
**CR à l'Institut Pasteur**  
**DR INSERM Espri26**  
**CR à l'IBCP**  
**DR à l'IBCP**

**Président**  
**Rapporteur**  
**Rapporteur**  
**Examineur**  
**Examineur**

Thèse préparée au sein de **l'Institut de Biologie Structurale Jean Pierre Ebel - UMR 5075**  
dans l' **Ecole Doctorale Chimie et Sciences du Vivant**



## RESUME EN FRANÇAIS :

*Streptococcus pneumoniae* est un pathogène majeur chez l'homme, responsable d'otites, de pneumonies, de septicémies et de méningites. Récemment des structures de type pilus ont été identifiées à la surface de *S. pneumoniae* et jouent un rôle important dans les étapes initiales de colonisation des tissus hôtes. Six gènes sont impliqués dans la formation de cette structure. Trois d'entre eux codent pour les protéines structurales ou pilines (RrgA, RrgB et RrgC) et trois autres gènes codent pour les enzymes, appelées sortases, qui catalysent l'association covalente des pilines (SrtC-1, SrtC-2 et SrtC-3). Des modèles de formation du pilus ont été proposés suite à des études de délétion génétique, mais aucune donnée biochimique permettant d'expliquer précisément la formation du pilus au niveau biomoléculaire n'est encore disponible.

L'étude individuelle des protéines impliquées dans la formation du pilus a permis la mise en évidence de ponts intramoléculeurs Lys-Asn stabilisateurs présents dans chacune des pilines. De plus, la résolution cristallographique de RrgA et RrgB permet de mieux comprendre les propriétés adhésives de cette structure mais également son mécanisme d'assemblage.

Comme le rôle de chacune des sortases reste imprécis, nous avons développé un système de co-expression permettant de tester toutes les combinaisons de pilines et de sortases. Celui-ci nous a permis d'identifier les spécificités de chacune des sortases, de générer des complexes covalents piline/piline mais également piline/sortase et ainsi d'obtenir des éléments clés dans la compréhension de la biogenèse de cette structure.

**TITRE EN ANGLAIS:** Molecular Mechanisms of the pneumococcal pilus biogenesis

## RESUME EN ANGLAIS

*Streptococcus pneumoniae* is the most common cause of otitis, sinusitis, pneumonia, sepsis and meningitis. Recently, pili were identified at the surface of *S. pneumoniae* and were proposed to play a role in the initial step of host tissues colonisation. Six genes are involved in the pilus formation. Three of them encode structural proteins named pilins (RrgA, RrgB and RrgC), the three others encode specific enzymes called sortases, which catalyse the covalent association of the pilins (SrtC-1, SrtC-2 and SrtC-3). Pilus formation models have been proposed based on genetic studies, but no biochemical data explaining precisely the molecular process of pilus formation is yet available.

The individual study of each pilin led to the identification of stabilising Lys-Asn intramolecular bonds in each protein. Moreover, the crystallographic structure of RrgA and RrgB provided precious informations regarding the adhesive properties as well as the mechanism of pilus assembly.

Since the role of each sortase remains unclear, our aim was to elucidate, step by step, the molecular mechanisms involved in the pilus biogenesis. Therefore, we have developed a co-expression platform allowing the production of the pilins subunits together with sortases in *E.coli*. This system allowed to decipher the substrate specificity of the sortases, to generate covalent pilin/pilin complexes, as well as pilin/sortase complexes and thus provided key information towards the understanding complex macromolecular process.

## MOTS-CLES/KEY WORDS

*S. pneumoniae*, pilus, pilines, sortases, ponts intramoléculeurs / *S. pneumoniae*, pilus, pilins, sortases, intramolecular bonds

## INTITULE ET ADRESSE DU LABORATOIRE

Institut de Biologie Structurale Jean Pierre Ebel (UMR 5075)

41, rue Jules Horowitz 38027 Grenoble Cedex 1



« *Force et honneur* »  
Devise romaine

*A ma famille,*



## REMERCIEMENTS

*Mes travaux de thèse ont été réalisés au sein du « Groupe Pneumocoque » dirigé par Thierry Vernet, à l'Institut de Biologie Structurale de Grenoble, sous la direction scientifique d'Anne Marie Di Guilmi.*

*Je souhaitais exprimer toute ma gratitude à Thierry Vernet pour m'avoir accueillie dans son laboratoire, pour sa confiance dans mon travail et pour s'être montré suffisamment réceptif à mon humour parfois si particulier.*

*Je remercie vivement Mme Shaynoor Dramsi, chargée de recherche à l'Institut Pasteur (Paris) et Mr. David O'Callagan, directeur de recherche à l'UFR de médecine de Nîmes d'avoir accepté d'être les rapporteurs de cette thèse. Aussi, je suis très reconnaissante envers le Pr. Marc Jamin de l'UVHCI de Grenoble, ainsi que le Dr. Laurent Terradot et le Dr. Christophe Grangeasse de l'IBCP de Lyon, pour l'honneur qu'ils me font de juger ce travail de thèse.*

*Je souhaitais tout particulièrement remercier ma directrice de thèse, Anne Marie Di Guilmi. Merci pour ta confiance, merci pour tes conseils avisés, merci pour ton écoute, et mille merci pour ton soutien à chacune des étapes de cette thèse.*

*Je souhaitais chaleureusement remercier toutes les personnes qui ont directement participé à la réussite de ce projet: Andréa Dessen, Clothilde Manzano, Thierry Izoré et Carlos Contreras-Martel. Un grand merci également à tous les stagiaires qui ont participé au projet « pilus », avec par ordre d'apparition: Rémi Terrasse, Lorraine Benier, Tony Fourny et Maria Fehringer...Et je tenais également à remercier Daphna Fenel et Guy Schoen, Jean-Pierre Andrieu, Isabel Bérard, Luca Signor et Eric Forest, Alexandra Kraut et Jérôme Garin.*

*Je remercie infiniment tous les membres du « Groupe Pneumocoque », passés et présents, pour leurs conseils, leur gentillesse, leur joie de vivre et pour avoir amélioré chaque jour l'ambiance au sein du laboratoire: Cécile Morlot, Claire Durmort, Audrey Legouellec, André Zapun, Benoit Gallet, Nordine Hellassa...Je n'oublie pas les vrais blondes et assimilées « Kinders » qui m'auront énormément fait rire durant ces trois années: Marjolaine Noirclerc-Savoie, Laure Roux, Elodie Loisel ainsi que les personnalités particulièrement attachantes: Soizic Masson, Cécile Attali, Florent Bernaudat et Julien Offant... Je remercie aussi tout particulièrement Cécile Frolet pour tous ses bons conseils et son soutien... Un grand merci également à tous les joyeux stagiaires qui auront fait un passage au laboratoire... Ainsi que les membres de l'institut qui ont rendu ces trois années plus agréables encore.*

*Je remercie infiniment tous les membres de ma famille pour leur soutien sans faille, mes parents, sans qui je ne serai jamais arrivée où je suis maintenant, mon frère Itab, ma sœur Fadoit et ma Lobna, parce qu'ils sont juste formidables...Et bien sûr tous mes amis, les membres de la Dream Team et de Radio Potins, avec une pensée particulière pour Audrey Albouy.*





# SOMMAIRE

REMERCIEMENTS	7
SOMMAIRE	9
TABLE DES ILLUSTRATIONS	13
TABLE DES ABREVIATIONS	15
INTRODUCTION	17
I <i>Streptococcus pneumoniae</i>	17
I.1 Présentation générale	17
I.1.1 Classification	17
I.1.2 Caractéristiques biologiques	18
I.2 La paroi bactérienne	19
I.2.1 La paroi de <i>S. pneumoniae</i>	20
I.2.1.1 Le peptidoglycane	20
I.2.1.2 Les acides teïchoïques et lipoteïchoïques	21
I.2.2 La capsule de <i>S. pneumoniae</i>	22
I.3 Données épidémiologiques	24
I.3.1 Prophylaxie	25
I.3.2 Moyens curatifs	26
I.3.3 Difficultés rencontrées dans le traitement et la prévention des infections à pneumocoque	27
I.3.3.1 Résistance aux antibiotiques	27
I.3.3.2 Les limites de la protection vaccinale	28
II. La pathogénicité du pneumocoque	31
II.1 La capsule	32
II.2 La pneumolysine	33
II.3 Les protéines de surface et leur rôle dans la virulence	34
II.3.1 La GAPDH : un exemple de protéine « moonlighting »	36
II.3.2 CBPE : un exemple de Choline-Binding Protein	37
II.3.2.1 Structure de la protéine	38
II.3.2.2 Rôle de la protéine dans la virulence	38
II.3.3 AdcAII : un exemple de lipoprotéines	39
II.3.3.1 Structure de la protéine	40
II.3.3.2 AdcAII est une adhésine	41
II.3.4 les protéines LPxTG	41
II.3.4.1 Topologie des protéines LPxTG	42
II.4 Les sortases	43
II.4.1 La sortase A de <i>S. aureus</i>	43
II.4.1.1 Identification	43
II.4.1.2 Le mécanisme de transpeptidation	43
II.4.1.3 Structure tridimensionnelle	45
II.4.1.4 Paramètres enzymatiques	46
II.4.2 La classification des sortases chez les bactéries Gram-positif	47
II.4.3 Les enzymes sortases chez <i>S. pneumoniae</i>	49
III. Structure et fonction des pili	51
III.1 Identification des pili chez les bactéries Gram-positif	52
III.1.1 Identification des pili chez le pneumocoque	53

III.1.2 Organisation génétique des pili chez les Gram-positifs	54
III.2 La fonction des pili	55
III.2.1 Chez le pneumocoque	55
III.2.2 Chez d'autres bactéries Gram-positives	57
III.3 Régulation transcriptionnelle du PI-1 chez <i>S. pneumoniae</i>	57
III.3.1 <i>rlrA</i> : l'activateur	57
III.3.2 <i>mgrA</i> : le répresseur	58
III.3.3 Les autres régulateurs du PI-1 chez le pneumocoque	58
III.4 Les pilines	60
III.4.1 Pilines majeures	61
III.4.2 Pilines accessoires	63
III.4.3 Des ponts intramoléculaires covalents au sein des pilines	65
III.4.4 Organisation structurale des pili	68
III.5 Structure et assemblage du pilus chez <i>S. pneumoniae</i>	69
III.5.1 Les pilines du PI-1	69
III.5.2 Modèles d'assemblage du pilus chez <i>S. pneumoniae</i>	71
III.6 Rôle des sortases	75
III.6.1 <i>C. diphtheriae</i> : les premiers éléments de réponse	75
III.6.2 Pourquoi trois sortases chez <i>S. pneumoniae</i> ?	77
III.6.2.1 la polymérisation de RrgB	77
III.6.2.2 l'incorporation des pilines accessoires	78
III.6.2.3 les structures des sortases de classe C	79
III.7 Conclusion	80
 MATERIELS & METHODES	 83
I. Souches, vecteurs et milieux de culture	83
I.1 Les souches	83
I.2 Les vecteurs	83
I.3 Les constructions	85
I.3.1 Constructions dans le vecteur pLIM01	85
I.3.2 Constructions dans les vecteurs pETDuet®, pACYCDuet® et pCDFDuet®	86
I.3.4 Récapitulatif des vecteurs construits pour cette étude	88
I.3.5 Récapitulatif des souches d'expression réalisées pour cette étude	88
I.4 Milieux de culture et antibiotiques	89
I.4.1 Milieux de culture	89
I.4.2 Antibiotiques	89
II. Techniques de biologie moléculaire	90
II.1 Clonage	90
II.2 Mutagenèse	91
II.3 Préparation des cellules « Ultra-compétentes » <i>E. coli</i> STAR	92
II.4 Transformation des cellules compétentes	92
II.5 Mini préparation plasmidique et séquençage	93
III. Techniques biochimiques	94
III.1 Production des protéines recombinantes en grand volume	94
III.2 Production des protéines recombinantes en mini-cultures	94
III.3 Purification des protéines recombinantes	94
III.4 Minipurification des surnageants de co-expressions	96
III.5 Analyses des protéines	96
III.5.1 Analyse des protéines par électrophorèse en conditions dénaturantes	96
III.5.2 Coloration des protéines au bleu de Coomassie	96

III.5.3 Transfert des protéines sur membrane de nitrocellulose	96
III.5.4 Immunorévélation des protéines sur membrane de nitrocellulose	97
III.5.7 Tests de stabilité par protéolyse à la trypsine et séquençage N-terminal	97
III.5.8 Tests de stabilité par Thermal Shift Assay (TSA)	98
IV Principes	99
IV.1 Principe du Thermal Shift Assay	99
IV.2 Principe de la plateforme de co-expression chez <i>E. coli</i>	100
 PRESENTATION DES TRAVAUX	 101
 RESULTATS	 103
I. Analyse structurale des pilines de <i>S. pneumoniae</i>	103
I.1 Structure de la piline accessoire RrgA	104
I.1.1. Bases structurales de la fonction d'adhésion de RrgA	104
I.1.2. Replissements de type IgG et CNA	107
I.1.3. Identification de deux ponts intramoléculaires dans RrgA	107
I.1.4. Rôle de stabilisation des ponts intramoléculaires	109
I.1.5. Analyse de la protéine recombinante RrgA par microscopie électronique à transmission	110
I.1.6. Production du domaine D4	111
I.2 Etude de la piline accessoire RrgC	112
I.2.1. Topologie de RrgC	112
I.2.2. Mise en évidence des ponts intramoléculaires de RrgC	114
I.2.3. Les deux ponts intramoléculaires stabilisent RrgC	115
I.2.3.1 Sensibilité à la digestion protéolytique	115
I.2.3.2 Stabilité thermique	117
I.3 Etude de la piline majeure RrgB	119
I.3.1. Topologie de RrgB	119
I.3.2. Mise en évidence des ponts intramoléculaires de RrgB	120
I.3.3. Les trois ponts intramoléculaires stabilisent RrgB	121
I.3.3.1 Sensibilité à la digestion protéolytique	121
I.3.3.2 Stabilité thermique	122
I.3.4. Structure cristallographique de la piline majeure RrgB	123
I.4 Conclusions concernant l'étude individuelle des pilines	127
II. Détermination du rôle des sortases	129
II.1 Fonction de la sortase C-1	129
II.1.1 La sortase C-1 associe covalentement RrgB et RrgC	129
II.1.2 La sortase C-1 reconnaît le motif IPQTG de RrgB	131
II.1.3 Les ponts intramoléculaires de RrgB jouent un rôle dans l'efficacité de reconnaissance des pilines par la SrtC-1	133
II.1.4 Conclusion	135
II.2 Fonction de la sortase C-2	136
II.2.1. La SrtC-2 catalyse deux complexes covalents différents	136
II.2.2. La SrtC-2 reconnaît le motif YPRTG de RrgA et le motif IPQTG de RrgB	137
II.2.3. Substrat privilégié de la SrtC-2 : le motif YPRTG de RrgA	138
II.2.4. Liaison intermoléculaire entre RrgA-RrgB	139
II.2.5. Rôle des ponts isopeptidiques dans la formation du complexe RrgA-RrgB	140
II.2.6. SrtC-2 forme un complexe acyl-enzyme stable avec RrgA	141
II.2.7. SrtC-2 multimérise le domaine D4 de RrgA	142
II.2.8. SrtC-2 multimérise RrgA	144

II.2.9 Conclusions concernant l'activité de la SrtC-2	146
II.3 Fonction de la SrtC-3	149
II.3.1. La SrtC-3 catalyse deux complexes: RrgB-RrgA et RrgB-RrgC	149
II.3.2. La SrtC-3 reconnaît le motif IPQTG de RrgB	150
II.3.3. La SrtC-3 forme un complexe acyl-enzyme stable avec RrgC	151
II.3.4 Conclusions concernant l'activité de la SrtC-3	153
DISCUSSION & PERSPECTIVES	157
REFERENCES BIBLIOGRAPHIQUES	165
ANNEXES	179
Annexe 1 : Récapitulatif des vecteurs construits pour cette étude	179
Annexe 2 : Liste des oligonucléotides utilisés pour réaliser les mutants	181
Annexe 3 : Récapitulatif des souches d'expression réalisées pour cette étude	182
PUBLICATIONS	185

## TABLE DES ILLUSTRATIONS

Figure 1: Photo de Microscopie Electronique à Balayage de la souche <i>S. pneumoniae</i> TIGR4. (IBS)	17
Figure 2: Représentation schématique de la paroi des bactéries à Gram positif et à Gram négatif	20
Figure 3: Représentation schématique de l'unité de base du peptidoglycane	20
Figure 4: Les étapes de la biosynthèse de la capsule chez <i>S. pneumoniae</i>	23
Figure 5: Analyse de la distribution des sérotypes des souches invasives de pneumocoque	25
Figure 6: Evolution des PSDPs en France pour la période 1984-2008	27
Figure 7: Représentation schématique de la localisation des protéines de surface de <i>S. pneumoniae</i> .	34
Figure 8: Immunofluorescence de monocouches de cellules A549 et EaHy infectées par la souche <i>S. pneumoniae</i> D39	39
Figure 9: Peptide signal de AdcAII.	40
Figure 10: Représentation schématique du « Cell Wall Sorting Signal »	42
Figure 11: Représentation schématique du mécanisme de transpeptidation médié par la sortase A permettant l'ancrage des protéines à LPxTG à la paroi des bactéries à Gram positive	44
Figure 12: Structure RMN de la sortase AΔN59 de <i>S. aureus</i> complexée au peptide modifié LPAT*	46
Figure 13: Pili du <i>PI-1</i> de <i>S. pneumoniae</i> .	53
Figure 14: Organisation génétique des îlots de pathogénicité codant pour les pili des bactéries Gram-positif	55
Figure 15: Immunolocalisation des trois pilines qui composent le pilus SpaABC de <i>C. diphtheriae</i>	61
Figure 16 : Alignements locaux présentant les motifs conservés trouvés dans les séquences des pilines majeures utilisées pour cette étude	62
Figure 17: Le domaine N2 de la piline accessoire Gbs52 de <i>S. agalactiae</i> présente une topologie de CnaB.	64
Figure 18: Les ponts intramoléculaires de la piline majeure Spy128 de <i>S. pyogenes</i>	65
Figure 19: Identification de ponts intramoléculaires chez les pilines majeures SpaA et BcpA	67
Figure 20: Modèles de l'organisation générale des pili chez les bactéries Gram-positif.	69
Figure 21: Représentation schématique de la topologie des protéines Rrg de <i>S. pneumoniae</i>	70
Figure 22: Immuno-localisation de la piline majeure RrgB du PI-1 de <i>S. pneumoniae</i>	70
Figure 23: Immuno-localisation des pilines du PI-1 de <i>S. pneumoniae</i>	71
Figure 24: Immuno-localisation des pilines RrgA et RrgB du PI-1 de <i>S. pneumoniae</i>	72
Figure 25: Immuno-localisation des pilines RrgA, RrgB et RrgC sur des pili purifiés provenant de la souche <i>S. pneumoniae</i> TIGR4	72
Figure 26: Immunolocalisation des pilines mineures RrgA et RrgC à partir de pili natifs purifiés de la souche TIGR4	73
Figure 27: Modèles de l'organisation générale des pili du pneumocoque	74
Figure 28: Analyse du pilus SpaABC de <i>C. diphtheriae</i>	75
Figure 29: Modèle de l'assemblage du pilus SpaABC de <i>C. diphtheriae</i>	77
Figure 30: Structures cristallographiques des sortases C-1 et C-3 de <i>S. pneumoniae</i>	79
Figure 31: Organisation schématique du vecteur pLIM01	83
Figure 32: Organisation schématique des vecteurs pETDuet®, pACYCDuet® et pCDFDuet®.	84
Figure 33: Représentation schématique des protéines natives et des constructions des pilines dans les vecteurs pLIM01.	85
Figure 34: Carte du vecteur pETDuet RrgA RrgB	86
Figure 35: Carte du vecteur pACYCDuet RrgC SrtC-1	87
Figure 36: Carte du vecteur pCDFDuet SrtC-2 SrtC-3	87
Figure 37: Schéma du principe de la méthode TSA	99
Figure 38: Arrangement schématique de la structure tertiaire de RrgA	105
Figure 39: L'Adhésine du pilus de <i>S. pneumoniae</i> montre des caractéristiques structurales d'origine multiple	106
Figure 40: Repliement de type CNA et IgG	107
Figure 41: Les ponts intramoléculaires de RrgA.	108
Figure 42: Les Domaines D2 and D4 sont stabilisés par les ponts intramoléculaires	109
Figure 43: Analyse de la protéine recombinante RrgA	110
Figure 44: Production du domaine D4 recombinant	111
Figure 45: Alignement des séquences CnaB (PF05738) des pilines de <i>S. pneumoniae</i> (RrgA, RrgB et RrgC), <i>S. agalactiae</i> (Gbs52 et Gbs104), <i>B. cereus</i> (BcpB), <i>E. faecalis</i> (EpbB et EpbC) et <i>C. diphtheriae</i> (SpaG)	113
Figure 46: Les ponts intramoléculaires de RrgC	114
Figure 47: Stabilité protéolytique de RrgC	115
Figure 48: Schéma récapitulatif des données obtenues lors des expériences de digestions par la trypsine	117
Figure 49: Thermostabilité de RrgC et de ses variants	118

Figure 50: Les ponts intramoléculaires de RrgB	119
Figure 51: Identification de la « D-Box »	121
Figure 52: Stabilité protéolytique de RrgB	121
Figure 53: Thermostabilité de RrgB et de ses variants	122
Figure 54: Structure cristallographique de RrgB	123
Figure 55: Comparaison des fibres natives et artificielles de RrgB	125
Figure 56: Assemblage de RrgB dans la fibre artificielle	126
Figure 57: Constructions utilisées dans la plateforme de co-expression <i>E. coli</i>	129
Figure 58: Catalyse par SrtC-1 du complexe covalent RrgB-RrgC	130
Figure 59: Spécificité de SrtC-1 dans la reconnaissance du substrat	133
Figure 60: Rôle des ponts intramoléculaires dans la formation du complexe covalent RrgB-RrgC	134
Figure 61: Schéma récapitulatif de la catalyse du complexe RrgB-RrgC par SrtC-1	135
Figure 62: Catalyse par SrtC-2 des complexes covalents RrgA-RrgB et RrgB-RrgC	136
Figure 63: Spécificité de SrtC-2 dans la reconnaissance des substrats	137
Figure 64: SrtC-2 a un substrat privilégié : le motif YPRTG de RrgA	138
Figure 65: La Lysine du motif piline de RrgB est impliquée dans la liaison intermoléculaire de RrgA-RrgB	139
Figure 66: Rôle des ponts intramoléculaires dans la formation du complexe covalent RrgA-RrgB	141
Figure 67: Existence d'un intermédiaire acyl-enzyme SrtC-2- RrgA stable	142
Figure 68: La sortase C-2 catalyse un complexe D4-RrgB et permet la multimérisation de D4	143
Figure 69: SrtC-2 multimérise RrgA	145
Figure 70: Analyse par MET en coloration négative des formes monomériques et multimériques de RrgA	146
Figure 71: Récapitulatif de la catalyse du complexe RrgA-RrgB par SrtC-2	147
Figure 72: Récapitulatif de la multimérisation de RrgA par SrtC-2	148
Figure 73: Catalyse par SrtC-3 des complexes covalents RrgB-RrgA et RrgB-RrgC	149
Figure 74: Spécificité de SrtC-3 dans la reconnaissance des substrats	151
Figure 75: Existence d'un intermédiaire acyl-enzyme RrgC-SrtC-3 stable	152
Figure 76: Analyse ESI-MS de l'acyl-enzyme SrtC-3*RrgC	153
Figure 77: Schéma récapitulatif des complexes catalysés par SrtC-3	154
Figure 78: Schéma récapitulatif de la formation de l'acyl-enzyme RrgC-SrtC-3	155
Figure 79: Modèle d'assemblage du pilus PI-1 de <i>S. pneumoniae</i>	162
Tableau 1: Récapitulatif des interactions des protéines de surface du pneumocoque avec les protéines de l'hôte identifiées par tests ELISA	36
Tableau 2: Récapitulatif des différentes classes de sortases: rôles et caractéristiques	49
Tableau 3 : Les différents types de pili des bactéries Gram-négatif.	51
Tableau 4: Informations relatives aux protéines natives et aux protéines recombinantes provenant des constructions réalisées dans le vecteur pLIM01	86
Tableau 5: Liste des amorces utilisés pour les clonages destinés aux co-expressions.	88
Tableau 6: Informations relatives aux protéines recombinantes obtenue à partir des constructions réalisées dans le vecteur pETDuet (pour RrgA et RrgB) pACYCDuet (pour RrgC et SrtC-1) et pCDFDuet (pour SrtC-2 et SrtC-3).	88
Tableau 7 : Mesure des masses de RrgA et de ses variants par ESI-MS	108
Tableau 8: Analyse par ESI-MS de la protéine recombinante native et de ses mutants	115
Tableau 9: Masses moléculaires et séquences N-terminales des variants RrgC natif et mutants digérés à la trypsine.	116
Tableau 10: Analyse des spectres de masse ESI-TOF de la protéine recombinante native et de ses mutants	120
Tableau 11: Résumé des peptides identifiées lors de l'étude du complexe RrgB-RrgC par MS/MS	131
Tableau 12: Récapitulatif des masses des protéines recombinantes seules ou en complexe.	132

## TABLE DES ABREVIATIONS

°C: degree centigrade  
6xHis: étiquette 6 Histidines  
Å: Angström  
ABC: ATP-Binding Casette  
ADN: Acide DésoxyriboNucléique  
Amp: Ampicilline  
ARN: Acide RiboNucléique  
BHI: Brain Heart Infusion  
CBD: Choline-binding Domain  
CBP: Choline-Binding Protein  
Cm: Chloramphénicol  
CNRP: Centre National de Référence du Pneumocoque  
CRP: C-Reactive Protein  
CSP: Competence-Stimulating Peptide  
CWSS: Cell Wall sorting Signal  
DMSO: DiMéthylSulfOxyde  
dNTP: désoxyriboNucléotides TriPhosphate  
EDTA: EthyleneDiamineTetraAcetic acid  
ELISA: Enzyme-Linked ImmunoSorbent Assay  
ESI-MS: ElectroSpray Ionisation-Mass Spectrometry  
FACS: Fluorescence-Activated Cell Sorting  
Fc: Fragment constant  
Fig.: Figure  
FRET: Fluorescence Resonance Energy Transfer  
GAPDH: GlycérAldhéhyde 3-Phosphate DésHydrogénase  
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid  
HK: Histidine Kinase  
HMW: High Molecular Weight  
HRP: HorseRadish Peroxidase  
Ig: Immunoglobuline  
IIP: Infections Invasives à Pneumocoque  
IPTG: Isopropyl-β-D-thio-galactopyranoside  
kDa: kilo Dalton  
Km: Constante de michaelis  
LB: Luria Bertina  
LTA: LipoTeichoic Acid  
MBR: Metal-Binding Receptor  
MCS: Multiple Cloning Site  
MET: Microscopie Electronique à Transmission  
MIDAS: Metal Ion-Dependant Adhesion Site  
MOPS : 4-MOrpholinePropaneSulfonic acid  
MS/MS: Spectrométrie de masse en tandem  
MSCRAMMs: Microbial Surface Components Recognizing Matrix Molecules  
NAG: N-Acetyl Glucosamine  
NAM: N-Acetyl Muramique  
NanA: Neuraminidase A  
NBD: Nucleotide Binding Domain



OMS: Organisation Mondiale de la Santé  
ORP: Observatoire Régional du pneumocoque  
PBP: Penicillin Binding Protein  
PBS: Phosphate Buffered Saline  
Pce: Phosphocholinestérase  
PCR: Polymerase Chain Reaction  
PDB: Protein Data Bank  
PI-1: Pathogenicity Islet-1  
PIPES: Piperazine -1,2bis(2-ethanesulfonic) acid  
PM : Poids Moléculaire  
qsp: quantité suffisante pour  
RBS: Ribosome Binding Site  
RMN: Résonance Magnétique Nucléaire  
RMSD: Root Mean Square Deviation  
RR: Régulateur de Réponse  
Rrg : RlrA-regulated gene  
SBR: Substrate Binding Receptor  
SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis  
Srt: Sortase  
STEM: Scanning Transmission Electron Microscopy  
STM: Signature Tagged Mutagenesis  
Str: streptomycine  
TA: Teichoic Acid  
TB: Terrific Broth  
TCS: Two Component System  
Tev: Tobacco etch virus  
THY: Todd-Hewitt Yeast  
Tm: Température de fusion  
TMD: TransMembrane Domain  
TNF: Tumor Necrosis Factor  
TSA: Thermal Shift Assay  
VWA: Von Willebrand A

# **Introduction**



## INTRODUCTION

### I *Streptococcus pneumoniae*

#### I.1 Présentation générale

##### I.1.1 Classification

En 1881, deux microbiologistes, Louis Pasteur en France et George M. Sternberg aux Etats-Unis isolent indépendamment à partir de salive humaine des bactéries qui se présentent sous forme de diplocoques. Alors que Sternberg utilise sa propre salive, Pasteur se sert de celle d'un enfant mort de la rage pour infecter des lapins. Les deux chercheurs démontrent le rôle pathogène de ces bactéries et l'identifient comme agent responsable de pneumonies. En 1926, cette bactérie est nommée *Diplococcus pneumoniae* en raison de ses caractéristiques morphologiques, puis est renommée *Streptococcus pneumoniae* en 1974 (du grec Strepto pour « chaîne » et coccus pour « graine ») car lors de sa croissance en milieu liquide, la bactérie se présente sous forme de chaînettes (Fig.1).

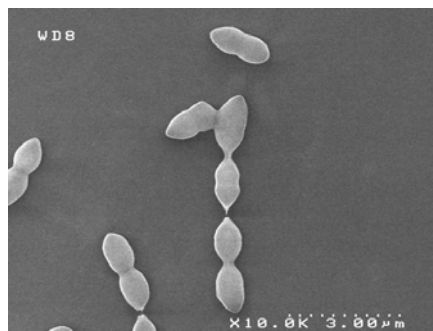


Figure 1: Photo de Microscopie Electronique à Balayage de la souche *S. pneumoniae* TIGR4. (IBS)

La première classification complète des streptocoques est proposée par Sherman en 1937. Plus récemment, l'analyse des séquences des gènes codant pour les ARNs 16S permet de classer les streptocoques en six groupes : Anginosus, Pyogenic (*S. pyogenes* et *S. agalactiae*), Bovis, Salivarius, Mutans et Mitis, ce dernier comprenant *S. pneumoniae*, *S. mitis* et *S. oralis* (Nobbs 2009). Ces espèces partagent près de 99% d'identité de séquence avec *S. pneumoniae* (pneumocoque) ce qui fait d'elles d'excellentes candidates au transfert horizontal de gènes. *S. pneumoniae* appartient à la division des *Firmicutes*, à la classe des *Coccus*, à l'ordre des *Lactobacillales*, à la famille des *Streptococcaceae* et au genre *Streptococcus*.

## I.1.2 Caractéristiques biologiques

Différentes propriétés métaboliques du pneumocoque sont exploitées dans les tests d'identification bactériologique. Le pneumocoque est une bactérie à métabolisme principalement anaérobie et, comme toutes les bactéries du genre *Streptococcus*, dépourvue de catalase, l'enzyme catalysant la dismutation de l'eau oxygénée. D'autre part, sur gélose au sang frais, les colonies de pneumocoque apparaissent entourées d'une plage légèrement verte qui résulte non pas de la lyse des hématies mais de la transformation de l'hémoglobine en biliverdine ; pour cette raison, le pneumocoque est qualifié de bactérie alpha-hémolytique. L'addition de bile ou de sels biliaires (déoxycholate de sodium) à une culture de pneumocoque active l'une de ses autolysines, LytA, une N-acetylmuramoyl-L-Ala amidase qui clive le peptidoglycane et mène à la lyse bactérienne : la rapidité de cette lyse induite est caractéristique du pneumocoque et donc parfois utilisée pour différencier le pneumocoque des autres streptocoques (Euzéby 2007). Enfin, *S. pneumoniae* est la seule espèce du genre *Streptococcus* à être sensible à un sel de cuivre, l'éthyl-hydrocupréine ou l'optochine, ce qui en fait un excellent test de typage bactérien.

Le pneumocoque présente une propriété de compétence naturelle, faisant intervenir une centaine de gènes sur les quelques 2000 gènes identifiés, et qui joue un rôle très important dans l'adaptation de la bactérie à son environnement. C'est d'ailleurs ce phénomène qui est à la base du mécanisme de transformation bactérienne mis en évidence par F. Griffith en 1928 qui a décrit les souches R et S et comparé leur pouvoir pathogène. En effet, lorsque le pneumocoque est cultivé sur une gélose supplémentée en sang, les colonies apparaissent sous deux formes : une forme transparente, lisse (S, Smooth) pour les bactéries encapsulées et une forme rugueuse (R, Rough) pour les bactéries dépourvues de capsule. Chez la souris, seules les colonies S sont létales. Griffith observe que l'injection de bactéries de type S préalablement tuées par la chaleur perdent leurs propriétés virulentes. En revanche, lorsque ces bactéries S préalablement chauffées, sont injectées conjointement avec des bactéries R vivantes, les souris meurent de septicémie. Les bactéries R, au contact de bactéries S inactivées, ont donc acquis un caractère pathogène qu'elles ne possédaient pas précédemment. Ce phénomène a été appelé transformation bactérienne et il a été par la suite reproduit chez plusieurs autres espèces bactériennes. Griffith en déduit également qu'il doit exister un « facteur transformant » probablement libéré par la chaleur et susceptible d'être intégré par d'autres bactéries, leur conférant ainsi de nouvelles propriétés génétiques. Ce n'est qu'en

1944 que la nature biochimique du « facteur transformant » est mis en évidence par Avery, McLeod et McCarthy : l'ADN.

Ce phénomène de compétence, très étudié chez le pneumocoque depuis cette découverte, peut être divisé en plusieurs phases: (i) l'induction de la compétence, (ii) la liaison de l'ADN double brin exogène à la surface cellulaire, (iii) la digestion de l'ADN double brin, (iv) l'import d'ADN linéaire simple brin et (v) la recombinaison de l'ADN homologue dans le chromosome (Nobbs 2009, Lacks 2004). L'induction du processus de « compétence » est initiée par des peptides CSP (Competence-Stimulating Peptide) qui servent de signal dans un mécanisme de détection du quorum par lequel les bactéries d'une même population peuvent communiquer et estimer leur nombre. En effet, lorsque la concentration en CSP extracellulaire atteint 1-10 ng/ml (ce qui correspond à une densité cellulaire de  $10^7$  cellules/ml) la compétence est initiée (Yother 2004, Tomasz 1964, Pozzi 1996, Steinmoen 2002). Ainsi, ce mécanisme naturel présente un avantage physiologique pour la bactérie puisqu'elle lui permet de s'adapter à son environnement par échange de gène et recombinaison homologue. C'est d'ailleurs ce phénomène qui est observé dans l'acquisition de résistances aux antibiotiques.

## **I.2 La paroi bactérienne**

En 1884, Christian Gram montre, grâce à une technique de coloration qui porte son nom, que les bactéries se divisent en deux grands groupes: les bactéries dites Gram-positif qui fixent le colorant et apparaissent de couleur violette et les bactéries dites Gram-négatif qui ne le fixe pas et gardent une coloration rose. Ce n'est que grâce au développement de la microscopie électronique dans les années 50 et la description des morphologies bactériennes que les différences observées lors des colorations Gram se trouvent expliquées.

Les bactéries Gram-positif, dont *S. pneumoniae* fait partie, possèdent une paroi bactérienne qui se compose d'une couche de peptidoglycane d'une épaisseur approximative de 20 à 80 nm dans laquelle sont enchâssées des acides teïchoïques (polymères contenant du ribitol, du glycérol phosphate et des sucres) et des acides lipoteïchoïques ancrés dans la membrane par leur partie lipidique. Cette paroi, selon la bactérie considérée, peut être elle-même surmontée d'une capsule polysaccharidique dont l'épaisseur varie entre 200 et 400 nm. Chez les bactéries Gram-négatif, la couche de peptidoglycane, plus fine (10-20 nm), est recouverte par une membrane externe dans laquelle sont présentes des protéines et des lipopolysaccharides, qualifiés d'endotoxines pour leurs propriétés inflammatoires (Fig. 2).

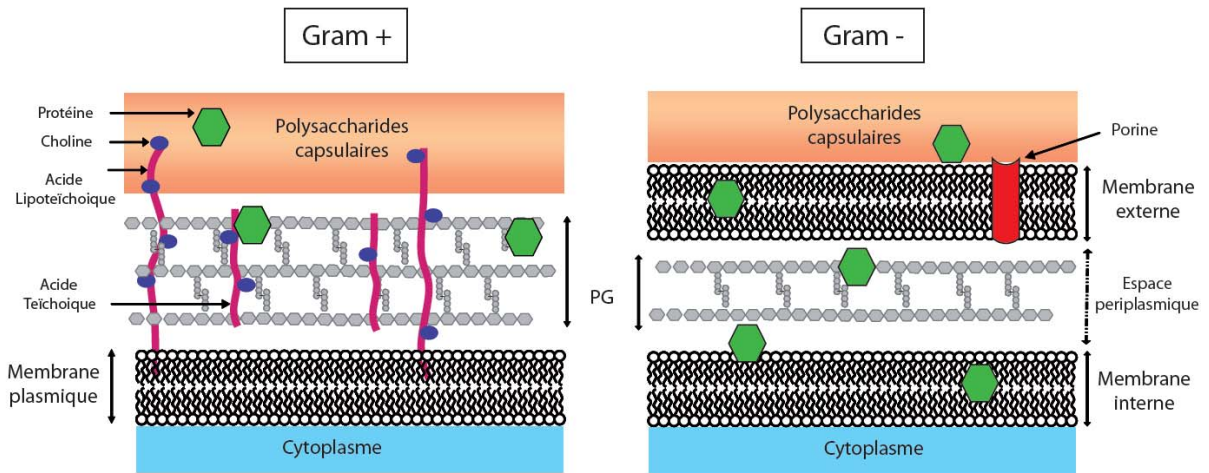


Figure 2: Représentation schématique de la paroi des bactéries à Gram positif et à Gram négatif

### I.2.1 La paroi de *S. pneumoniae*

Chez le pneumocoque, la paroi a une épaisseur de 20 nm et se compose de trois éléments : un maillage de peptidoglycane représentant une plateforme d'ancrage pour de nombreuses protéines de surface, des acides teïchoïques et des acides lipoteïchoïques.

#### I.2.1.1 Le peptidoglycane

Le peptidoglycane est formé par des chaînes glycanes, composées par l'alternance de résidus *N*-Acetyl-Glucosamine (NAG) et *N*-Acetyl-Muramique (NAM) et reliées entre elles par des liaisons osidiques  $\beta$ 1-4. Ces polymères saccharidiques sont réticulés, au niveau des groupements D-lactyl du résidu NAM, par des pentapeptides dont la séquence contient des acides-aminés de série L et D (*L*-Ala-*D*- $\gamma$ -Glu-*L*-Lys-*D*-Ala-*D*-Ala) (Fig.3).

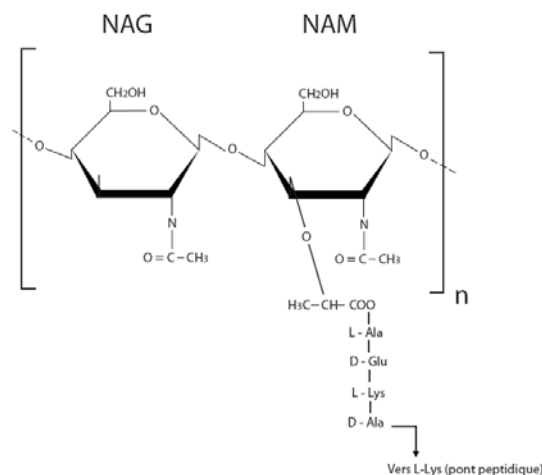


Figure 3: Représentation schématique de l'unité de base du peptidoglycane

Le lipide II est le précurseur du peptidoglycane et se compose d'une chaîne aliphatique undecaprenyl couplée, via un pyrophosphate, au NAM-pentapeptide, lui-même associé à un groupement NAG. Le lipide II est produit dans le cytoplasme puis est exporté du côté extracellulaire par un mécanisme encore inconnu. Le lipide II est alors pris en charge par des glycosyltransférases membranaires qui catalysent l'élongation de la chaîne glycane. Des transpeptidases interviennent ensuite pour former des liaisons peptidiques entre le résidu L-Lys d'une première chaîne et le premier résidu D-Ala d'une autre chaîne. Ces réactions conduisent à la réticulation tridimensionnelle du peptidoglycane, lui conférant ainsi des propriétés de résistance et d'élasticité qui permettent à la bactérie de résister aux fortes variations de pressions osmotiques internes et externes auxquelles elle est soumise. Afin de maintenir l'intégrité des cellules filles lors de la division bactérienne, la synthèse du peptidoglycane doit être coordonnée aux processus d'invagination de la membrane et de ségrégation de l'ADN (Den Blaauwen 2008).

Les protéines qui catalysent les réactions de glycosyltransfert et de transpeptidation permettant d'obtenir le maillage tridimensionnel du peptidoglycane sont appelées PBPs pour « Penicillin Binding Proteins ». Ces enzymes ont en effet été identifiées comme cibles de la pénicilline, antibiotique de la classe des  $\beta$ -lactamines, avant même que leur fonction physiologique dans la synthèse du peptidoglycane ne soit compris (Tipper et Strominger 1965, Blumberg 1972). L'intégrité du peptidoglycane est un élément essentiel à la survie de la bactérie : l'inhibition de sa synthèse est donc létale, expliquant le succès des antibiotiques de la famille des  $\beta$ -lactamines. Ces molécules se lient de façon covalente à la sérine active du site catalytique des domaines transpeptidases des PBP et inhibent la réaction de transpeptidation. Découvert plus récemment, la vancomycine, antibiotique de la classe des glycopeptides, inhibe également la réaction de transpeptidation mais en se liant à la Lysine du pentapeptide du lipide II, bloquant la reconnaissance de l'enzyme pour son substrat.

### **I.2.1.2 Les acides teïchoïques et lipoteïchoïques**

En 1930, les acides teïchoïques (TA) et lipoteïchoïques (LTA) étaient qualifiés respectivement de polysaccharide C et lipocarbohydrate (ou antigène F) et leur relation structurale déjà décrite (Tillet 1930, Goebel 1943). Contrairement à d'autres bactéries Gram-positif et Gram-négatif, les TAs et LTAs chez le pneumocoque sont formés de chaînes composées d'éléments de répétitions identiques dont le motif est : D-Glucose, 2-acetamido-4-amino-2,4,6-trideoxy-D-Galactose (2x), N-Acetyl-D-Galactosaminy (2x), ribitol-5-phosphate. Ces séquences sont liées les unes aux autres par un pont phosphodiester entre



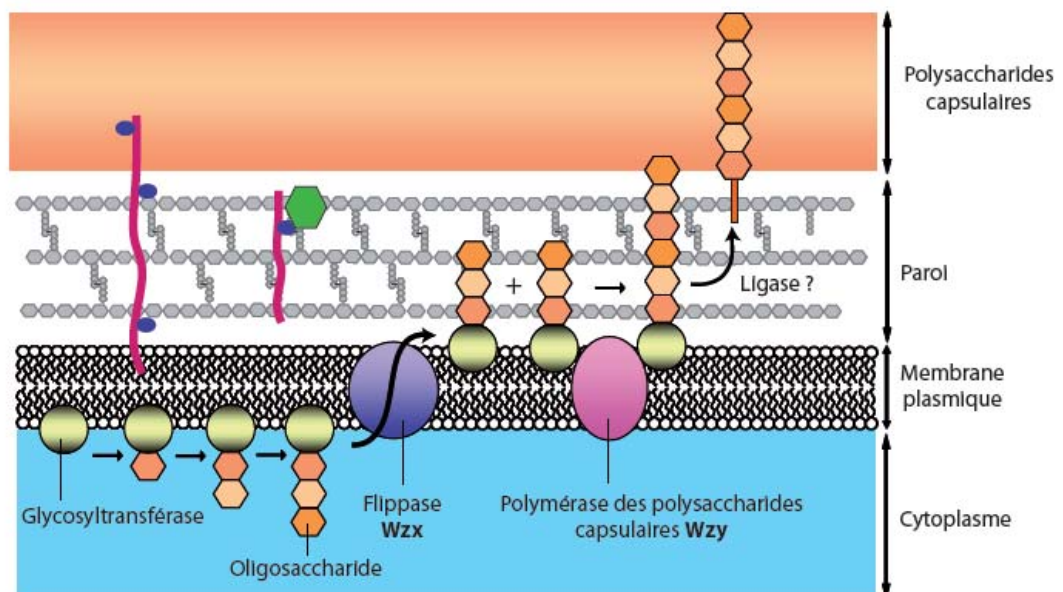
l'oxygène 5 du ribitol et l'oxygène 6 du résidu glucosyl de la répétition adjacente. Les TAs sont liés à 15 - 30% des NAMs du peptidoglycane via des liaisons phosphodiester alors que les LTAs sont liés à la membrane plasmique via les têtes polaires des lipides (Fisher 2000). Une particularité des TA et LTA du pneumocoque, mais également retrouvée chez *S. oralis* ou encore *H. influenzae*, est la présence de résidus phosphorylcholines liés par une liaison phosphodiester à l'oxygène 6 du résidu N-Acetyl-D-galactosaminyll (Tomasz 1967, Briles 1973). Par ailleurs, le nombre de résidus phosphocholines par répétition est souche dépendant. A titre d'exemple chez la souche R6, la majorité des répétitions porte deux résidus phosphocholines alors que chez la souche Rx1, seul un résidu phosphocholine est présent par répétition (Tomasz 2000).

### **I.2.2 La capsule de *S. pneumoniae***

Dès 1880, Pasteur note la présence d'une fine couche d'un composant de nature non identifiée localisée à la surface de la bactérie, qui sera plus tard défini comme la capsule polysaccharidique. Dochez et Avery parviennent à isoler ce matériel biologique qu'ils supposent protéique, comme tous les composés présentant un pouvoir antigénique identifiés jusqu'alors. Finalement, en 1925 Avery démontre que la capsule est en réalité constituée de polysaccharides dont l'épaisseur varie entre 200 et 400 nm selon les souches (Avery 1925, Sorensen 1988, Kamerling 2000). Bien que les mécanismes de liaisons de la capsule au peptidoglycane ne soient pas encore totalement élucidés, il semble que ces deux composants soient liés covalentement l'un à l'autre. La capsule est un polymère composé par la répétition d'unités mono- ou oligo-saccharidiques plus ou moins complexes (polymères branchés, chaînes latérales additionnelles). Dix-huit saccharides différents et groupement associés sont retrouvés dans les différents types de capsule. Le rôle de la capsule dans les processus physiopathologiques est très important et sera décrit au chapitre II.1.

A l'heure actuelle, 91 sérotypes de capsule ont été identifiés. Ceux-ci diffèrent selon leur composition en polysaccharides et les liaisons qui les unissent (Lynch 2010). Deux systèmes de nomenclature des sérotypes ont été développés : le système américain basé sur l'ordre chronologique d'isolement des souches cliniques et le système danois, définitivement adopté, qui tient compte de la composition en polysaccharides et donc des caractéristiques antigéniques et structurales capsulaires. Les 91 sérotypes sont ainsi classés en 46 sérogroupes (1 à 46 avec seuls les sérogroupes 26 et 30 non représentés) associés à une lettre A, B ou C qui est fonction de la spécificité antigénique. Enfin dans un même séro groupe les lettres F (First désignant le premier sérotype d'un serogroupe), L (Lederle), N (Neufeld) et V

(Valdemar) peuvent désigner le sérotype considéré (Brisou 2004). Malgré la diversité capsulaire, tous les gènes impliqués dans la synthèse sont organisés de façon similaire dans un locus génétique *cps* localisé entre les deux gènes *dexB* et *AliA*. Les quatre premiers gènes du locus *cps* (*cpsA-D*) sont communs à tous les sérotypes, à l'exception des sérotypes 3 et 37. Les protéines codées par ces gènes ont toutes été définies comme des protéines de régulation de la synthèse capsulaire (Morona 2000). La région centrale du locus *cps* comprend des gènes qui codent pour des glycosyltransférases qui assemblent les unités oligosaccharidiques sur le lipide accepteur. Cette région centrale contient également le gène codant pour une flippase (*Wzx*) qui transfère les unités oligosaccharidiques vers la face externe de la membrane plasmique ainsi que le gène codant pour la polymérase (*Wzy*) qui lie les unités polysaccharidiques les unes aux autres (Fig.4). Enfin, les gènes codant pour les enzymes indispensables à la synthèse et l'activation des précurseurs oligosaccharidiques sont localisés à la fin du locus *cps*.



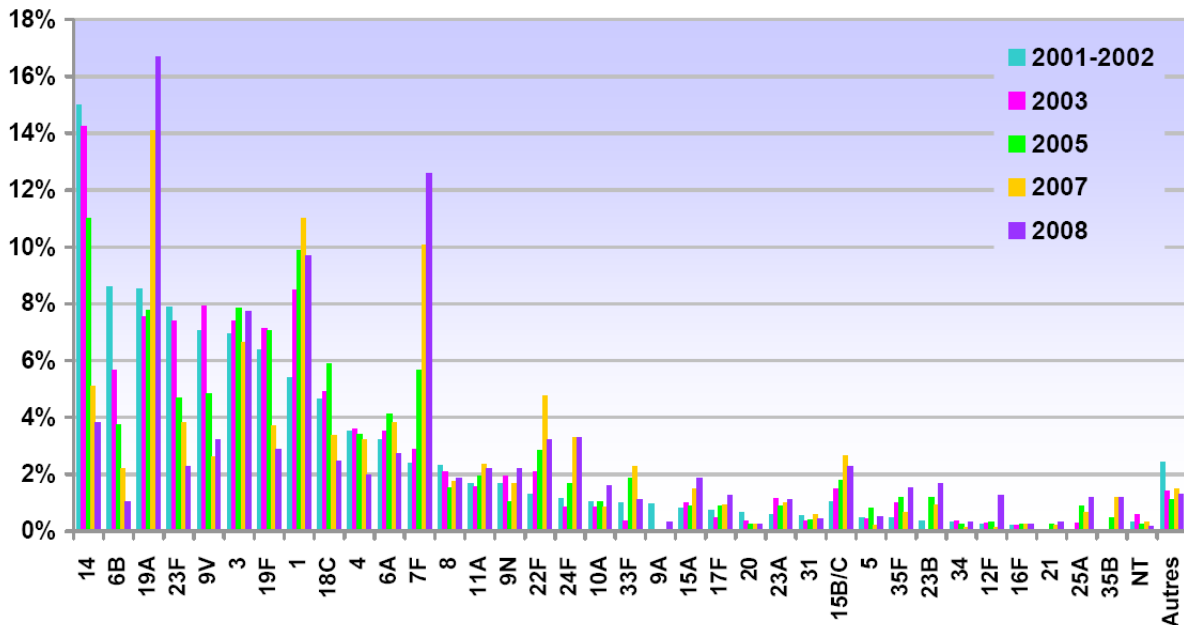
**Figure 4: Les étapes de la biosynthèse de la capsule chez *S. pneumoniae*.** La biosynthèse de la capsule est régulée par le locus *cps*. (i) Les glycosyltransférases assemblent les unités oligosaccharidiques au niveau de la face interne de la membrane plasmique, dans le cytoplasme. (ii) La Flippase *Wzx* transporte les répétitions d'oligosaccharides au niveau de la face externe de la membrane plasmique. (iii) *Wzy* polymérise les répétitions d'oligosaccharides pour former les polysaccharides capsulaires qui sont liés covalentement au peptidoglycane par un mécanisme encore inconnu.

### **I.3 Données épidémiologiques**

Le pneumocoque est une bactérie commensale des voies aériennes supérieures et notamment de la sphère oropharyngée qu'il colonise de façon asymptomatique dès le troisième mois de vie du nourrisson. Néanmoins, lorsque les défenses immunitaires de l'hôte sont affaiblies, chez les jeunes enfants, les personnes âgées ou encore les individus immunodéprimés, la migration de la bactérie vers des régions stériles peut mener à l'établissement d'infections non invasives comme les sinusites, les otites, les bronchites ou encore les pneumonies. La migration bactérienne peut conduire au développement de maladies dites invasives telles que les méningites et les septicémies lorsque des barrières tissulaires sont traversées par la bactérie. Il est à noter que l'épidémiologie des infections invasives à pneumocoque (IIP) varie selon les régions géographiques. Les jeunes enfants sont principalement affectés dans les pays en développement, reflétant la pauvreté, la malnutrition, la co-morbidité et bien entendu l'accès restreint aux soins médicaux (Lynch 2010). Dans les pays développés, les personnes âgées et immunodéprimées sont les populations essentiellement touchées par les IIP. Un certain nombre de facteurs de risques associés aux IIP ont été déterminés : les abus d'alcool, les infections grippales récentes, l'insuffisance cardiaque, le diabète, l'asthme ou encore l'exposition à la fumée de cigarettes sont des facteurs augmentant l'incidence des IIPs (Lynch 2009). Ainsi, selon l'OMS, le pneumocoque serait responsable de près de 1,6 million de mort par an dans le monde, dont 1 million ne concernerait que des enfants de moins de 5 ans. En France, en 2008, 6679 cas d'IIPs ont été répertoriés dont 611 méningites et 6068 bactériémies. Entre 2007 et 2008, tout âge confondu, l'incidence des IIPs a augmenté de 14% (Dorléans 2010). En revanche, si on s'intéresse à la tranche d'âge de 0 à 23 mois, l'incidence des IIPs est en forte baisse (32,5%), conséquence probable de la grande couverture vaccinale du Prevnar® (ORP 2010).

Sur les 91 sérotypes identifiés à ce jour, seuls 20 sont responsables de 80% des IIPs. Il s'agit principalement des sérotypes 14, 4, 1, 6A, 6B, 3, 8, 7F, 23F, 18C, 19F et 9V. Chez les jeunes enfants, ce nombre est moins important, les sérotypes prédominants sont : 6, 14, 18, 19 et 23F. La représentativité des sérotypes infectieux varie selon les régions géographiques mais également au cours du temps. Des phénomènes de remplacement sont mis en évidence avec l'émergence de sérotypes peu représentés ou non-vaccinaux, par rapport aux sérotypes majeurs et vaccinaux. Le rapport émis par le Centre National de Référence des Pneumocoques permet de suivre l'évolution, entre 2001 et 2008, de la distribution des sérotypes des souches isolées d'hémocultures et de liquide céphalo-rachidien (Fig.5) : tout âge confondu, ce sont les

sérotypes 19A, 7F, 1 et 3, non inclus dans le Prevenar®, qui sont dorénavant prédominants (Fig.5).



**Figure 5: Analyse de la distribution des sérotypes des souches invasives de pneumocoque (Hémocultures, LCR) tout âge confondu entre 2001 et 2008.** En 2001-2002 (n=2719), en 2003 (n=1390), 2005 (n=1236), 2007 (n=1489) et en 2008 (n=1176) (Source : Rapport cnpr 2009).

### I.3.1 Prophylaxie

La prophylaxie correspond à l'ensemble des mesures (vaccins, tests de dépistage, hygiène) destinées à prévenir l'apparition de maladies. A l'heure actuelle dans les pays développés, la vaccination est la principale mesure mise en place pour prévenir les infections à pneumocoque. Le premier vaccin anti-pneumocoque mis sur le marché en 1983 est le Pneumovax®. Ce vaccin commercialisé par Merck est composé d'un mélange de 23 antigènes capsulaires correspondant aux sérotypes les plus représentés et/ou les plus infectieux : 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F et 33F. Il est destiné aux personnes âgées de plus de 65 ans ainsi qu'aux individus à risques âgés entre 2 et 64 ans. Ce vaccin a prouvé son efficacité chez les jeunes adultes et les personnes âgées chez lesquels il confère une protection de 60-80% contre les IIPs. En revanche, chez les individus immunodéprimés cette efficacité est moindre. Ce vaccin 23-valent est peu immunogénique chez l'enfant de moins de deux ans. En effet, les antigènes capsulaires polysaccharidiques qui le composent activent uniquement les lymphocytes B, dont l'efficacité est limitée chez les jeunes enfants dont le système immunitaire est immature. C'est la raison pour laquelle en 2000 le Prevenar® (Wyeth), destiné aux enfants de moins de deux ans, est

mis sur le marché aux Etats-Unis. Les premières recommandations pour la vaccination par le Prevenar® en France débutent en 2003 mais restent peu suivies, la seconde en 2006 est plus efficace et s'accompagne d'une augmentation du nombre de vaccination : la couverture vaccinale passe de 44% des enfants de moins d'un an vaccinés en 2004 à 80% en 2007 (ORP 2010). Le Prevenar® est un vaccin conjugué heptavalent comportant les 7 sérotypes les plus infectieux chez l'enfant: 4, 6B, 9V, 14, 18C, 19F et le 23F. Initialement destiné aux seuls jeunes enfants susceptibles de développer des infections invasives, son utilisation s'est étendue et s'adresse dorénavant à tous les enfants de moins de 2 ans. Depuis mai 2010, un nouveau vaccin anti-pneumococcique polysidique conjugué 13-valent (Prevenar 13®) est commercialisé en France par Pfizer. Celui-ci comporte les mêmes sérotypes que la version précédente mais complétée par les six sérotypes émergents: 1, 3, 5, 6A, 7F et 19A.

### **I.3.2 Moyens curatifs**

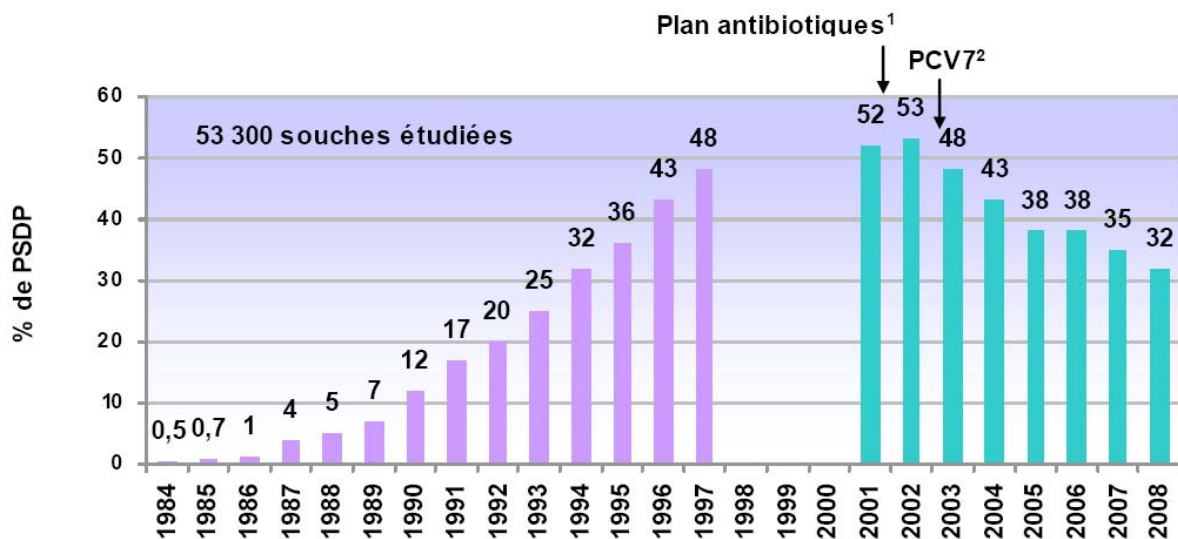
En 1928, Alexander Fleming découvre par hasard qu'un champignon, *Penicillium notatum*, a le pouvoir d'inhiber la prolifération de bactéries. Avec l'aide de chimistes, il isole la substance antibactérienne qu'il nomme pénicilline et teste l'efficacité de cette molécule dans le traitement d'infections bactériennes. Ce n'est que plusieurs années plus tard qu'un groupe de chimistes de l'université d'Oxford produit la pénicilline en grande quantité, qui est alors administrée à grande échelle et avec succès chez les soldats américains pendant la seconde guerre mondiale (Sulek 1968, Chain 1993, Ligon 2004). Dès lors, et jusqu'aux années 80, on parle de l'« âge d'or » des antibiotiques.

Le pneumocoque est sensible à la plupart des antibiotiques actifs sur les bactéries Gram-positif, à savoir les  $\beta$ -lactamines (pénicilline, céphalosporine, amoxicilline, etc...), les macrolides, les tétracyclines, le chloramphénicol, la rifampicine, le cotrimoxazole et les glycopeptides (Brisou 2004). Chacun de ces antibiotiques cible un processus essentiel au développement ou à la survie de la bactérie. Les  $\beta$ -lactamines et les glycopeptides par exemple inhibent la synthèse de la paroi bactérienne tandis que la tétracycline, le chloramphénicol et les macrolides ont un effet bactériostatique en inhibant la synthèse protéique. Le choix du traitement antibiotique dépend essentiellement de la pathologie considérée et des recommandations des autorités nationales de santé. Dans les cas de pneumopathies et d'otites moyennes aiguës à pneumocoque, c'est la prescription d'amoxicilline en monothérapie qui est privilégiée ; dans le cas de méningites, l'association de deux antibiotiques ciblant la paroi est préconisée : le céfotaxime associé à la vancomycine ou à l'amoxicilline (Brisou 2004).

### I.3.3 Difficultés rencontrées dans le traitement et la prévention des infections à pneumocoque

#### I.3.3.1 Résistance aux antibiotiques

En 1970, les premières souches de pneumocoque insensibles à la pénicilline sont identifiées en Afrique du Sud, en Australie et en Espagne. La dissémination des souches résistantes s'est étendue jusqu'à atteindre dans les années 2000 un taux supérieur à 50% des souches cliniques isolées, comme tel est le cas en France (Fig.6) (Decousser 2004, Lynch 2009). Ces phénomènes de diminution de sensibilité à la pénicilline s'expliquent par un remodelage des protéines interagissant avec l'antibiotique, qui se produit par recombinaison génétique (Laible 1991, Laible 1987).



**Figure 6: Evolution des PSDPs en France pour la période 1984-2008.** Données extraites du rapport 2009 du CNRP (Centre National de Référence du Pneumocoque) . En violet: données recueillies par P. Geslin pour la période 2001-2007. En turquoise: données recueillies par E. Varon et L. Gutmann. ¹Plan national pour préserver l'efficacité des antibiotiques, nov 2001 [http://www.sante.gouv.fr/htm/actu/34\\_01.htm](http://www.sante.gouv.fr/htm/actu/34_01.htm); ²Introduction du vaccin anti-pneumococcique conjugué heptavalent Prevenar® (PCV7). (PSDP : Pneumocoque de Sensibilité Diminuée à la Pénicilline)

Dans le cas de la résistance aux  $\beta$ -lactamines, les bactéries expriment des PBPs mutées, qui conservent leur fonction de transpeptidation du peptidoglycane, mais dont l'affinité pour les  $\beta$ -lactamines est diminuée par rapport à la souche de type sauvage. Ces mutations sont acquises par transfert horizontal et recombinaison d'ADN issu de souches génétiquement proches du pneumocoque telles que *S. mitis*. La compétence naturelle du pneumocoque est capitale dans ce processus et la forte consommation d'antibiotique favorise l'émergence de souches de sensibilité diminuée aux antibiotiques. Le phénomène de résistance est observé pour chaque famille d'antibiotique et il a été récemment montré que la résistance aux

macrolides augmente conjointement avec celle des pénicillines (Lynch 2009). Des profils de multi-résistances ont également été décrits. Il s'agit de la perte de sensibilité d'une même souche vis-à-vis d'au moins trois familles différentes d'antibiotiques. Une étude menée dans 15 pays d'Europe entre 2004 et 2005 a montré que 15,8% des isolats de pneumocoques étaient multi-résistants (Brandileone 2006). Quatre ans plus tôt, en Asie ce taux était de 26,8% (Hsieh 2009). La flambée des résistances aux antibiotiques serait due à la dissémination mondiale d'un faible nombre de clones. A l'heure actuelle, seuls six sérotypes sont responsables de plus de 80% des résistances aux  $\beta$ -lactamines ou aux macrolides: 6A, 6B, 9V, 14, 19F, 23F (Lynch 2009).

En France, les autorités nationales prennent conscience de l'urgence de la situation et mettent en place en 2001 le « plan antibiotiques ». Celui-ci a pour but notamment d'améliorer la surveillance des isolats cliniques résistants et multi-résistants, en renforçant l'action des centres régionaux d'observation de plusieurs bactéries pathogènes, dont le pneumocoque. La mise en place d'actions générales d'information, non seulement des professionnels par le biais de l'Afssaps, mais également de la population avec des spots télévisés diffusant le célèbre slogan « Les antibiotiques, c'est pas automatique » marquent les esprits. Alors qu'en 2001, 51% des pneumocoques montrent une sensibilité diminuée à la pénicilline, en 2008 ce pourcentage chute à 32% (Fig.6). Malgré ces efforts, le taux de résistance global envers tout type d'antibiotique reste élevé (Varon 2009).

### **I.3.3.2 Les limites de la protection vaccinale**

Depuis l'introduction du vaccin heptavalent, l'incidence mondiale des IIPs et des pneumonies causés par les sérotypes vaccinaux a diminué chez les enfants et les adultes, ce qui est probablement lié à une diminution du portage nasopharyngé chez les enfants vaccinés, réduisant ainsi la transmission aux enfants non vaccinés et aux adultes (Lynch 2009, Tyrrell 2006, Greenberg 2009). Cependant, une augmentation notable de sérotypes de remplacement, non vaccinaux, est apparu (Carvalho 2009, Moore 2008). En France, chez les enfants vaccinés de moins de 2 ans, l'augmentation de l'incidence des IIPs dues aux sérotypes non vaccinaux a été moins importante que la diminution de l'incidence des IIPs dues aux sérotypes vaccinaux, conduisant ainsi à un bilan positif, bien que cet effet favorable ait tendance à s'affaiblir en 2007-2008. En revanche, dans les autres tranches d'âges, l'effet indirect de la vaccination n'a pas suffi à compenser l'effet de remplacement sérotypique par les sérotypes non vaccinaux (Dorléans 2010).

Bien que le nouveau vaccin anti-pneumococcique 13-valent couvre les sérotypes 1, 3, 5, 6A, 7F et 19A, il apparaît évident que l'utilisation de ce type de vaccination engendrera à terme l'émergence de nouveaux sérotypes non-vaccinaux. Seule la production d'un vaccin 91-valent pourrait empêcher définitivement l'émergence des phénomènes de remplacement. Or, les coûts associés à ce type de produits rendent sa production non envisageable. De plus, les conséquences de l'éradication de cette bactérie commensale chez l'homme pourrait à terme favoriser la colonisation de l'organisme hôte par d'autres pathogènes. Par conséquent, il demeure nécessaire de développer de nouvelles méthodes de prévention des infections à pneumocoque. Les vaccins à base de protéines sont une piste intéressante car elles sont communes à un grand nombre de sérotypes et leurs séquences sont conservées, deux caractéristiques qui devraient limiter l'émergence des phénomènes de remplacement. Les principaux candidats pour ces vaccins sont des protéines exposées à la surface de la cellule bactérienne et impliquées dans la virulence du pneumocoque (Cao 2007, Seib 2009).





## II. La pathogénicité du pneumocoque

*S. pneumoniae* est une bactérie commensale de l'homme mais qui dans certains cas, est à l'origine de pathologies de gravité variable. Les processus physiopathologiques du pneumocoque, qu'il s'agisse des mécanismes moléculaires et cellulaires mis en place au cours de l'étape de colonisation et ceux développés lors des étapes invasives sont encore peu connus.

En 1999, Casadevall et Pirofsky décrivent la pathogénicité comme la capacité de la bactérie à causer des dommages chez son hôte (Casadevall et Pirofski 1999). La pathogénicité est associée à des facteurs dits de virulence qui sont définis comme les composants du microorganisme qui endommagent l'hôte mais inclut aussi les composants essentiels à la viabilité du pathogène au sein de l'hôte infecté. En 2007, Medzhitov propose que les acteurs bactériens de virulence correspondent à des facteurs d'adaptation du microorganisme à des environnements spécifiques (Medzhitov 2007). Ainsi, selon la niche colonisée, les facteurs de virulence peuvent présenter plusieurs type d'activité : (i) la pénétration des épithélia, (ii) l'attachement à la surface des cellules ou à la matrice extracellulaire, (iii) l'invasion de compartiments intracellulaires, (iv) l'évasion des mécanisme de défense de l'hôte (Medzhitov 2007).

La composition des facteurs de virulence des bactéries, et notamment ceux du pneumocoque, est variée. Il peut s'agir de protéines, de polysaccharides, de lipides ou de petites molécules qui vont être essentielles dans les interactions avec l'hôte. Chez *S. pneumoniae*, la plupart de ces facteurs sont sécrétés ou localisés en surface de la bactérie afin de faciliter leur contact avec les composants de l'hôte.

Le chapitre suivant n'a pas pour but de faire une présentation exhaustive de tous les facteurs de virulence du pneumocoque identifiés à ce jour. Seuls les fonctions de la capsule et de la pneumolysine y seront développés, reflétant le rôle essentiel de ces composants dans la virulence du pneumocoque. Quant aux nombreuses protéines de surface impliquées dans les interactions de *S. pneumoniae* avec l'hôte, chaque famille de protéines de surface sera illustrée au travers d'exemples, qui sont, ou qui ont été des sujets de recherche traités au laboratoire.

## II.1 La capsule

Comme décrit précédemment, 91 types de capsule structurellement et sérologiquement distincts ont été identifiés à ce jour (Lynch 2010). La plupart des sérotypes présentent des polysaccharides capsulaires chargés négativement au pH physiologique, qui forment une couche plus ou moins épaisse isolant la bactérie du milieu environnant (Lee 1991).

Etant la couche la plus externe de la bactérie, elle représente la première barrière à entrer en contact avec les constituants de l'hôte et joue donc un rôle important dans l'évasion du pneumocoque face aux mécanismes de défense de l'organisme. L'activation des voies classique et alternative du complément est particulièrement mobilisée par le système immunitaire inné pour lutter contre la prolifération du pneumocoque. Une fonction essentielle de la capsule est d'empêcher l'opsonisation de la bactérie par le fragment C3b, issu du clivage de C3 produit lors de l'activation du complément, inhibant ainsi la phagocytose bactérienne par les neutrophiles (Hyams 2010, Hyams 2010, Kim 1999, Magee 2001).

La modulation de l'épaisseur de la capsule semble jouer un rôle dans l'interaction de la bactérie avec les tissus épithéliaux, ce qui est particulièrement important lors de la colonisation des tissus respiratoires (MacLeod 1953, Yother 2004, Bootsma 2007, Nelson 2007). Des études montrent que des souches non-encapsulées présentent une meilleure adhérence aux cellules épithéliales A549 *in vitro* que des dérivés isogéniques exprimant les capsules des sérotypes 3 ou 19F qui elles-mêmes réduisent la quantité de polysaccharides capsulaires lorsqu'elles entrent en contact avec les cellules (Hammerschmidt 2005). L'adhésion accrue des bactéries aux cellules humaines, observée lorsque la capsule est moins épaisse ou absente, est probablement à mettre en relation avec une meilleure accessibilité des adhésines protéiques sous-jacentes.

Ces données suggèrent par ailleurs que la synthèse de la capsule doit pouvoir être régulée et ceci en réponse à une modification de l'environnement dans lequel se trouve la bactérie. Ces questions sont complexes et à l'heure actuelle peu d'éléments de réponse sont connus. La biosynthèse de la capsule semble néanmoins régulée par trois protéines: CpsB, CpsC et CpsD. CpsB est une phosphotyrosine phosphatase (dépendante du manganèse). CpsC est une protéine membranaire qui interagit avec la polymérase des polysaccharides capsulaires (Wzy). Enfin, CpsD est une tyrosine kinase dont la fonction d'autophosphorylation est activée via l'interaction avec CpsC. Morona et ses collaborateurs suggèrent que la phosphorylation de CpsD régulerait négativement la production de capsule (Morona 2000, Morona 2002, Morona

2003). Au contraire, selon les travaux de Bender et de ses collègues la phosphorylation de CpsD mènerait à une augmentation de la biosynthèse de la capsule (Bender 2003). Bien que ces résultats soient contradictoires il n'en demeure pas moins que la phosphorylation de CpsD est l'élément central de la régulation de l'expression de la capsule.

En conclusion, la capsule polysaccharidique présente de nombreuses fonctionnalités, (i) de protection et d'échappement face au système immunitaire, (ii) de diversité pour la bactérie, qui nécessite d'être hautement régulée en fonction de l'environnement dans lequel se trouve la bactérie.

## **II.2 La pneumolysine**

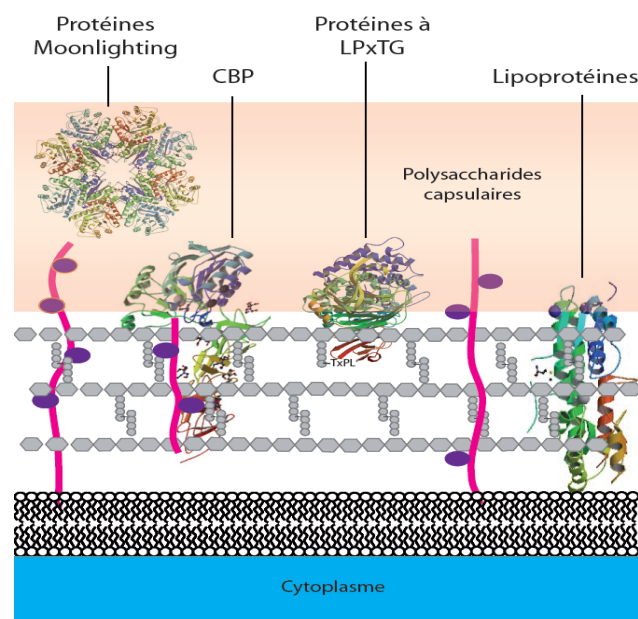
La pneumolysine est une protéine de 52 kDa retrouvée chez tous les isolats cliniques de pneumocoque. Cette protéine appartient à la famille des Cytolysines Dépendantes du Cholestérol car elle est capable de lyser toutes les cellules mammifères qui contiennent du cholestérol dans leur membrane. L'activité lytique de cette protéine se traduit par la formation de larges pores (260 angströms de diamètre) dans la membrane de la cellule cible, générés par l'oligomérisation de plus de 44 monomères de pneumolysine (Morgan 1995, Tilley 2005). Durant sa conversion du stade monomère soluble en oligomère membranaire, la pneumolysine subit des changements de conformation importants (Tilley 2005). Selon une étude menée sur des tissus neuronaux, la formation des pores par la pneumolysine induit une augmentation intracellulaire de calcium dans la cellule cible et le relargage du « Apoptosis-Inducing Factor » conduisant à la mort cellulaire (Braun 2002, Stringaris 2002). La pneumolysine peut également activer la voie classique du complément via sa liaison au fragment Fc des immunoglobulines engendrant ainsi une réponse pro-inflammatoire (Mitchell 1991, Paton 1984). Mais la pneumolysine est aussi capable de diminuer les réponses immunitaires et inflammatoires de l'hôte par ses effets cytotoxiques directs sur les cellules phagocytaires (Jedrzejewski, 2007)

Des études ont montré que des souches de pneumocoque mutantes n'exprimant pas la pneumolysine sont moins virulentes que la souche de type sauvage, qu'elles soient inoculées chez la souris par voie nasale ou systémique (Berry 1989). Dans des modèles d'infection pulmonaire, les souches mutantes induisent une inflammation moindre que celle générée par la souche parentale (Canvin 1995, Kadioglu 2000). De façon intéressante, une souche où la pneumolysine est mutée pour ses activités lytiques et d'activation du complément, apparaît plus virulente qu'une souche dans laquelle l'expression de la pneumolysine est totalement

abolie, suggérant que la pneumolysine présenterait une activité additionnelle, non identifiée à ce jour (Kadioglu 2008).

### II.3 Les protéines de surface et leur rôle dans la virulence

Depuis le début des années 2000, soit dès la publication de la séquence du génome des souches R6 et TIGR4, de nombreux cribles génétiques ont été réalisés afin d'identifier les protéines du pneumocoque potentiellement impliquées dans les interactions avec l'hôte (Zysk 2000, Hava 2002, Polissi 1998, Wizemann 2001, Rigden 2003, Tettelin 2001, Hoskins 2001). Un grand nombre de candidats ainsi caractérisés correspondent à des protéines localisées en surface de la bactérie, qu'elles soient sécrétées ou liées à la paroi, et pour un certain nombre d'entre elles, un rôle dans la virulence a été établi. Les protéines de surface peuvent être classées en 4 grands groupes. Le premier, atypique, correspond à des protéines qui ne présentent pas de peptide signal, ni aucun motif spécifique d'ancrage et qui sont en outre connues pour une fonction enzymatique dans le métabolisme bactérien, notamment la glycolyse. La double fonction remplie par ces protéines selon leur localisation (virulence en surface et métabolisme dans le cytoplasme) explique leur qualification de « moonlighting ». Les trois autres groupes réunissent des protéines qui présentent un peptide signal permettant leur exportation à travers la membrane plasmique et un motif d'ancrage spécifique dont la séquence est très conservée: les Choline-binding proteins (CBPs), les lipoprotéines et les protéines LPxTG (Fig.7)



**Figure 7: Représentation schématique de la localisation des protéines de surface de *S. pneumoniae*.** Pour les protéines Moonlighting : l'énolase (PDB 1W6T; Ehinger *et al.* 2004). Pour les Choline-Binding Protein: CBPE (PDB 2BIB; Hermoso *et al.* 2005). Pour les protéines à LPxTG : la hyaluronate lyase (PDB 1EGU; Li *et al.* 2000). Pour les lipoprotéines : AdcAII (PDB 3CX3; Loisel *et al.* 2008)

Huit génomes de souches de pneumocoques ont été totalement séquencés et sont actuellement disponibles : R6, D39, TIGR4, Hungary 19A-6, JJA, 70585, P1031 et Taïwan 19F-14. De part la conservation des motifs de liaison à la paroi des CBPs, lipoprotéines et protéines LPxTG, et grâce à l'utilisation d'outils bioinformatiques pour définir les peptides signaux, il est possible d'identifier l'ensemble des protéines présentes en surface de chacune de ces souches de pneumocoque. Par exemple, le séquençage de la souche de référence R6 en 2001 indiquait que 471 protéines sont potentiellement localisées en surface de la bactérie (tout mode de fixation à la paroi confondu) car présentant dans leur séquence un signal peptide (Hoskins 2001). Il est important de noter que ce nombre ne tient pas compte des protéines membranaires ni de celles exportées en surface par des mécanismes inconnus (ex : moonlighting). En conclusion, le profil d'expression de protéines en surface de cette bactérie Gram-positif est extrêmement riche, ce qui rend complexe l'étude des interactions entre le pneumocoque et l'hôte. Même si le rôle de chacune de ces protéines peut être étudié de façon indépendante, de nombreuses questions resteront encore en suspens : dans quelles conditions physico-chimiques, dans quels tissus humains, à quelle étape du processus infectieux ces protéines sont-elles exprimées ? Dans quelle mesure une intervention coordonnée de ces protéines est-elle nécessaire ?

Dans le but d'appréhender de façon un peu plus globale les interactions entre le pneumocoque et son hôte, un travail de criblage à moyen débit a été effectué au laboratoire. L'ensemble du répertoire des CBPs et des protéines LPxTG des souches R6 et TIGR4 a été produit, purifié et exploité dans un test d'interaction avec des composants de l'hôte, permettant ainsi d'identifier de nouvelles interactions (Frolet 2010) (Tableau 1). La pertinence et la fonction physiologique de ces données sont en cours de réalisation.

**Tableau 1: Récapitulatif des interactions des protéines de surface du pneumocoque avec les protéines de l'hôte identifiées par tests ELISA.** Les protéines de surface utilisées sont des protéines recombinantes. En saumon les CPBs, en turquoise les protéines à LPxTG (Adapté de Frolet *et al.* 2010)

	Collagènes	Collagène IV	Fibrinogène	Laminine	Elastine	CRP	Facteur H	Plasminogène
CbpA							+	
CbpI					+	+		
CbpM						+		
CbpJ						+		
CbpL	+				+	+		
CbpF					+			+
CbpE			+	+			+	+
SpuA								+
Eng								+
PrtA		+						+
ZmpB		+						
NanA	+	+	+					
Spr1806								+

### II.3.1 La GAPDH : un exemple de protéine « moonlighting »

La glycéraldéhyde 3-phosphate déshydrogénase (GAPDH) est une enzyme cytoplasmique qui joue un rôle très important dans la glycolyse, en catalysant la conversion du glycéraldéhyde 3-phosphate en D-Glycérate 1,3 bisphosphate. Or, de nombreux travaux rapportent la présence de cette enzyme, ainsi que plusieurs autres protéines de la glycolyse comme l'énolase, à la surface externe de différents types cellulaires, bactéries, parasites, levures, cellules mammifères. Dans tous les cas, le mécanisme d'export de ces enzymes glycolytiques n'est pas élucidé. Dès 1971, Tanner et Gray identifient la présence de la GAPDH en surface de cellules mammifères et en 1995, Glaser et Gross montrent que cette enzyme est impliquée dans les phénomènes de fusion membranaire (Tanner 1971, Glaser 1995).

Ces dernières années, l'exposition en surface de la GAPDH a été mise en évidence chez de nombreux microorganismes pathogènes et son implication dans la virulence a été décrite. Chez *S. pyogenes* par exemple, la GAPDH localisée à la surface cellulaire est impliquée dans

l'adhérence de la bactérie aux cellules hôtes et présente des propriétés antiphagocytaires (Jin 2005, Boël 2005, Madureira 2007). Chez *S. pneumoniae*, la GAPDH a également été identifiée à la surface bactérienne et intervient dans le recrutement du système plasminogène-plasmine, facilitant par ce biais l'invasion des tissus épithéliaux et endothéliaux par le pneumocoque (Ling 2004, Bergmann 2004).

Une des premières lignes de défense de la réponse innée contre tout agent du non-soi, qu'il s'agisse de microorganismes ou de cellules apoptotiques, fait intervenir la voie classique du complément. Des facteurs présents en surface de la cellule à détruire sont reconnus par le complexe C1 (formé de C1q, d'un dimère de C1r et d'un dimère de C1s), ce qui, par une activation en cascade de protéases, conduit à la production de fragments C4b et C3b dont le dépôt à la surface de l'agent « étranger » conduit à la lyse de la cellule ainsi ciblée. Une question importante concerne l'identification des molécules reconnues par C1q en tant qu'étiquette du non-soi. L'équipe dirigée par Philippe Frachet à l'IBS s'intéresse à cette problématique dans un modèle de cellules apoptotiques humaines et a montré que la GAPDH humaine, présente à la surface de cellules HeLa rendues apoptotiques, est un ligand de C1q (Frachet non publié). Un projet de recherche est donc en cours au laboratoire pour étudier l'interaction de la GAPDH présente à la surface du pneumocoque avec C1q et le rôle de cette reconnaissance dans l'échappement du pneumocoque face au système immunitaire.

### **II.3.2 CBPE : un exemple de Choline-Binding Protein**

Les CBPs sont des protéines exportées du côté extracellulaire via la machinerie Sec. Toutes les CBPs sont associées de façon non covalente aux résidus choline des acides teïchoïques et lipoteïchoïques via les Choline-binding Domains (CBDs), formés par la répétition de motifs dont la séquence consensus est **GWVKDNGTWYYLNSSGAMAT** (Fernandez-Tornero 2001). Par analyse des génomes des souches R6 et TIGR4, il a été identifié respectivement 10 et 15 gènes codants pour des CBPs (Tettelin 2001, Hoskins 2001, Frolet 2010). Ces protéines sont les mieux caractérisées au niveau structural et fonctionnel et pour leurs propriétés antigéniques. Parmi elles, la protéine CBPE a fait l'objet du travail de thèse de Cécile Attali au laboratoire (2004-2007).



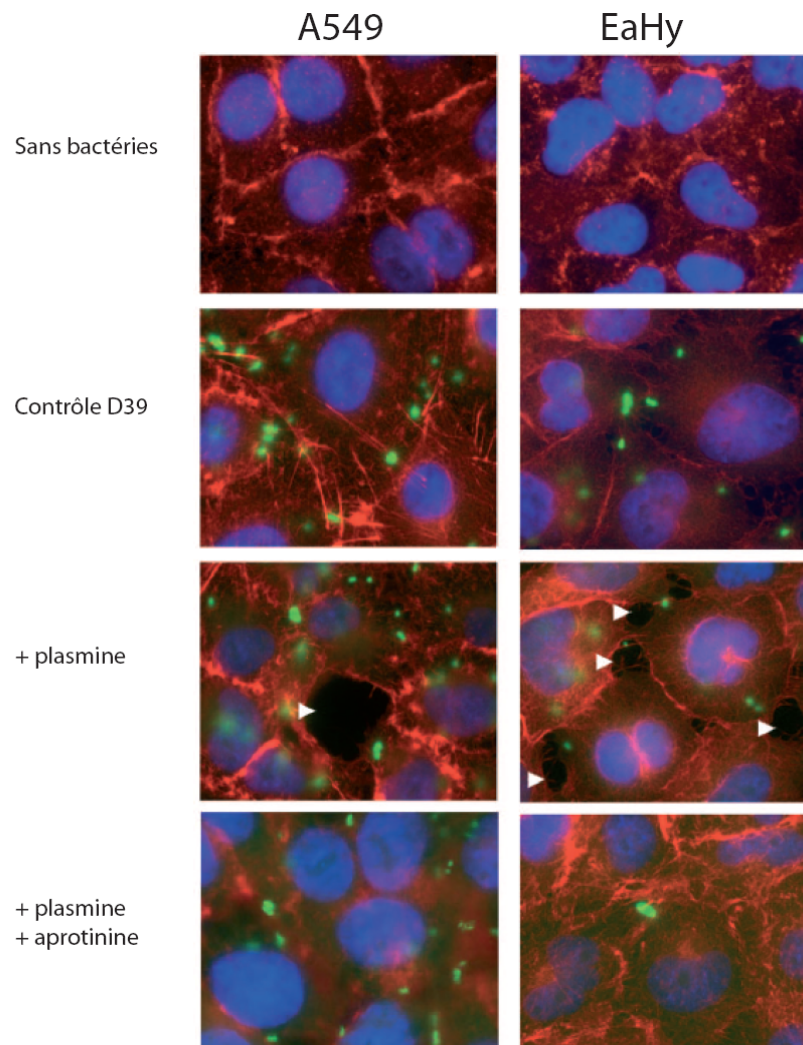
### **II.3.2.1 Structure de la protéine**

CBPE est une protéine de 627 acides aminés qui présente une organisation modulaire composée par : un peptide signal, un domaine catalytique phosphorylcholinestérase (Pce), une région linker puis le domaine CBD formé par la répétition de dix motifs consensus (Garau 2005). L'activité enzymatique Pce permettrait de libérer 20% à 30% des résidus phosphorylcholine provenant des TAs et LTAs de la paroi bactérienne, mais bien que mesurée, le rôle physiologique de cette activité Pce n'est pas encore clairement comprise (de las Rivas 2001, Vollmer 2001, Hermoso 2005). Les structures cristallographiques de la protéine entière et de la seule partie catalytique Pce complexée avec une phosphorylcholine, ont permis, (i) de classer CBPE parmi la superfamille des metallo- $\beta$ -lactamases bien que ses propriétés catalytiques la rapproche plus de la famille des metallo-phosphatases, (ii) de déterminer les résidus impliqués dans la liaison du substrat et dans la catalyse, (iii) d'expliquer pourquoi la quantité de résidus de phosphorylcholine libérés est limitée (Garau 2005, Hermoso 2005). Ce dernier point s'explique grâce à la modélisation de reconnaissance des TAs par CBPE indiquant que seuls les résidus de phosphorylcholine situés à l'extrémité des TAs seraient accessibles au site actif de l'enzyme (Garau 2005).

### **II.3.2.2 Rôle de la protéine dans la virulence**

En 2000, Gosink et ses collaborateurs démontrent que les bactéries mutantes  $\Delta cbpE$  présentent un défaut de colonisation du nasopharynx et que cette anomalie résulte d'une diminution de l'adhérence des bactéries aux cellules humaines due à l'absence de CBPE (Gosink 2000). Une autre fonction de CBPE dans la virulence du pneumocoque a été démontrée au laboratoire par Cécile Attali qui a mis en évidence la capacité de CBPE à lier le plasminogène et à l'activer en plasmine (Attali 2008). En plus de son rôle dans la fibrinolyse, le système protéolytique plasminogène-plasmine intervient dans la migration des cellules humaines en dégradant la matrice extracellulaire entourant les cellules ainsi que les jonctions intercellulaires favorisant le passage de cellules à travers les tissus (Waltz 1997). Dans son étude, Cécile Attali montre que la plasmine détournée par le pneumocoque via sa liaison à CBPE permet la dégradation de la matrice extracellulaire (Attali 2008). De plus, lorsque les bactéries, encapsulées ou non (D39 ou R6) sont préincubées avec la plasmine, elles dégradent les jonctions intercellulaires de cellules endothéliales EaHy et épithéliales A549 (Fig.8). Cette démonstration est un élément clé de la compréhension des mécanismes de détournement du

système plasminogène-plasmine mis en place par la bactérie pour faciliter son invasion à travers les tissus humains, et le rôle de CBPE dans ce processus.

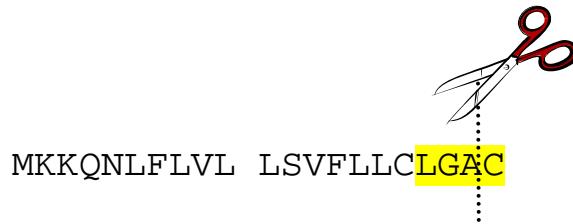


**Figure 8: Immunofluorescence de monocouches de cellules A549 et EaHy infectées par la souche *S. pneumoniae* D39.** En vert, les bactéries marquées au FITC. En rouge, le réseau VE-Cadhérine des cellules et en bleu le noyau des cellules marqué au DAPI. La flèche blanche représente les « fissures » des jonctions intercellulaires (Attali *et al.* 2008)

### II.3.3 AdcAII : un exemple de lipoprotéines

Toutes les protéines appartenant à la classe des lipoprotéines présentent à leur extrémité N-terminale un motif  $Lx_1x_2C$  ou Lipobox qui constitue un site d'attachement covalent aux têtes polaires des phospholipides membranaires. Ce motif a ainsi permis de prédire la présence de 42 lipoprotéines potentielles codées dans le génome de la souche R6 (Hutchings 2009). La Lipobox est reconnue par la peptidase de type II qui clive à la fois le peptide signal et le motif d'ancrage, entre le résidu  $x_2$  et la cystéine, générant ainsi la lipoprotéine mature

(Fig.9). La liaison covalente se forme alors entre ce résidu cystéine et le diacylglycérol des lipides membranaires (Jedrzejjas, 2007).



**Figure 9: Peptide signal de AdcAII.** (Q97R34 ou locus SP\_1002 dans génome TIGR4). Le motif Lipobox est surligné en jaune.

La protéase de type II ou Lsp (pour Lipoprotein signal peptidase) n'a été identifiée que très récemment chez *S. pneumoniae* par analogie de séquence avec celle de *E. coli* (Khandavilli 2008). Cette étude montre que Lsp est essentielle à la fonction d'un certain nombre de lipoprotéines dont les transporteurs de type ABC (pour ATP-Binding Casette). La superfamille des transporteurs ABC a été identifiée dans les années 80 chez les eucaryotes, les procaryotes et les archaebactéries et intervient dans l'import et l'export de substrats variés (Chen 1986, Gros 1986). Ces transporteurs se composent de plusieurs protéines qui s'organisent selon une topologie conservée : deux domaines transmembranaires (TMD pour TransMembrane Domain) générant un canal dans la membrane plasmique et deux domaines cytoplasmiques (NBD pour Nucleotide Binding Domain) qui sont impliqués dans l'hydrolyse de l'ATP, source d'énergie du transport. Chez les bactéries, dans le cas des systèmes d'import, une troisième protéine extracellulaire est indispensable, le SBR (pour Substrate Binding receptor). Celui-ci a pour rôle de fixer le substrat (ex: vitamines, ions...) présent dans l'environnement bactérien afin de le présenter au TMD du système ABC. Neuf classes de SBR ont été identifiées à ce jour dont la dernière nommée MBR pour Metal-binding Receptor et qui regroupe des protéines ayant un rôle présumé dans la fixation des ions manganèse et zinc (Tam et Saier 1993, Dintilhac 1997). Au laboratoire, Claire Durmort et sa doctorante Elodie Loisel se sont intéressées à une protéine nommée AdcAII qui présente une forte identité de séquence avec des MBRs connus, notamment PsaA et AdcA chez *S. pneumoniae*.

### II.3.3.1 Structure de la protéine

AdcAII est une protéine de 305 acides aminés qui présente à son extrémité N-terminale, en plus de son peptide signal, le motif Lipobox spécifique des lipoprotéines (Fig.9). Grâce à la résolution de la structure cristallographique de AdcAII, Elodie Loisel a confirmé

l'appartenance d'AdcAII au groupe des MBRs et prouvé que cette protéine est capable de lier le zinc (Loisel 2008). En revanche, contrairement aux autres transporteurs de type ABC qui présentent une organisation génétique en opéron, comprenant les gènes codants pour les TMD, NBD et MBR, l'analyse du locus *adcAII* révèle une organisation atypique.

### II.3.3.2 AdcAII est une adhésine

La séquence primaire d'AdcAII présente une similarité de plus de 60% avec des protéines référencées Lmb (pour laminin binding protein). Ces protéines, chez *S. agalactiae* et *S. pyogenes* sont impliquées dans l'adhérence aux cellules endothéliales et épithéliales humaines et cela via la laminine, une protéine de la matrice extracellulaire (Spellerberg 1999, Elsner 2002). Lors de test d'interaction par ELISA, Elodie Loisel a montré que *S. pneumoniae* est capable de lier la laminine. En revanche, la souche mutante  $\Delta$ *adcAII* présente un défaut d'adhérence (environ 25%) indiquant ainsi (i) que AdcAII est impliquée dans l'interaction de la bactérie avec cette protéine, (ii) qu'il existe d'autres composants du pneumocoque qui interviennent dans cette interaction. Par ailleurs l'utilisation de la protéine recombinante AdcAII, toujours dans des tests d'interaction par ELISA, permet de valider la capacité de cette protéine à lier la laminine, mais aussi que cette liaison est inhibée en présence de Zinc. Des études RMN complémentaires expliquent ce phénomène ; lorsque AdcAII est sous sa forme apo (sans zinc) elle est capable de lier la laminine, en revanche lorsqu'elle est sous sa forme holo (liée au zinc) des changements conformationnels au sein de la structure s'opèrent menant à une diminution de fixation à la laminine d'un facteur 20 (Loisel en soumission).

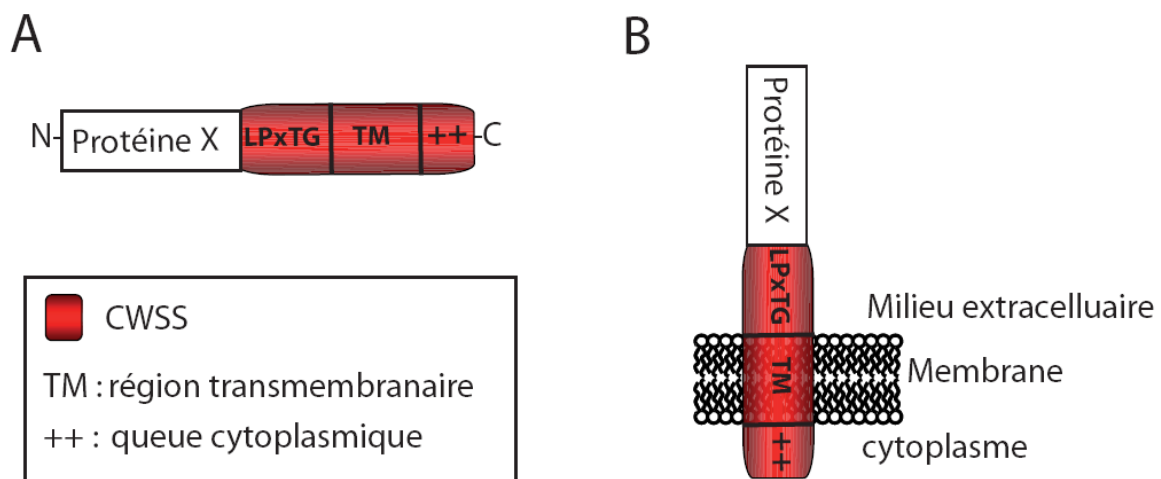
### II.3.4 les protéines LPxTG

En 1972, Sjöquist montre que la protéine A de *Staphylococcus aureus* est covalamment liée à la paroi bactérienne (Sjöquist 1972) et en 1988, Pancholi et Fischetti font le même constat pour la protéine M de *S. pyogenes* (Pancholi et Fischetti 1988 et 1989). Le mécanisme d'ancrage de ces protéines est partiellement décrit : le motif LPxTG (L pour Leucine ; P pour Proline ; x pour acide aminé indifférent ; T pour Thréonine ; G pour Glycine) présent à l'extrémité C-terminale de la protéine est l'élément clé d'ancrage de la protéine au peptidoglycane et cette liaison de type transpeptidation est catalysée par une enzyme membranaire (Pancholi et Fischetti 1988 et 1989). En 1999, le crible génétique de mutants de *S. aureus* présentant une virulence diminuée et déficients dans l'ancrage de protéines

« reporters » mène à l'identification du gène responsable de la fixation des protéines LPxTG au peptidoglycane, il s'agit du gène *srtA* codant pour la sortase A (Mazmanian 1999)

### II.3.4.1 Topologie des protéines LPxTG

Chez les bactéries Gram-positif, les protéines LPxTG, ou plus précisément leurs précurseurs, présentent une organisation similaire, à savoir : un peptide signal à leur extrémité N-terminale qui leur permet d'être exportées à la membrane via la machinerie Sec et un CWSS (pour « Cell Wall Sorting Signal ») à leur extrémité C-terminale qui se compose du pentapeptide LPxTG, suivi d'une région hydrophobe qui sert d'ancrage membranaire et d'une queue cytoplasmique chargée positivement (Schneewind 1992) (Fig.10). Il est à noter qu'un motif additionnel peut être retrouvé au niveau du peptide signal chez certaines protéines de surface, le motif YSIRK-G/S, son absence n'empêche pas l'ancrage à la paroi mais le diminue (Rosenstein 2000, Tettelin 2001, Bae 2003). Chez *S. aureus*, l'utilisation de différentes constructions de l'enterotoxine B, protéine normalement sécrétée dans le milieu, fusionnée aux différents éléments du CWSS a permis de mettre en évidence que le motif LPxTG est le site de coupure protéolytique de l'enzyme sortase alors que la région transmembranaire et la queue cytoplasmique servent de signaux de rétention de la protéine à la membrane durant son export via la machinerie Sec (Schneewind 1993).



**Figure 10: Représentation schématique du « Cell Wall Sorting Signal ».** **A.** Représentation schématique de la topologie du CWSS d'une protéine X à LPxTG. Le CWSS se localise à l'extrémité C-terminale de la protéine. Il contient le motif LPxTG, suivi d'une région hydrophobe qui correspond au domaine transmembranaire de la protéine, et de plusieurs résidus chargés positivement. **B.** Représentation schématique du rôle du CWSS dans la localisation membranaire de la protéine. Le motif LPxTG se trouve du côté extracellulaire. La région hydrophobe est enchâssée dans la membrane plasmique. La queue cytoplasmique contient les résidus chargés positivement.

Sur les quelques 500 protéines potentiellement présentes à la surface de la bactérie, seules une vingtaine ont été caractérisées de façon détaillée pour leur fonction dans la virulence en tant que ligands de composés humains. Pour d'autres, leur implication dans ces mécanismes a été décrite mais leur mode de fonctionnement demeure encore inconnu.

Un certain nombre de ces protéines de surface impliquées dans la virulence ont été qualifiées de candidats vaccins. Ces composants sont probablement les plus intéressantes cibles de futurs vaccins car contrairement aux polysaccharides capsulaires, elles sont plus largement conservées parmi les différents sérotypes (Giefing 2008, Seib 2009).

## **II.4 Les sortases**

### **II.4.1 La sortase A de *S. aureus***

#### **II.4.1.1 Identification**

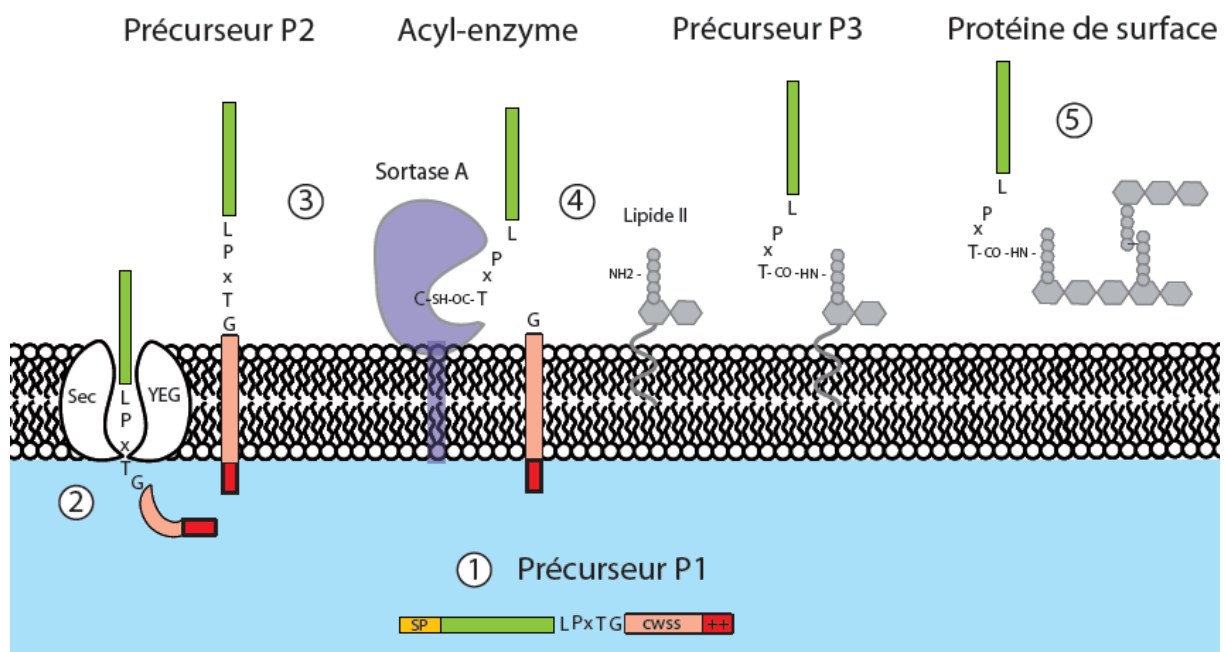
Comme décrit précédemment, l'identification du gène responsable de l'ancrage des protéines LPxTG provient de l'étude de mutants de *S. aureus* déficients pour cette fonction (Mazmanian 1999, Mazmanian 2002). Le gène *srtA* code pour l'enzyme nommée sortase A qui présente des orthologues chez toutes les bactéries Gram-positif (Mazmanian 1999, Hlangovan 2001, Pallen 2001). L'alignement des séquences des différentes sortases révèle la présence d'un peptide signal associé à un domaine d'ancrage membranaire, d'une région linker peu conservée (résidus 26-59 pour SrtA de *S. aureus*) et du domaine catalytique (résidus 60-206) (Hlangovan 2001).

#### **II.4.1.2 Le mécanisme de transpeptidation**

En 1995, Schneewind et ses collaborateurs décrivent le système d'ancrage des protéines LPxTG chez *S. aureus* en fusionnant la Maltose-binding protein de *E. coli* au CWSS de la protéine A. La protéine fusion est ancrée à la paroi de *S. aureus* car son extraction nécessite l'utilisation d'une muramidase (la lysostaphine, endopeptidase glycine-glycine). La protéine ainsi purifiée est ensuite digérée par la trypsine et les peptides générés sont analysés par dégradation d'Edman couplé à la spectrométrie de masse. Les séquences obtenues révèlent que le CWSS de la protéine A est clivé entre la thréonine et la glycine et que le groupement

carboxylique de la thréonine et covalentement lié à une amine de la chaîne latérale du pentaglycine du peptidoglycane (Schneewind 1995).

En 2002, Perry met en évidence plusieurs états intermédiaires lors du processus de liaison covalente des protéines au peptidoglycane (Fig.11). La première espèce P1 correspond au précurseur de la protéine présentant son peptide signal en N-terminal. Le clivage de ce peptide et l'export via la machinerie Sec mène à l'espèce P2, qui localisée dans la membrane, est prise en charge par la sortase A. La reconnaissance du motif LPXTG et le clivage de la liaison peptidique entre les résidus T et G conduit à la formation de l'intermédiaire acyl-enzyme. Celui-ci subit une attaque nucléophile par une amine de la chaîne latérale du lipide II. Ce complexe covalent protéine-lipide II correspond à l'espèce P3 qui est pris en charge par les enzymes de synthèse du peptidoglycane, les Penicillin-Binding Proteins, pour être incorporé au réseau de peptidoglycane (Perry 2002, Ton-That 2002, Ruzin 2002). Ce modèle se base essentiellement sur l'étude de la SrtA de *S. aureus* mais reste néanmoins tout à fait acceptable pour les autres bactéries Gram-positif.



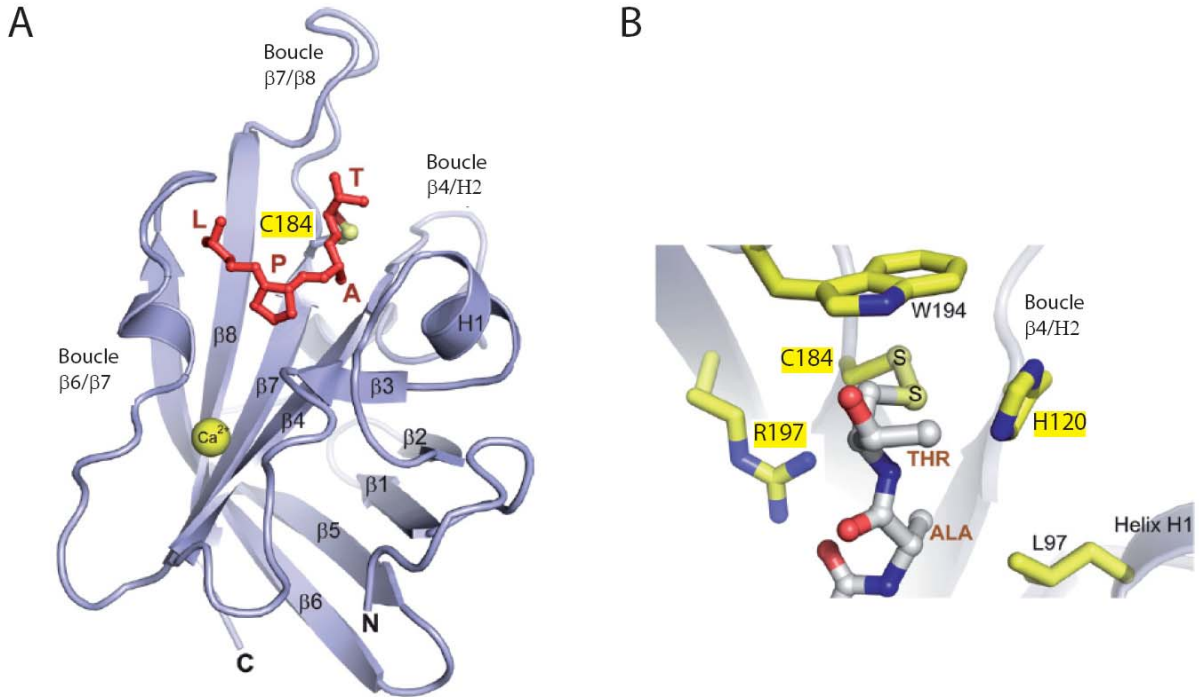
**Figure 11: Représentation schématique du mécanisme de transpeptidation médié par la sortase A permettant l'ancrage des protéines à LPxTG à la paroi des bactéries à Gram positive.** Etape 1. La protéine de surface est cytoplasmique : précurseur P1. Elle présente à son extrémité C-terminale son CWSS et sa queue chargée positivement. A son extrémité N-terminale son peptide signal va lui permettre d'être exporté à la membrane. Etape 2. Le précurseur P1 va transiter via la machinerie Sec où son peptide signal est clivé. La protéine se retrouve alors enchâssée dans la membrane grâce à son CWSS : Précurseur P2. Etape 3 : La sortase A membranaire prend en charge le LPxTG du précurseur P2, clive la liaison Thr-Gly menant ainsi à l'obtention d'un intermédiaire acyl-enzyme. Etape 4 : l'intermédiaire subit une attaque nucléophile d'un groupement amine du lipide II. Cette réaction mène à la liaison covalente entre le lipide II et la protéine de surface : Précurseur P3. Etape 5. Le précurseur P3 est pris en charge pour être incorporé à la synthèse du PG.

### II.4.1.3 Structure tridimensionnelle

La sortase A de *S. aureus* est une protéine de 206 acides aminés qui adopte une topologie de protéine membranaire de type II : une hélice transmembranaire à l'extrémité N-terminale et un domaine catalytique exposé du côté extracellulaire. Différentes études ont montré le rôle catalytique du groupement thiol de la cystéine 184 comme site nucléophile de clivage du motif LPxTG. La substitution de la cystéine 184 en alanine empêche l'activité de la sortase *in vivo* et *in vitro* (Ton-That 2000). De plus, le traitement de la sortase avec de l'hydroxylamine (NH<sub>2</sub>OH) dissocie l'intermédiaire acyl-enzyme ce qui mène au relargage des protéines de surface présentant des thréonines hydroxamates à leur extrémité C-terminale (Ton-That 1999). Enfin, le méthyle méthanesulfonate, un réactif qui forme des ponts disulfures avec les groupements thiols, empêche le clivage du motif LPxTG par la sortase (Ton-That 1999).

La structure tridimensionnelle de la protéine a été initialement déterminée par RMN révélant une organisation en feuillets  $\beta$  parallèles et antiparallèles, incluant deux petites hélices et de nombreuses boucles (Ilangovan 2001) (Fig.12). La structure de la sortase A permet de mieux comprendre son mécanisme enzymatique. En effet, le site actif de la protéine est formé d'une triade catalytique comprenant la cystéine 184, l'histidine 120 et l'arginine 197 et l'efficacité catalytique de l'enzyme dépend de l'intégrité de chacun de ces résidus (Ilangovan 2001, Marraffini 2004). En effet, la substitution des résidus H120 ou R197 en alanine cause une forte diminution de l'activité de l'enzyme (Marraffini 2004). La structure révèle également la présence d'un site de liaison au calcium localisé à proximité du site actif qui stabilise la boucle située entre les feuillets  $\beta_6/\beta_7$  (Ilangovan 2001). En 2009, l'étude par RMN de la SrtA complexée à un peptide LPAT (dont le groupement carboxylique de la thréonine a été remplacé par un groupement thiol permettant de mimer artificiellement l'intermédiaire acyl-enzyme) propose que la reconnaissance du peptide LPxTG induit la fermeture et l'immobilisation de la boucle  $\beta_6/\beta_7$  du site actif (Suree 2009) (Fig.12).





**Figure 12: Structure RMN de la sortase AAN59 de *Staphylococcus aureus* complexée au peptide modifié LPAT\*.** A. Représentation en ruban de la structure de SrtA $\Delta$ N59 avec en rouge le peptide LPAT\* dont la thréonine est modifiée (le groupement carboxylique de la thréonine est substitué par un groupement thiol générant ainsi un pont disulfure entre la thréonine modifiée et la cystéine 184 du site actif de SrtA). Surligné en jaune la cystéine catalytique. B. Zoom sur le site actif de la sortase A. Surlignés en jaune les résidus de la triade catalytique. (Suree *et al.* 2009)

#### II.4.1.4 Paramètres enzymatiques

En 2001, Ton-That et ses collaborateurs étudient les paramètres cinétiques de la sortase A de *S. aureus* en mesurant l'hydrolyse de la liaison peptidique entre la thréonine et la glycine du peptide LPxTG. La sortase A est incubée avec un pentapeptide modifié  $\alpha$ -LPETG-d où  $\alpha$  est un fluorophore et d est un « quencher » de fluorescence. Un signal FRET (Förster Resonance Energy Transfer) s'opère au niveau du pentapeptide du à la proximité du fluorophore et de son « quencher ». En revanche, lorsque la sortase A clive le pentapeptide entre les résidus Thr et Gly, le fluorophore est éloigné de son quencher, il n'y a plus de signal FRET. Cette réaction d'hydrolyse peut être suivie au cours du temps, en mesurant l'augmentation de la fluorescence émise par le composé  $\alpha$ -LPET à la longueur d'onde de 420 nm. Ces expériences ont permis de déterminer les constantes cinétiques d'hydrolyse de la sortase A:  $K_m = 116 \mu\text{M}$  ;  $V_{\text{max}} = 0,48 \mu\text{M}\cdot\text{s}^{-1}$  ;  $K_{\text{cat}} 0,57 \text{ s}^{-1}$  et  $K_{\text{cat}}/K_m = 4,91 \times 10^{-3} \mu\text{M}^{-1}\cdot\text{s}^{-1}$  (Ton-That 2002)

Comme indiqué précédemment, l'activité enzymatique de la sortase A se fait en deux étapes : une première correspond à l'hydrolyse du substrat qui porte le motif LPxTG et à la

formation du complexe acyl-enzyme (acylation), la seconde est l'attaque nucléophile de cet intermédiaire par l'accepteur final (ex: le pentaglycine du peptidoglycane chez *S. aureus*). En 2003, Huang réalise les mêmes expériences d'hydrolyse du peptide mais ajoute au mélange réactionnel de la triglycine, composé qui avait été identifié dans des travaux antérieurs comme pouvant jouer le rôle d'accepteur dans les réactions de transpeptidation *in vitro* (Ton-That 2000). Contrairement aux données obtenues précédemment, Huang rapporte une valeur du  $K_m$  d'hydrolyse de  $20\mu\text{M}$ , soit une affinité d'environ 5 fois plus élevée. Par ailleurs, un  $K_m$  de  $117 \pm 17 \mu\text{M}$  correspondant uniquement à l'étape d'acylation peut être mesuré et explique les données obtenues par Ton-That trois années auparavant. Enfin la dernière phase de transpeptidation ( $K_m = 41 \mu\text{M}$ ) nécessite que l'acyl-enzyme soit déjà formée pour que la triglycine lie l'enzyme (Huang 2003).

A l'heure actuelle, la sortase A de *S. aureus* demeure la seule sortase pour laquelle les études structurales ainsi que les caractérisations enzymatiques sont les plus avancées.

#### II.4.2 La classification des sortases chez les bactéries Gram-positif

L'identification de la sortase A chez *S. aureus* a permis de mettre en évidence la présence d'orthologues chez les autres bactéries Gram-positif (Ilangovan 2001, Pallen 2001). En 2005, Shaynoor Dramsi et coll. proposent un système de classification qui permet de regrouper les sortases en 4 classes en fonction notamment de leurs topologies et de leurs substrats respectifs (Dramsi 2005). L'étude réalisée porte sur 61 sortases identifiées à partir de génomes bactériens totalement séquencés. L'analyse phylogénétique des alignements de séquences de ces sortases met en évidence 4 familles, SrtA, SrtB, SrtC et SrtD qui présentent un certain nombre de caractéristiques communes (Tableau 2). En effet, toutes les sortases présentent un peptide signal à leur extrémité N-terminale et un site catalytique TLXTC. En outre, trois domaines conservés peuvent être identifiés, les domaines D1, D2 et D3 dont deux contiennent les résidus de la triade catalytique, à savoir l'histidine dans le domaine D2, la cystéine et l'arginine dans le domaine D3.

Les sortases de classe A (SrtA), dont le prototype est la sortase A de *S. aureus*, regroupe toutes les sortases dites « ménagères » dont le rôle est de lier à la paroi bactérienne les protéines présentant un motif LPxTG et impliquées dans l'adhésion et l'invasion des cellules hôtes et dans l'évasion au système immunitaire (Patterson 2005, Kharat 2003). A l'exception de deux espèces bactériennes, *Enterococcus faecalis* V583 et *Streptococcus thermophilus*

LMG18311, une seule copie de ce gène est retrouvée par génome et celui-ci est toujours isolé génétiquement (Dramsi 2005).

Les sortases de classe B (SrtB), dont le prototype est la sortase B de *S. aureus*, regroupe les enzymes qui ont pour substrats des protéines impliquées dans l'acquisition de fer. Contrairement à la sortase A, elle ne reconnaît pas les protéines à LPxTG au sens large mais présente une spécificité de reconnaissance pour des pentapeptides spécifiques (ex: NPQTN pour la sortase B de *S. aureus*).

Les sortases de classe C (SrtC) sont les plus nombreuses et ont pour rôle de participer à l'assemblage de sous-unités protéiques (ou pilines) qui forment les pili. En outre, elles possèdent la caractéristique d'être codées dans les mêmes opérons que leurs substrats. Chez certaines bactéries, le nombre de SrtC et de pilines est le même (ex: Les sortases C-1, C-2 et C-3 de *S. pneumoniae* et les trois pilines RrgA, RrgB et RrgC). Cette classe faisant l'objet de notre étude, les sortases qui la composent seront détaillées dans la dernière partie de l'introduction.

Enfin, la classe D (SrtD) regroupe des sortases qui sont impliquées dans des cycles de différenciation de la bactérie telle que la sporulation. Chez *B. anthracis*, la SrtD (appelée SrtC avant la classification proposée par S. Dramsi) est présente dans un locus codant pour un système de régulation à deux composantes et pour une protéine BasI qui présente un motif LPNTA. Un gène *basH* localisé ailleurs dans le génome code pour la protéine BasH qui présente également ce motif LPNTA. Ces deux protéines BasH et BasI ont été identifiées comme étant les substrats de la SrtD. Tous ces gènes sont exprimés durant la phase de sporulation, BasI et BasH sont ancrés à la paroi de la cellule mère et de la pré-spore (Marrafini 2006).

**Tableau 2: Récapitulatif des différentes classes de sortases: rôles et caractéristiques.** En jaune, les sortases de la classe C qui vont faire l'objet de la dernière partie de l'introduction. (Dramsi *et al.* 2005, Marraffini *et al.* 2006, Maresso et Schneewind 2008, Dramsi *et al.* 2008, Clancy *et al.* 2010)

Sortases (nombre)	Exemple	Rôle	Description
Classe A (16)	SrtA ( <i>S. aureus</i> )	- Adhésion - Évasion au SI - Internalisation - Reconnaissance par les phages	- Site catalytique TLXTC (X= Val, Ile ou Thr) - Pentapeptide reconnu: LPxTG - Gène isolé dans le génome
Classe B (7)	SrtB ( <i>S. aureus</i> )	- Acquisition du fer	- Segments B1, B2 et B3 - Site catalytique TLXTC (X généralement Ser) - Pentapeptides reconnus: nombreux (ex: NPQTN) - Gène souvent en opéron avec les gènes des protéines cibles
Classe C (24)	SrtC-1 ( <i>S. pneumoniae</i> )	- Assemblage des pili	- Site catalytique TLXTC (résidu proline conservé en aval de ce motif) - Pentapeptides reconnus: nombreux (ex: IPQTG) - Gène en opéron avec les gènes des protéines cibles
Classe D (14)	SrtD ( <i>B. anthracis</i> )	- Cycles de différenciation (formation de mycélium, sporulation...)	- Site catalytique TLXTC - Pentapeptides reconnus: nombreux (ex: LPNTA)

### II.4.3 Les enzymes sortases chez *S. pneumoniae*

Forts de la connaissance des éléments conservés des protéines LPxTG, Tettelin et Hoskins ont identifié 19 et 13 gènes codant potentiellement pour ce type de protéines dans les souches de pneumocoque TIGR4 et R6 (Tettelin 2001, Hoskins 2001). Certains gènes sont communs aux deux souches, telle que la  $\beta$ - galactosidase, ou PsaA (Kharat 2003, Frolet 2010). Certains gènes sont en revanche spécifiques à chacune des souches (ex : NanA ou encore PclA uniquement identifiés chez R6 Frolet 2010). Trois gènes ne sont présents que dans la souche TIGR4 et possèdent un motif LPxTG non conforme : *rrgA*, *rrgB* et *rrgC* présentant respectivement les motifs YPRTG, IPQTG et VPDTG. La présentation de ces trois gènes sera plus largement développée dans la dernière partie de l'introduction.

Chez le pneumocoque, un gène homologue à celui de la SrtA de *S. aureus* est identifié chez la souche R6 (SPR1098) ainsi que chez la souche TIGR4 (SP1218) et présentent 99% d'identité entre eux contre 18% d'identité avec la sortase A de *S. aureus*. Chez TIGR4, trois gènes supplémentaires sont annotés comme des sortases potentielles *srtC-1*, *srtC-2* et *srtC-3* et seront décrits dans la dernière partie de l'introduction (Tettelin 2001, Kharat 2003). En 2003, Kharat cherche à identifier le rôle de la SrtA chez le pneumocoque et notamment à savoir si comme chez *S. aureus*, elle est responsable de l'ancrage des protéines LPxTG. Il construit donc des souches mutantes  $\Delta$ *srtA* chez R6 et s'intéresse à la localisation de deux protéines LPxTG, la  $\beta$ -Galactosidase et la neuraminidase A (NanA). Cette étude montre que




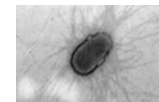
60 à 90% de la  $\beta$ -Galactosidase est relarguée dans le milieu de culture chez les bactéries délétées de *srtA* alors que chez les bactéries sauvages la protéine est associée à la paroi bactérienne, des résultats identiques sont obtenus pour la protéine NanA (Kharat 2003). Les mêmes auteurs montrent que la souche  $\Delta$ *srtA* présente globalement un défaut dans l'adhérence et l'invasion du pneumocoque dans un modèle de cellules pharyngées, ce qui est probablement dû à l'absence de toutes les protéines LPxTG à la surface de la souche mutante (Kharat 2003). En 2005, Patterson et Mitchell poursuivent l'étude et prouvent que l'absence de la sortase A chez une souche de pneumocoque diminue sa virulence dans des modèles de pneumonie et de bactériémie. De plus, cette protéine contribue également à la colonisation nasopharyngée *in vivo* et son rôle dans l'adhérence des bactéries aux cellules humaines *in vitro* est prouvée comme étant dépendante de la présence de la capsule (Patterson 2006). Chez le pneumocoque, il n'existe pas de chaîne latérale de pentaglycine, et de ce fait il semblerait que se soit l'amine libre de la L-Lysine du lipide II qui interviendrait dans l'ancrage des protéines LPxTG au peptidoglycane (Kharat 2003).

### III. Structure et fonction des pili

Les pili (du latin pilus qui signifie « poil »), également appelés fimbriae (terme dérivé du latin pour « frange ») sont des structures fibrillaires présentes à la surface de cellules. Au début des années 1950, l'observation par microscopie électronique de bactéries Gram-négatif met en évidence de fines structures allongées et flexibles qui s'étendent depuis la surface bactérienne vers le milieu environnant (Houwink 1950, Duguid 1955). En raison de leur apparence, l'équipe de Brinton les qualifie de pili, alors que celle de Duguid préfère les nommer fimbriae. Les deux termes étant synonymes, c'est celui de pili qui sera utilisé dans ce manuscrit. Dès lors, et jusqu'à aujourd'hui, l'étude des pili bactériens est un domaine d'investigation très dynamique questionnant leur structure, leur mode d'assemblage, les modifications post-traductionnelles qu'ils subissent, la régulation de leur expression et bien entendu leur rôle dans les interactions avec l'organisme hôte, qu'il s'agisse d'une relation commensale ou pathogène.

La grande variété des pili exprimés par les bactéries Gram-négatif a rendu nécessaire leur classification qui tient compte notamment de leur mécanisme d'assemblage. Quatre groupes ont été défini : les pili de type I, de type IV, les « curli-pili » et la famille des pili CS1 (Proft and Baker 2009, Telford 2006) (Tableau 3). Indépendamment des particularités propres à chacun de ces groupes, une caractéristique commune et spécifique aux pili des bactéries Gram-négatif est la liaison non covalente reliant les sous-unités protéiques ou pilines.

**Tableau 3 : Les différents types de pili des bactéries Gram-négatif.**

	<b>Pili Type I</b>	<b>Pili Type IV</b>	<b>« Curli pili »</b>	<b>Pili CS1</b>
<b>Exemples</b>	<i>E. coli</i> (EPEC)	<i>E. coli</i> , <i>Neisseria</i> et <i>Pseudomonas</i>	<i>E. coli</i> (certaines souches)	<i>E. coli</i> (ETEC)
<b>Microscopie électronique</b>	 Telford <i>et al.</i> 2006	 Telford <i>et al.</i> 2006	 Telford <i>et al.</i> 2006	 Sakellaris <i>et al.</i> 1999
<b>Longueur et diamètre</b>	1-2 $\mu\text{m}$ ; 7 nm (sur la longueur) 2-3 nm (à l'extrémité)	1-2 $\mu\text{m}$ ; 5-6 nm	1-2 $\mu\text{m}$ ; 1-2 nm	1-2 $\mu\text{m}$ ; 5-7 nm
<b>Biogenèse du pilus</b>	Voie Chaperone/Usher	Système de sécrétion de type II	Extracellulaire	Voie alternative Chaperone/Usher

L'identification des pili en surface des bactéries Gram-positif date des années 70 et leur analyse a nécessité d'adapter le protocole utilisé pour les bactéries Gram-négatif (Yanagawa et Honda 1976). Dans ce cas, l'extraction des pilines est réalisée par traitement à la chaleur et leur identification individuelle par analyse sur gel SDS-PAGE. Ce protocole appliqué aux bactéries Gram-positif n'étant que très peu efficace, des méthodes alternatives ont été développées. L'une utilise une enzyme muramidase qui clive la paroi cellulaire de la bactérie et une seconde fait appel à la solubilisation de l'extrait bactérien par l'acide formique. Dans le premier cas, l'analyse par immuno-blot des échantillons traités permet l'identification de bandes individuelles formant une échelle de haut poids moléculaires, dans le second cas, seule une « trainée » est visible. Ces observations sont à l'origine de la proposition selon laquelle les pili des bactéries Gram-positif sont composés de sous-unités liées covalentement les unes aux autres (expliquant leur résistance au SDS) et que la fibre elle-même est liée covalentement à la paroi cellulaire.

Les bactéries Gram-positif ont ainsi développé un système unique assurant l'expression à leur surface d'une structure fibrillaire extrêmement résistante. Ces pili sont-ils représentés chez toutes les bactéries Gram-positif et quelles sont leur composition ? Comment sont-ils régulés ? Quels sont les rôles des pili chez ces bactéries ? Quels sont les mécanismes d'assemblage de ces structures ? Des réponses à ces questions sont proposées ici, en établissant un parallèle entre les découvertes de majeure importance réalisées chez certaines bactéries Gram-positif et notre sujet d'étude, le pneumocoque.

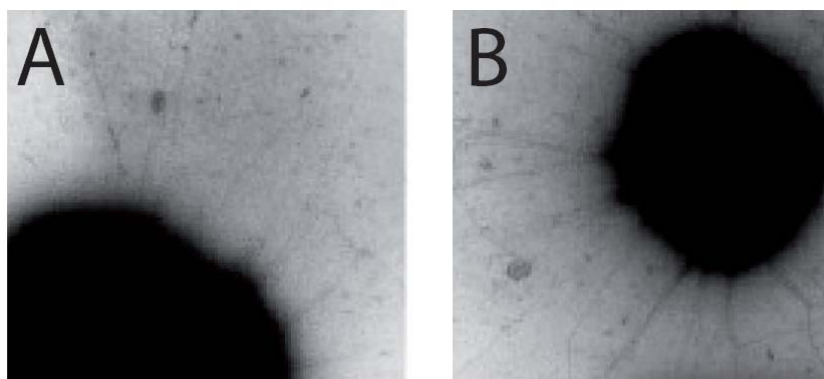
### **III.1 Identification des pili chez les bactéries Gram-positif**

Comme indiqué précédemment, les premières identifications de structures fibrillaires à la surface de bactéries Gram-positif datent des années 1970. L'observation par microscopie électronique de souches de *Corynebacterium diphtheriae* met en évidence des structures flexibles et très fines, de diamètre compris entre 2.5-3.5 nm et de longueur variant entre 0.2-1.6 µm (Yanagawa et Honda 1976). Dans les années 1980, des pili sont observés à la surface de spores de *Bacillus cereus* mais ce n'est qu'au début des années 2000 que ce champ d'activité prend de l'ampleur et que les premières études génétiques et moléculaires sont réalisées chez *C. diphtheriae*, *Actinomyces naeslundii*, *S. agalactiae*, *S. pyogenes*, *E. faecalis* et *B. cereus* (Ton That 2003, Wu 2001, Lauer 2005, Mora 2005, Nallapareddy 2006, Budzik 2007).

### III.1.1 Identification des pili chez le pneumocoque

En 2002, l'équipe de Camilli effectue un crible d'insertion de 6149 transposons dans la souche TIGR4 et identifie des éléments déterminants pour la virulence de la bactérie dans des modèles murins de pneumonie (Hava 2002). Parmi ces facteurs de virulence, un opéron de 12 kb flanqué de séquences d'insertion attire leur attention car les sept gènes composant cet îlot génétique présentent des similarités de séquences avec des éléments déjà connus. Cet îlot est nommé *rlrA*, pour Rof-like régulateur car le gène *rlrA*, codé de façon divergente par rapport aux autres gènes, est un homologue du régulateur RofA chez *S. pyogenes*. Cet opéron est qualifié *a posteriori* PI-1 (Pathogenicity Islet-1), car un second pilus PI-2 a été identifié en 2008 (Bagnoli 2008). Par souci de clarté, nous utiliserons dans ce manuscrit la dénomination PI-1.

Les trois gènes en aval de *rlrA* présentent une similarité de séquences avec la famille des MSCRAMMs (Microbial Surface Components Recognizing Adhesive Matrix Molecules) et codent pour des protéines à motif LPxTG. Ceux-ci sont nommés *rrgA*, *rrgB* et *rrgC* (pour RlrA-Regulated gene). Les trois derniers gènes de l'opéron codent pour des sortases. Etant donné la présence de la sortase ménagère SrtA, ces trois sortases supplémentaires sont dénommées *srtB*, *srtC* et *srtD* lors de leur identification (Barocchi 2006). La classification des sortases proposée par S. Dramsi en 2005 permet de classer ces trois sortases dans la classe C, c'est la raison pour laquelle nous utiliserons cette nomenclature : *srtB*, *srtC* et *srtD* sont dorénavant dénommés respectivement *srtC-1*, *srtC-2* et *srtC-3*. Le rôle de cet opéron dans la formation des pili chez le pneumocoque est démontré en 2006 par observation en microscopie électronique de souches TIGR4 (Barocchi 2006) (Fig.13 et 14).



**Figure 13:** Pili du PI-1 de *S. pneumoniae*. **A.** Coloration négative de la souche TIGR4. **B.** Coloration négative de la souche TIGR4Δ*mgrA* qui exprime une quantité plus importante de pili (Barocchi *et al.* PNAS 2006)



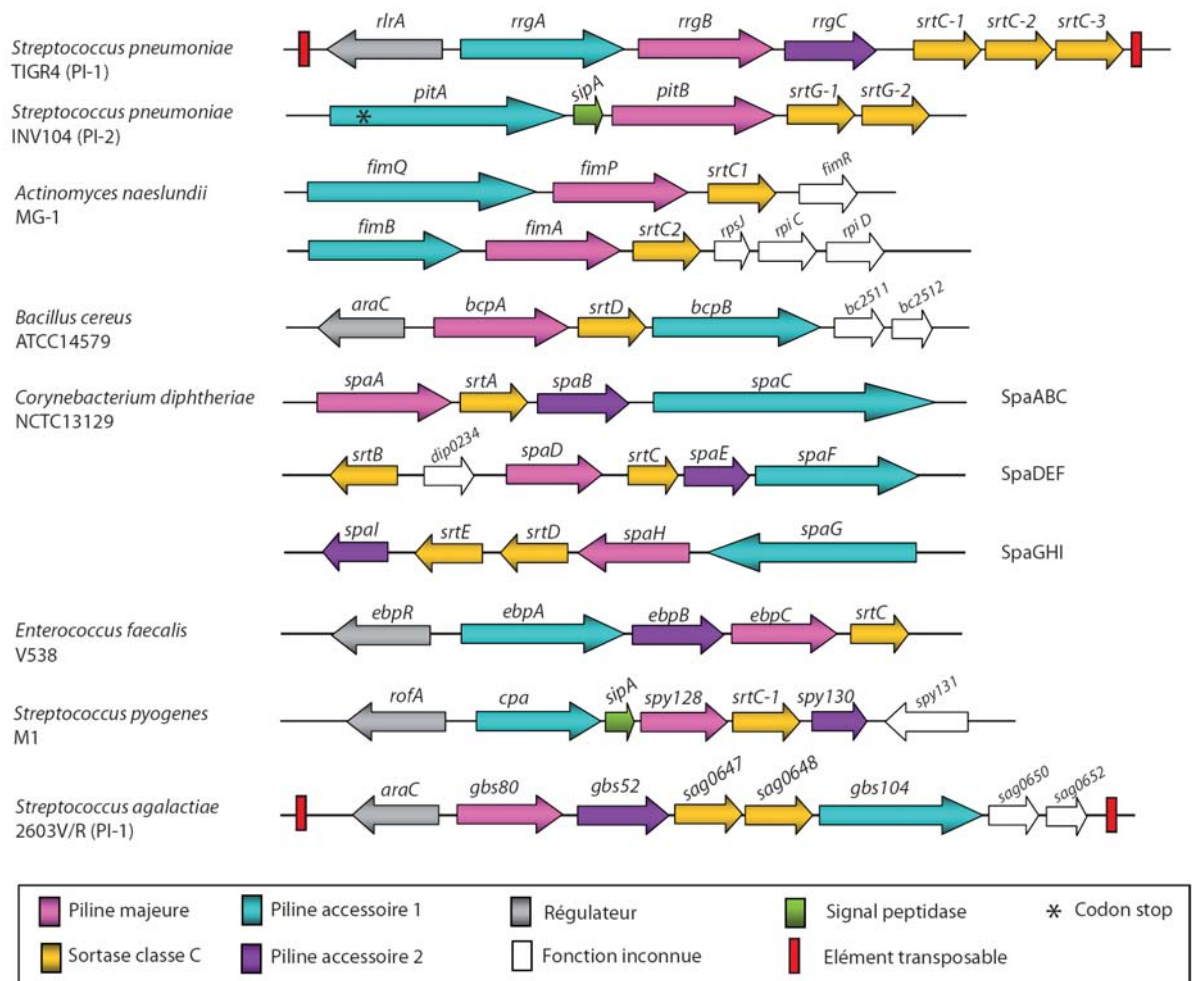
Comme déjà mentionné, un opéron codant pour un second pilus, PI-2, a été identifié dans la souche clinique INV104 (Bagnoli 2008) (Fig.14). Cinq gènes sont présents dans cet élément génétique de 6575 pb et certains d'entre eux présentent une similarité de séquence avec les éléments du pilus FCT-3 identifié chez *S. pyogenes* (Mora 2005, Telford 2006). Le premier gène code pour la protéine PitA, similaire à l'adhésine de FCT-3. *pitA* est suivi par deux éléments génétiques fonctionnellement dépendants, *sipA* et *pitB*. Il est proposé que SipA soit une signal peptidase s'associant à la piline majeure PitB. Enfin, les deux derniers gènes identifiés codent pour deux sortases, SrtG1 et SrtG2 (Bagnoli 2008).

PI-1 n'est présent que dans 30 % des souches de pneumocoque et dans 50% des souches résistantes aux antibiotiques (Aguar 2008, Moschioni 2008). La présence du PI-2 est associée aux sérotypes non vaccinaux émergents avec une distribution d'environ 16% dans une collection de 305 isolats cliniques (Bagnoli 2008).

### III.1.2 Organisation génétique des pili chez les Gram-positifs

Le nombre croissant de génomes bactériens séquencés permet de comparer l'organisation génétique des opérons codant pour les pili et de tenter d'en extraire des informations concernant les mécanismes d'assemblage et/ou la fonction de certaines pilines et sortases (Fig.14). L'expression « piline majeure » désigne la sous-unité constituant la fibre du pilus et « piline accessoire » une autre protéine associée au pilus. Ces notions seront décrites de façon plus détaillée dans la suite du manuscrit.

Un gène codant pour la piline accessoire 1 et pour la piline majeure sont présents dans toutes les espèces bactériennes étudiées, ce qui n'est pas le cas de la piline accessoire 2, suggérant une fonction essentielle uniquement pour les deux premières pilines (Fig.14). Le nombre de sortases est variable, une ou deux sortases sont communément retrouvées par opéron. Le cas du pneumocoque, qui présente trois sortases dans le PI-1 reste actuellement unique. Cette observation résume à elle seule une question importante concernant le mode d'action des sortases et leur spécificité de substrat : comment une seule sortase peut-elle catalysée l'association covalente de deux, voire de trois pilines ? Réciproquement, pourquoi trois sortases sont-elles nécessaires à l'assemblage du pilus chez le pneumocoque ? Un gène codant pour un régulateur de l'opéron de type *rofA* (M1 de *S. pyogenes*) est identifié chez *S. pneumoniae* (*rlrA*), *B. cereus* (*araC*), *E. faecalis* (*ebpR*) et *S. agalactiae* (*araC*), suggérant qu'une régulation de l'expression des gènes pilines et sortases doit avoir lieu *in vivo*.



**Figure 14: Organisation génétique des îlots de pathogénicité codant pour des pili chez les bactéries Gram-positif.** (Wu *et al.* 2001 ; Budzik *et al.* 2007 ; Ton-That et Schneewind 2004 ; Nallapareddy *et al.* 2006 ; Mora *et al.* 2005 ; Lauer *et al.* 2005 ; Hava *et al.* 2002 ; Bagnoli *et al.* 2008)

## III.2 La fonction des pili

### III.2.1 Chez le pneumocoque

Comme décrit dans le chapitre II, le pneumocoque a développé un grand nombre de facteurs lui permettant de survivre dans des environnements de nature variée, d'augmenter sa capacité à envahir l'organisme hôte ou encore d'échapper au système immunitaire. Etant donnée sa localisation en surface et les similarités de séquences des pilines accessoires 1 avec les MSCRAMMs, le pilus peut potentiellement jouer un rôle dans la virulence de la bactérie. L'utilisation de souches TIGR4 de type sauvage et mutante pour la délétion de l'îlot de pathogénicité PI-I indique que la souche mutante est moins virulente que la souche sauvage dans des modèles murins de colonisation, de pneumonie et de bactériémie. Ce défaut de virulence est restauré lorsque l'expression de l'îlot PI-1 est complétement (Barocchi 2006). Par

ailleurs, l'infection intra-péritonéale de souris par la souche TIGR4 conduit à une réponse inflammatoire TNF élevée dans laquelle le pilus PI-1 paraît jouer un rôle (Barocchi 2006). Dans un modèle cellulaire *in vitro*, l'introduction de PI-1 dans des souches encapsulées D39 (souches de pneumocoque naturellement dépourvues de PI-I) augmente la capacité de la bactérie à adhérer aux cellules épithéliales pulmonaires humaines A549 (Barocchi 2006). Ces données indiquent que le pilus peut être qualifié de facteur de virulence car impliqué dans la colonisation des tissus épithéliaux par le pneumocoque mais également dans les processus pathologiques.

Les premiers éléments de réponse concernant l'identification des composants du pilus directement impliqués dans ces fonctions d'interactions avec l'hôte proviennent des cribles STM (Signature Tagged Mutagenesis) réalisés par Hava en 2002. Dans cette étude, l'utilisation d'un mutant TIGR4 $\Delta$ *rlrA* conduit à un défaut de virulence localisée dans le nasopharynx. Or, ce régulateur contrôle l'expression de tous les gènes du PI-1 (voir chapitre suivant) ce qui suggère qu'au moins l'un d'entre eux est impliqué dans l'interaction de la bactérie avec les muqueuses de l'appareil respiratoire. L'analyse de simple et double mutants indique que seules les souches présentant des transposons insérés dans *rrgA* ou *srtC-3* montrent des défauts de virulence dans un modèle murin de compétition au niveau des poumons, alors que des défauts de colonisation sont observés chez les doubles mutants *rrgA* et *srtC-1* (Hava 2002).

Ces travaux préliminaires identifiant RrgA comme un élément important dans la fonction d'adhésion du pneumocoque ont été confirmés ultérieurement. La souche TIGR4 $\Delta$ *rrgA* présente un taux d'adhérence aux cellules épithéliales A549 huit fois moins élevé que la souche sauvage ou la souche TIGR4 $\Delta$ *rrgBC*, qui par ailleurs présentent toutes deux des niveaux d'adhérence comparables (Nelson 2007). Par contre, dans une souche TIGR4 exprimant une copie supplémentaire de *rrgA*, l'adhérence est plus élevée d'un facteur 1,6 par rapport à la souche parentale (Nelson 2007). D'autre part, la forme recombinante de la protéine RrgA adhère directement aux cellules respiratoires et peut entrer en compétition avec la bactérie piliée pour l'interaction avec les récepteurs cellulaires (Nelson 2007). Hilleringmann et ses collaborateurs montrent par ailleurs par des tests d'interaction en phase solide que la protéine recombinante RrgA reconnaît des composants de la matrice extracellulaire dont la fibronectine, le collagène I et la laminine (Hilleringman 2008). En conclusion, RrgA porte les fonctions adhésives du pilus envers les tissus humains, en reconnaissant des récepteurs cellulaires dont la nature reste encore méconnue. RrgA semble

également interagir avec des composants de la bactérie car son rôle dans la formation de biofilms a aussi été montré (Munoz-Elias 2008).

### III.2.2 Chez d'autres bactéries Gram-positives

Comme décrit pour le pneumocoque, la majorité des pili exprimés à la surface des bactéries Gram-positif interviennent dans les étapes d'adhérence aux cellules de l'organisme hôte. Chez *C. diphtheriae*, les pilines SpaB et SpaC du pilus *spaABC* sont impliquées dans l'adhérence de la bactérie aux cellules épithéliales humaines évoquant un rôle lors de l'étape de colonisation (Mandlik 2007). D'autre part, le taux d'adhérence est dépendant du type cellulaire employé indiquant une fonction possible du pilus de *C. diphtheriae* dans le tropisme cellulaire (Mandlik 2007). Les pili exprimés en surface de la souche M1 de *S. pyogenes* permettent l'adhésion des bactéries à des cellules épithéliales issues d'amygdales mais aussi à des kératinocytes, ces deux tissus représentant les deux principaux sites d'infections de *S. pyogenes* chez l'homme (Abbot 2007). Le rôle du pilus dans la formation de biofilms a été également décrit chez *S. pyogenes* (Manetti 2007). Bien que la fonction des pili demeure encore imprécise et étudiée principalement dans des modèles cellulaires *in vitro* ou des modèles murins d'infection, les pili sont exprimés *in vivo* dans l'organisme hôte comme le montre l'exemple de patients ayant déclarés une pharyngite à *S. pyogenes* et chez lesquels des anticorps anti-piline ont été détectés (Manetti 2007).

## III.3 Régulation transcriptionnelle du PI-1 chez *S. pneumoniae*

### III.3.1 *rlrA* : l'activateur

Dès l'identification de l'îlot de pathogénéicité PI-1, la fonction de régulateur transcriptionnel positif du gène *rlrA* est proposé (Hava 2002). En effet, lorsque ce gène est muté ou délété, le niveau des ARNm de chacun des 6 gènes du PI-1 est drastiquement diminué (Hava 2002). De plus, chez une souche contenant une seconde copie de *rlrA* inductible au maltose, la même équipe a montré par des expériences de « RNAse Protection Assay » qu'en présence de maltose, les quantités d'ARNm du gène *rlrA* natif sont plus importantes, indiquant que RlrA a aussi une capacité d'autoactivation (Hava 2003). Il n'est cependant toujours pas connu quels sont les facteurs qui agissent sur l'expression de *rlrA* lui-même.

La transcription du PI-1 est initiée au niveau de quatre sites promoteurs différents, placés en amont de chacun des gènes *rlrA*, *rrgA*, *rrgB* et *srtC-1*. Ces informations suggèrent

que les autres gènes de l'opéron, *rrgC*, *srtC-2* et *srtC-3* sont probablement co-transcrits à partir d'un promoteur distal. Trois des quatre sites de fixation de RlrA (riches en AT) ont été identifiés et ont permis de déterminer la séquence consensus suivante : **AYNTTTTATCAA** (Hava 2002). Cette séquence consensus de fixation de RlrA a été retrouvée dans 153 autres sites du génome de TIGR4 : 29 sont présents dans des régions promotrices et 14 sont localisés à 70 pb en amont d'un site d'initiation de transcription (Hava 2002). Cette étude suggère que l'expression d'un grand nombre de gènes, autres que ceux du locus PI-1, est également sous la dépendance de RlrA.

### III.3.2 *mgrA* : le répresseur

En 2003, Camilli et son équipe identifient par criblage STM de la souche TIGR4 un régulateur transcriptionnel codé par le locus *sp1800* (Hemsley 2003). Ce gène est appelé *mgrA* (pour Mga-like repressor A) en raison des 51% de similarité et 25% d'identité qu'il présente avec le régulateur *mga* présent chez les streptocoques de groupe A. Contrairement à son orthologue, le régulateur MgrA n'affecte aucunement les gènes de son environnement proche. En revanche, des gènes éloignés dans le génome sont surexprimés chez un mutant de délétion *mgrA*: il s'agit de l'opéron PI-1 (Hemsley 2003). Les auteurs proposent que MgrA pourrait agir soit sur chaque site promoteur de l'îlot soit uniquement sur *rlrA* ce qui affecterait indirectement les autres gènes du PI-1. La répression de l'expression du PI-1 par MgrA a été validée par Barrochi et ses collaborateurs en 2006 car l'analyse par FACS et par microscopie électronique à transmission de la souche TIGR4 $\Delta$ *mgrA* montre que celle-ci produit une quantité plus importante de pili que la souche sauvage (Barrochi 2006) (Fig.13).

### III.3.3 Les autres régulateurs du PI-1 chez le pneumocoque

Chez toutes les bactéries, l'homéostasie des ions divalents est essentielle à leur survie et/ou à leur croissance. Selon les différents compartiments de l'hôte, les concentrations locales en ions peuvent être extrêmement variables : à titre d'exemple, la concentration en  $Mn^{2+}$  dans la salive est de 36 $\mu$ M contre 20nM dans le sérum et le plasma. C'est la raison pour laquelle les microorganismes possèdent des systèmes de transport finement régulés de ces solutés.

Chez *S. pneumoniae*, le transport de  $Mn^{2+}$  est réalisé par une perméase de type ABC codée par les gènes *psaBCA*. Le  $Mn^{2+}$  est critique pour de nombreux processus métaboliques de la bactérie ce qui explique que l'inactivation de l'un des gènes de l'opéron *psaBCA* mène à d'importants défauts de virulence et à une limitation de la croissance bactérienne (McAllister

2004). Cet opéron est régulé par une protéine PsaR, qui agit uniquement lorsque la concentration en  $Mn^{2+}$  est importante. De façon intéressante, PsaR présente également un effet au niveau des gènes du PI-1 qui se traduit par une répression transcriptionnelle du régulateur *rlrA* conduisant donc à une diminution de l'expression des pili (Johnston 2006). PsaR agit comme un senseur du  $Mn^{2+}$  de l'environnement bactérien, il est possible de proposer les mécanismes suivants de régulation de l'expression des gènes du pilus : dans des compartiments comme le nasopharynx où la concentration en  $Mn^{2+}$  est supposée importante, PsaR aurait un effet négatif sur la transcription des gènes du PI-1, à l'inverse, dans des compartiments pauvres en  $Mn^{2+}$  comme les poumons ou la circulation sanguine, PsaR serait inactif permettant alors l'expression des gènes du PI-1.

D'autres éléments, également connus pour être des senseurs de l'environnement bactérien et contrôlant l'expression des gènes, sont proposés réguler les gènes de l'îlot *rlrA* : les systèmes à deux composantes TCS08 et TCS09 (Song 2009, Hendriksen 2007). De façon générale, les TCS sont composés par une protéine membranaire, HK Histidine Kinase qui est le senseur de l'environnement et une protéine RR cytoplasmique, régulatrice de la réponse. En réponse à la stimulation de la protéine HK, la protéine RR est phosphorylée et se lie aux promoteurs des gènes cibles. Le génome du pneumocoque contient 13 TCS complets et un régulateur orphelin (Lange 1999, Tettelin 2001). Des travaux ont montré que le système à deux composants TCS08 est un élément essentiel dans les processus de croissance et de survie du pneumocoque *in vivo* (Throup 2000). En 2009, l'équipe de Song constate que la double délétion des gènes *rr08* et *hk08* ou la simple délétion de la composante régulatrice *rr08* induit la surexpression des gènes du pilus de deux à sept fois supérieure à celle observée chez la souche sauvage, ce qui suggère que ce système TCS08 régule négativement les gènes du PI-1. Par ailleurs, cette régulation a la particularité d'être plus marquée durant la phase de croissance logarithmique tardive (Song 2009).

Le système TCS09 a également pour cible le PI-1 (Hendriksen 2007). L'étude, réalisée avec un mutant de délétion *rr09*, a permis de comparer les profils d'expressions de *rrgA* et *rlrA* en fonction du milieu dans lequel les bactéries sont cultivées (répression en THY et surexpression BHI) mais également en fonction de la phase de croissance considérée (une forte répression à des densités optiques faibles et une surexpression lors des phases plus tardives). Dans un modèle murin d'infection comparant TIGR4 et TIGR4 $\Delta$ *rr09*, les niveaux d'expression de *rrgA* s'avèrent 10 à 90 fois plus élevés dans la souche sauvage, ce qui suggère bien que RR09 présente un effet activateur de l'expression du PI-1.

Enfin en 2008, Rosch démontre que le PI-1, qui est un élément probablement acquis par recombinaison homologue, s'est intégré aux réseaux de régulation transcriptionnel préexistants dans la bactérie qui contrôlent l'expression des facteurs de virulence importants dans l'adhésion et l'invasion (Rosch 2008).

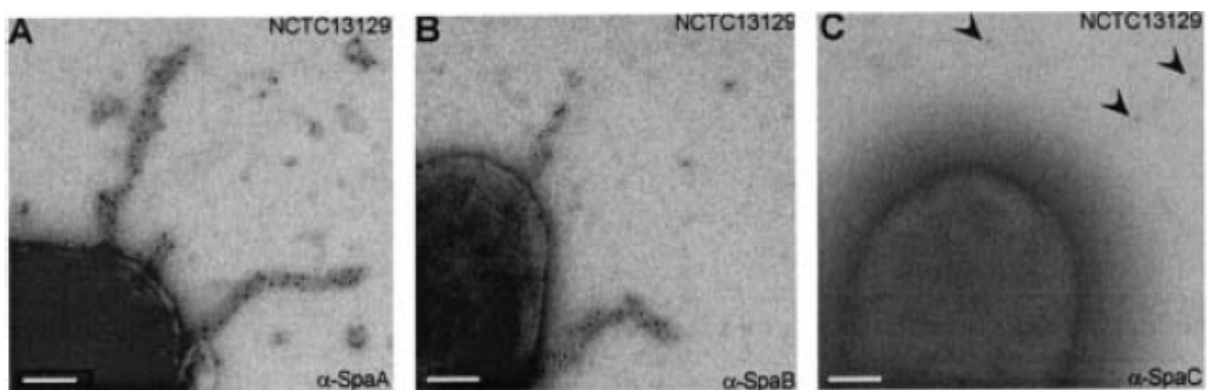
### **III.4 Les pilines**

Les pilines sont les sous unités protéiques structurales composant le pilus. Il s'agit dans tous les cas de protéines appartenant à la famille des protéines LPxTG, décrites au chapitre II.4 : elles sont exportées vers l'espace extracellulaire via un peptide signal mais retenues dans la membrane plasmique au niveau du segment transmembranaire présent en C-terminal. Le motif de type LPxTG est localisé du côté extracellulaire, juste en amont de l'ancrage membranaire. La conservation de la séquence et de la topologie du CWSS, contenant le motif LPxTG, le segment hydrophobe et une région chargée positivement, facilite l'identification des gènes codant pour les pilines dans les îlots de pathogénicité, qui, selon l'espèce bactérienne considérée, sont présents au nombre de deux ou trois.

L'identification génétique des îlots de pathogénicité regroupant en opéron des gènes codant pour des protéines exposées en surface bactérienne via le motif LPxTG et les enzymes sortases catalysant l'association covalente de ces protéines par transpeptidation, a suggéré très fortement la présence de pili en surface des bactéries Gram-positif. Ceux-ci ont été effectivement observés en microscopie électronique à transmission par coloration négative (Fig.13). L'étape successive dans la caractérisation des pili a consisté en leur détection immunologique. Les pilines ont été produites de façon recombinante et des anticorps polyclonaux ont été générés contre chacune de ces sous-unités afin de les marquer spécifiquement. Deux techniques sont employées pour détecter la présence des pilines. La première consiste en l'extraction par la mutanolysine, ou autre muramidase, de la paroi bactérienne contenant les polymères de pili. Ces extraits sont analysés par électrophorèse sur gel de polyacrylamide et après transfert des protéines sur une membrane de nitrocellulose, les pilines sont détectées par immunomarquage. Le profil de détection des pilines consiste en une échelle de haut poids moléculaire, traduisant l'association covalente des pilines au sein du pilus. La seconde technique de détection des pilines consiste en un immunomarquage des pili en surface des bactéries couplée à l'observation par microscopie électronique. Ces méthodes d'investigation par immunomarquage des pili natifs, associées à la génétique microbienne, ont

permis de proposer des modèles d'assemblage en localisant les pilines majeures et accessoires et de définir le rôle des sortases dans ces processus.

Pour illustrer ces propos, le pilus SpaABC chez *C. diphtheriae* est pris en exemple car il a fait l'objet de la toute première étude publiée dans ce domaine (Ton-That 2003). L'immunomarquage des pilines SpaA, SpaB et SpaC montre qu'elles sont distribuées de façon distincte au sein du pilus (Fig.15). SpaA est majoritairement représentée tout au long de la fibre alors que SpaB et SpaC sont localisées de manière ponctuelle (Ton-That 2003). Les mêmes expériences effectuées sur des souches mutantes indiquent que la délétion de *spaA* abolit la formation du pilus contrairement aux délétions de *spaB* ou *spaC* (Ton-That 2003). En conclusion, Ton-That et son équipe proposent que la structure fibrillaire du pilus SpaABC chez *C. diphtheriae* est formé par la polymérisation de la piline majeure SpaA et que les pilines accessoires SpaB et SpaC décorant la fibre, ne sont pas essentielles à l'assemblage du polymère SpaA.



**Figure 15: Immunolocalisation des trois pilines qui composent le pilus SpaABC de *C. diphtheriae*.** A. Immuno-localisation de la piline SpaA. B. Immuno-localisation de la piline SpaB. C. Immunolocalisation de la piline SpaC. Chacun des échantillons est observé par MET après avoir été marqué par les antisera spécifiques eux-même couplés à des billes d'or de 12nm. L'échelle représente une distance de 0.2 $\mu$ m. (Ton-That et Schneewind 2003)

### III.4.1 Pilines majeures

Le rôle du motif LPxTG dans l'association covalente des pilines majeures pour former la fibre a été attestée pour la première fois chez *C. diphtheriae* : la mutation du motif LPxTG de SpaA abolit la polymérisation de SpaA (Ton-That 2003).

L'étude exhaustive des différents pili exprimés chez *C. diphtheriae*, SpaABC, SpaDEF et SpaGHI permet d'identifier SpaD et SpaH comme les homologues de SpaA (Ton-That 2003). La comparaison des séquences de ces trois pilines majeures révèle en plus du motif LPxTG, la présence d'un motif conservé, YPKN qualifié par Ton-That de « motif piline » dont la lysine est capitale dans la formation de la fibre SpaA, sa mutation en alanine ou en



arginine inhibe toute polymérisation (Ton-That 2003). Lorsque les séquences YPKN et LPxTG sont intégrées dans la séquence d'une protéine sécrétée de *S. aureus* exprimée chez *C. diphtheriae*, la polymérisation de cette protéine a lieu, indiquant ainsi la fonction essentielle de ces éléments conservés dans la polymérisation covalente catalysée par les sortases pili-spécifiques (Ton-That 2003). Le motif piline n'étant pas retrouvé dans la séquence des pilines accessoires de *C. diphtheriae*, celui-ci devient un élément d'identification non ambigu des pilines majeures.

En 2004, des travaux effectués par la même équipe permettent l'identification d'un second motif conservé retrouvé dans la plupart des pilines majeures : YxLxE<sub>1</sub>TxAPxGY (Ton-That 2004). Du fait du haut niveau de conservation du glutamate dans cette séquence, ce motif est appelé E-Box. De façon intéressante, la mutation de ce glutamate n'a aucun effet sur la polymérisation de SpaA. En revanche ce résidu s'avère indispensable à l'incorporation de SpaB dans le polymère de SpaA et est partiellement nécessaire à l'ancrage de SpaC.

En conclusion, l'identification de ces motifs conservés permet de qualifier les pilines majeures (Fig.16).

	« Motif piline »	« E-Box »	« Motif LPxTG »
SpaA	187-VYPKH-191	442-FCLVETATASGY-453	491-LPLTG-495
SpaD	176-AYPKN-180	402-FCLKETKAPAGY-413	458-LPMTG-462
SpaH	199-VFPKN-203	472-YVLVETEAPTGF-483	520-LPLTG-524
FimP	179-VYPKN-183	446-YCLVETKAPEGY-457	494-LPLTG-498
FimA	232-VYPKN-236	483-YVLVETKAPAGY-494	530-LPLTG-534
BcpA	159-MYPKN-163	468-YQIFETKAPT-Y-478	519-LPVTG-523
EpbC	183-IYPKN-187	544-YYLEETVAPDDY-555	591-LPSTG-595
Gbs80	196-IYPKN-200	467-YKLKETKAPEGY-478	521-IPNTG-525
RrgB	180-VYPKN-184	573-YYLEETKQFAGY-584	628-IPQTG-632
PitB	200-NAPKV-204	324-YTVTETGV-AGY-334	380-VTPTG-384
Consensus	VYPKN	YxLxE <sub>1</sub> TxAPxGY	LPxTG

**Figure 16 : Alignements locaux présentant les motifs conservés trouvés dans les séquences des pilines majeures utilisées pour cette étude.** Alignements réalisés avec le logiciel Kalign (2.0) disponible sur le site EMBL-EBI (<http://www.ebi.ac.uk/Tools/kalign/index.html>). Les résidus hautement conservés sont surlignés en turquoise. Les résidus essentiels sont indiqués en rouge. Les numéros d'accèsion Uniprot des pilines majeures sont les suivants : SpaA (Q6NF81) SpaD (Q6NK05) SpaH (Q6NEP2) de *C. diphtheriae* NCTC13129 ; FimP (ANA\_2510) FimA (ANA\_0024) de *A. naeslundii* MG-1 ; BcpA (Q81D71) de *B. cereus* ATCC14579 ; EpbC (Q836L8) de *E. faecalis* V538 ; Gbs80 (Q8EOS9) de *S. agalactiae* 2603V/R ; RrgB (Q97SC2) de *S. pneumoniae* TIGR4.

A l'exception de l'opéron de *S. pyogenes* où ni le motif piline et ni le motif E-Box n'ont été identifiés par alignement de séquence, ces éléments sont présents chez toutes les autres bactéries. Dans PitB, la piline majeure du PI-2 de *S. pneumoniae*, ces motifs sont divergents de la séquence consensus de référence (Fig.16) (Bagnoli 2008).

Dans les alignements de séquences initialement réalisés par Ton-That, il est proposé que EbpC soit la seule piline majeure chez *E. faecalis* (Ton-That 2004). Or, l'étude ultérieure des trois pilines EbpA, EbpB et EbpC met en évidence que chacune d'elles présente les motifs spécifiques des pilines majeures (Nallapareddy 2006). Etant donné le manque d'études biochimiques supplémentaires, il est impossible de trancher sur la véritable identité de la piline majeure dans l'opéron *ebp*.

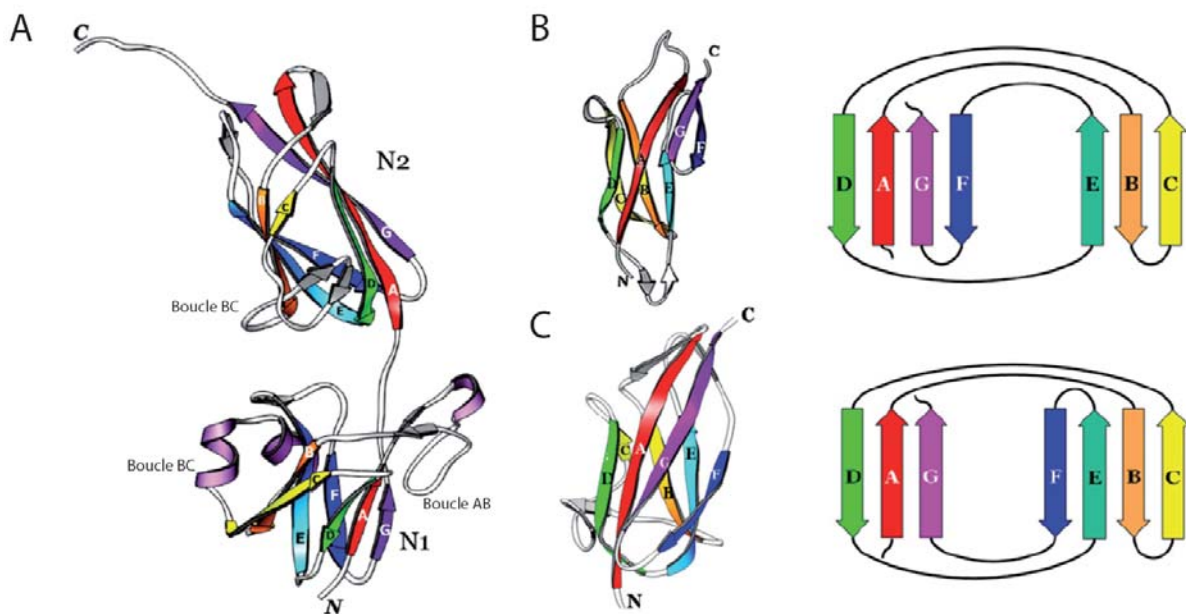
### III.4.2 Pilines accessoires

Comme indiqué précédemment les pilines dites mineures ou accessoires ne présentent pas de motif piline dans leur séquence (sauf exception : *E. faecalis*) et leur délétion n'empêche pas la formation de la fibre (Ton-That 2003, Budzik 2007, Mora 2005, Mishra 2007, Dramsi 2006, LeMieux 2006, Bagnoli 2008). La localisation des pilines accessoires n'est pas aussi clairement définie que celle des pilines majeures, on les retrouve soit décorant le pilus sur toute sa longueur, soit présentes à chacune des extrémités. Certaines d'entre elles ont pour fonction d'ancrer le pilus à la surface du peptidoglycane ou de signaler la terminaison de l'élongation. La fonction d'adhésine associée à une des pilines accessoires est par contre assez bien décrite chez plusieurs espèces bactériennes.

Dans la souche NEM316 de *S. agalactiae*, la piline accessoire PilA (l'orthologue de Gbs104 dans la souche 2603V/R) est responsable de l'adhésion de la bactérie à des cellules épithéliales alvéolaires humaines *in vitro* et cela indépendamment de la formation de la fibre (Dramsi 2006). L'analyse bioinformatique de la séquence de PilA indique l'existence d'un domaine adhésif von Willebrand de type A (VWA) (Konto-Ghiorghi 2009). Il est bien établi que les domaines VWA dans les protéines eucaryotes jouent un rôle important et très répandu dans les interactions protéine-protéine via un site d'adhésion ion-dépendant (MIDAS pour metal ion-dependant adhesion site) : ce sont notamment des ions  $Mn^{2+}$  et  $Mg^{2+}$  qui stabilisent le domaine I de l'intégrine  $\alpha 2\beta 1$  (Gotwals 1999). Cette fonction est retrouvée lorsque le domaine VWA est présent dans des protéines bactériennes : une souche mutante  $PilA\Delta VWA$  montre une diminution d'adhérence à des cellules épithéliales pulmonaires (A549) ou intestinales (TC7) en comparaison avec la souche sauvage (Konto-Ghiorghi 2009).

Chez la souche *S. agalactiae* 2603V/R, les deux pilines accessoires, Gbs104 et Gbs52 (orthologues respectifs de PilA et PilC dans la souche NEM316) possèdent des fonctions d'adhésine car des souches mutantes  $\Delta gbs104$  ou  $\Delta gbs52$  présentent une adhérence réduite à des cellules pulmonaires épithéliales par rapport à la souche sauvage (Krishnan 2007). Le site de liaison aux cellules eucaryotes dans Gbs52 a été identifié et se localise dans le domaine

N2. Cette analyse moléculaire a pu être menée grâce à la résolution de sa structure cristallographique (Krishnan 2007). Gbs52 se compose de deux domaines, N1 et N2, qui présentent tout deux une structure de type immunoglobuline inversé (Fig.17). Bien que leur topologie en sandwich  $\beta$  soit proche, il existe quelques différences structurales. En effet, les feuillet  $\beta$  du domaine N1 sont plus courts que ceux du domaine N2 et sont décorés de deux hélices AB et BC. De plus, le domaine N2 présente une longue queue C-terminale riche en proline absente dans le domaine N1 (Fig. 17, A). Outre ces différences, les domaines N1 et N2 se composent de 7 brins  $\beta$  antiparallèles organisés en deux feuillet dont l'organisation est similaire à la topologie des domaines de répétition CnaB, présents dans plusieurs pilines de bactérie Gram-positif et identifiés initialement dans l'adhésine Cna (pour collagen-binding protein) de *S. aureus* (Deivanayagam 2000) (Fig.17, B et C). La protéine Cna est responsable de l'adhérence de *S. aureus* au collagène, elle se compose d'une région A qui lie le collagène et de régions B, répétées, appelées CnaB, dont le rôle est de présenter la région A de façon optimale pour favoriser l'adhésion aux tissus hôtes (Deivanayagam 2000).



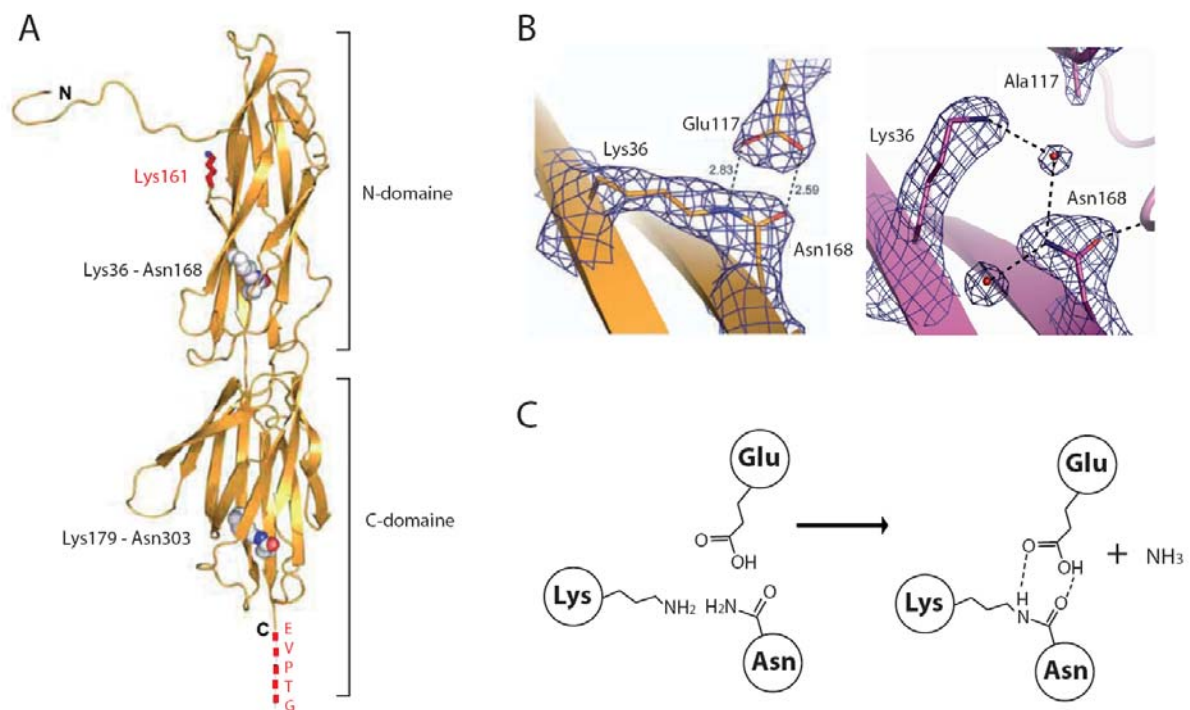
**Figure 17: Le domaine N2 de la piline accessoire Gbs52 de *S. agalactiae* présente une topologie de CnaB.**  
 A. Structure cristallographique complète de Gbs52. B. Structure et topologie du domaine CnaB de la protéine Cna de *S. aureus*. Le code couleur utilisé pour la structure (à gauche) permet de visualiser l'organisation schématique des feuillet  $\beta$  (à droite). C. Structure et topologie du domaine N2 de la protéine Gbs52. Le code couleur est identique à celui utilisé pour le CnaB de *S.aureus*. (Krishnan *et al.* 2007)

Chez *C. diphtheriae*, les pilines mineures semblent également impliquées dans l'adhérence des bactéries aux cellules humaines. En effet, dans le cas du pilus spaABC, la délétion individuelle des pilines mineures SpaB et SpaC, ainsi que leur double délétion, diminue jusqu'à 90% la fixation des bactéries à des cellules pharyngées humaines (Mandlik

2007). Le rôle de SpaB dans le contrôle de la longueur du polymère SpaA et dans l'ancrage de celui-ci à la paroi bactérienne a été proposé sur la base des phénotypes observés dans la souche  $\Delta spaB$  : les fibres de pili sont plus longues que dans la souche parentale et sont libérées dans le milieu de culture (Mandlik 2008).

### III.4.3 Des ponts intramoléculeaires covalents au sein des pilines

La résolution de la structure cristallographique de la piline majeure de *S. pyogenes* Spy128, a mis en évidence une modification post-traductionnelle particulière : la formation d'une liaison covalente entre les chaînes latérales de résidus Lys et Asn (Kang 2007) (Fig.18).



**Figure 18: Les ponts intramoléculeaires de la piline majeure Spy128 de *S. pyogenes*.** **A.** Structure complète et topologie de Spy128 présentant les ponts intramoléculeaires. La structure de Spy128 est représentée en ruban et les résidus impliqués dans les deux ponts intramoléculeaires sont représentés par des sphères. La Lysine équivalente à celle trouvée dans les motifs pilines est indiquée en rouge ainsi que le motif de type LPxTG. **B.** Zoom sur le pont intramoléculeaire du N-domaine, à gauche chez Spy128, à droite chez le mutant Spy128 E117A (PDB 3B2M et 3GLD respectivement). Les résidus impliqués dans les ponts sont représentés en bâton et boules. **C.** Réaction chimique permettant la formation du pont intramoléculeaire (Kang *et al.* 2007, Kang *et al.* 2009).

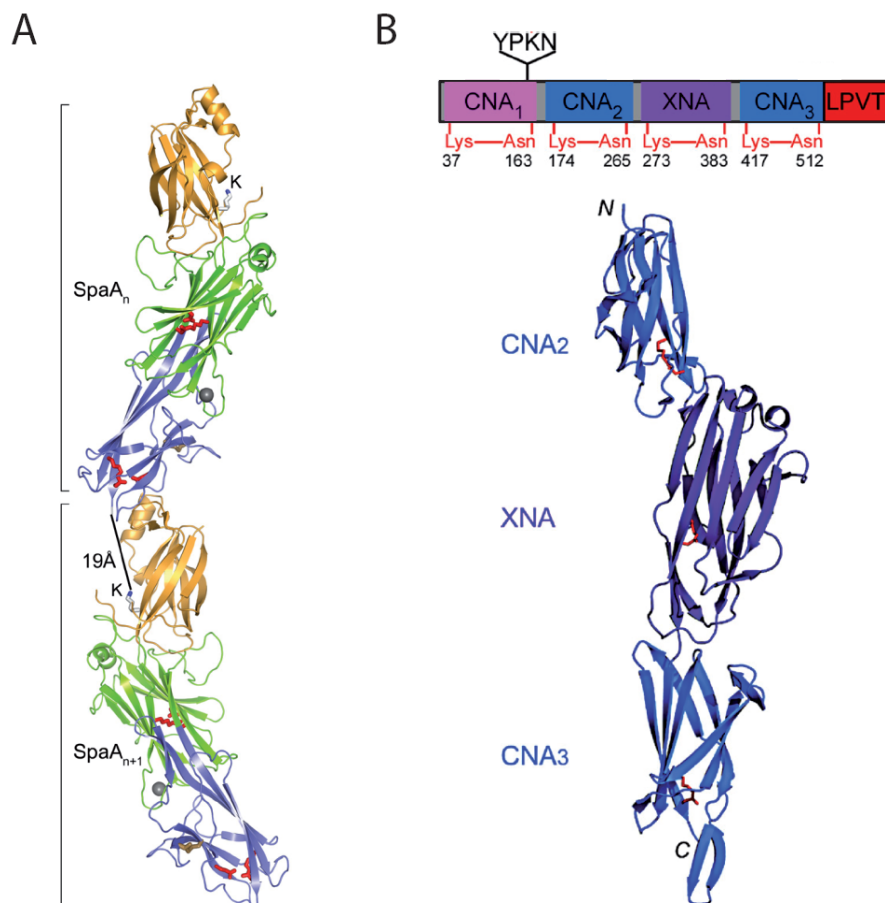
La piline Spy128 est une molécule allongée d'une longueur de 98Å formée par deux domaines, N et C qui présentent des structures de type immunoglobuline inversé (Fig.18, A). Au sein de chacun de ces domaines est identifié un pont covalent entre les résidus Lys36-Asn168 et Lys179-Asn303 (Fig.18) (Kang 2007). Ces résidus sont entourés par des acides aminés aromatiques ce qui modifierait le pKa local et permettrait à ces ponts de

s'autocatalyser (Wikoff 2000). Les glutamates 117 et 258 stabilisent chacune de ces liaisons en établissant des ponts hydrogènes avec les groupements NH et C=O (Fig.18, C). La formation du pont covalent Lys-Asn conduit à la perte d'un ion ammonium, soit 17 Da. La mesure de la masse de Spy128 par spectrométrie de masse est cohérente avec la perte de deux ions ammonium, confirmant la présence des deux ponts intramoléculaires dans la protéine en solution. Le rôle de chacun des résidus impliqués dans la formation des ponts a été vérifié par mutation des résidus Lys, Asn et Glu en Ala : la masse des protéines mutées démontre l'absence de formation des ponts intramoléculaires au sein de chacune d'elles (Kang 2007). De plus, la résolution de la structure cristallographique de la piline mutée Spy128 E117A confirme que la mutation du résidu glutamate abolit la liaison Lys-Asn et mène à un changement conformationnel des résidus de leur proche environnement (Kang 2007) (Fig.18, B). Ces ponts intramoléculaires covalents jouent un rôle essentiel dans la stabilisation de la piline Spy128. En effet, une diminution de stabilité, mesurée par dichroïsme circulaire et par sensibilité à la digestion tryptique est observée pour les protéines mutées E117A et E258A (Kang 2007, Kang 2009). En 1946, Rebecca C. Lancefield extrait et caractérise des antigènes, dits antigènes T, de *S. pyogenes* particulièrement résistants à la digestion protéolytique (Lancefield 1946). Il faut attendre près de 60 ans pour comprendre la raison d'une telle stabilité : les antigènes T sont les pili, polymères dont les sous-unités sont associées de façon covalente, et dont chaque sous-unité est elle-même stabilisée par une liaison interne covalente.

Des liaisons intramoléculaires covalentes de même type sont présentes dans les structures cristallographiques d'autres pilines, l'adhésine Cna de *S. aureus* et le domaine N2 de la piline Gbs52 de *S. agalactiae*. Or les auteurs respectifs de ces travaux ne les ont pas eux-mêmes identifiés, c'est Kang et ses collaborateurs, qui, ayant observés la présence des ponts dans Spy128 ont recherché des structures analogues dans les bases de données (PDB : Protein Data Bank) (Kang 2007, Deivanayagam 2000, Krishnan 2007). L'alignement des séquences de pilines majeures met en évidence une forte conservation des résidus Lys, Asn et Glu impliqués dans les ponts covalents, ce qui conduit Kang à proposer que ces ponts intramoléculaires sont caractéristiques des pilines de bactéries Gram-positif et peuvent être comparés à la fonction de stabilisation des ponts disulfures.

Dernièrement, deux structures cristallographiques de pilines majeures SpaA et BcpA trouvées respectivement chez *C. diphtheriae* NCTC13129 et *B. cereus* ATCC14579 ont été résolues (Kang 2009, Budzik 2009) (Fig.19). La piline majeure SpaA est une molécule allongée formée de trois domaines, représentés en doré, vert et bleu sur la figure 19A. Le

domaine central présente un pont intramolécule impliquant les résidus Lys199 et Asn 321, alors que le domaine C-terminal (en bleu) présente un pont intramolécule (Lys363-Asn482) ainsi qu'un autre élément stabilisateur : un pont disulfure (Fig.19A) (Kang 2009). Pour la piline BcpA, seule la structure cristallographique d'une forme tronquée de la protéine a été obtenue (Fig.19B). Celle-ci présente trois domaines nommés CNA2, XNA et CNA3 qui présente chacun un pont intramolécule. De façon intéressante, l'étude d'une forme recombinante de BcpA complète, ne permet d'identifier que les trois ponts intramolécules présents dans les domaines CNA2, XNA et CNA3. En revanche, lorsque BcpA est polymérisée un dernier pont intramolécule est identifiable dans le domaine CNA1, en plus du pont intermolécule impliquant la Thréonine du motif LPVT d'un monomère de BcpA et la Lysine du motif piline, YPKN d'un second monomère. Ces résultats suggèrent donc que contrairement aux trois ponts intramolécules des domaines CNA2, XNA et CNA3, celui du domaine CNA1 n'est formé que lorsque la piline est polymérisée, augmentant ainsi d'autant la stabilité de la fibre générée (Budzik 2009).



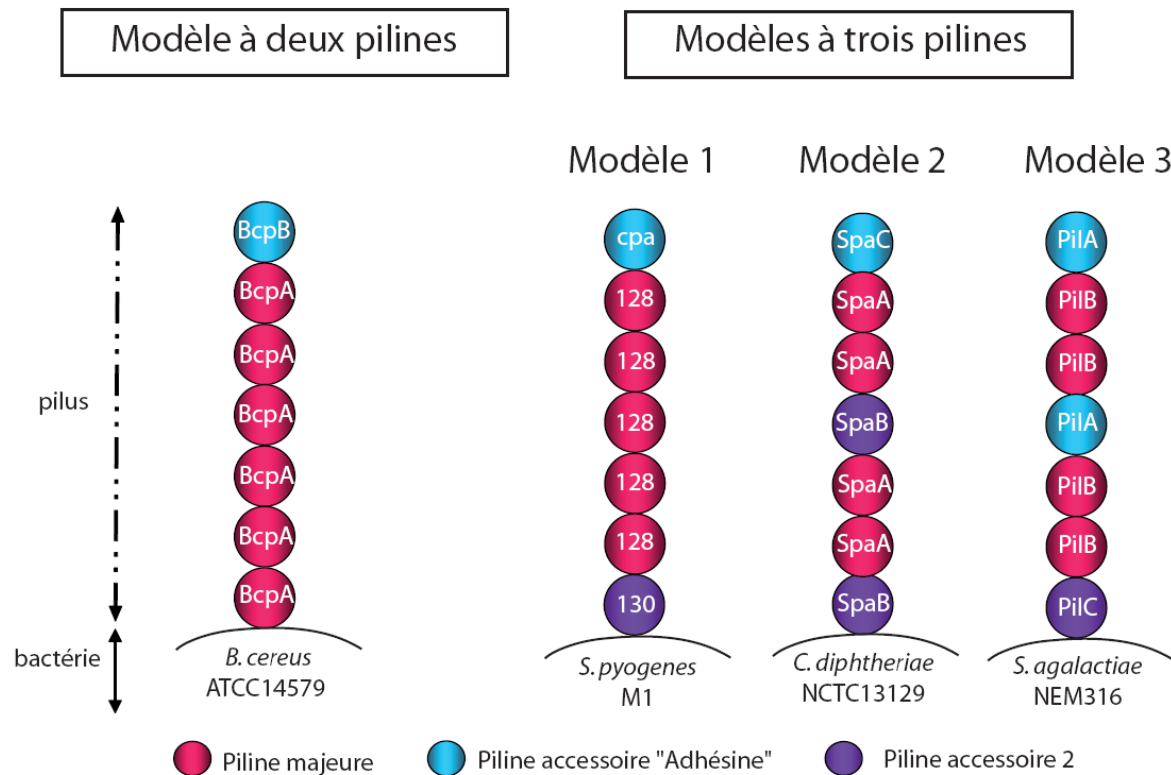
**Figure 19: Identification de ponts intramolécules chez les pilines majeures SpaA et BcpA.** A. Structure cristallographique de SpaA, la piline majeure du pilus SpaABC de *C. diphtheriae*. Le pont disulfure est représenté en jaune (Kang *et al.* 2009). B. Topologie globale et structure cristallographique partielle de BcpA, la piline majeure de *B. cereus*. (Budzik *et al.* 2009). Dans les deux cas, les ponts intramolécules sont indiqués en rouge.

### III.4.4 Organisation structurale des pili

Le nombre de pilines accessoires présentes dans les îlots de pathogénéicité chez chaque espèce bactérienne est variable. Cette donnée influence par conséquent la structure du pilus et son mode d'assemblage.

Chez les bactéries qui ne présentent qu'une seule piline accessoire, comme par exemple chez *B. cereus* et *A. naeslundii*, un modèle unique d'organisation du pilus est proposé. L'analyse par microscopie électronique à transmission couplée à l'immunomarquage indique que la polymérisation de la piline majeure forme le squelette du pilus, alors que la piline accessoire semble localisée à l'extrémité distale de cette structure (Budzik 2009, Mishra 2007) (Fig.20, Modèle à deux pilines). Chez *B. cereus*, la fibre résulte de la polymérisation de BcpA et la piline accessoire BcpB y est associée. Les motifs piline et LPVTG de BcpA sont indispensables à l'étape de polymérisation (Budzik 2007, Budzik 2008) et le motif IPNTG de BcpB permet son incorporation à la fibre BcpA (Budzik 2009). Chez *A. naeslundii*, le squelette du pilus est composé de FimA pour le pilus I et FimP pour le pilus II avec respectivement à leur extrémité les adhésines FimB et FimQ (Mishra 2007, Chen 2007).

Chez les bactéries qui présentent deux pilines accessoires, les modèles d'assemblage sont plus complexes et diffèrent selon l'espèce bactérienne. Le modèle le plus simple est celui proposé chez *S. pyogenes* où la piline majeure Spy128 forme le squelette du pilus et les deux pilines accessoires Spy130 et Cpa sont localisées respectivement, à la base du pilus et à l'extrémité distale, ce qui est en accord avec la fonction d'adhésine de Cpa (Linke 2010) (Fig.20, Modèles à trois pilines - Modèle 1). Dans le cas du pilus SpaABC de *C. diphtheriae*, un modèle de l'organisation des pilines a été proposé en 2008 (Fig.20, Modèles à trois pilines - Modèle 2). La polymérisation de SpaA construit le squelette du pilus, l'adhésine SpaC est localisée uniquement à l'extrémité distale, alors que la seconde piline accessoire SpaB peut être incorporée dans la fibre de SpaA et/ou à la base du pilus : dans ce cas, SpaB interviendrait comme signal de fin d'élongation et d'ancrage de la fibre à la paroi (Mandlik 2008). Le dernier modèle à trois pilines est celui proposé chez la souche NEM316 de *S. agalactiae* Fig.20, Modèles à trois pilines - Modèle 3). L'adhésine PilA, outre sa localisation à l'extrémité distale du pilus, serait également incorporée dans la fibre de PilB alors que PilC est essentiellement positionné à la base du pilus (Nobbs 2008, Konto-Ghiorghi 2009).



**Figure 20: Modèles de l'organisation générale des pili chez les bactéries Gram-positif.** Le modèle à deux pilines est issu des données recueillies chez *B. cereus* (Budzik *et al.* 2009) et chez *A. naeslundii* (Mishra *et al.* 2007). Les modèles à trois pilines proviennent des études menées sur le pilus SpaABC de *C. diphtheriae* (Mandlik *et al.* 2008), les pili de *S. pyogenes* (Linke *et al.* 2010) et les pili de *S. agalactiae* (Nobbs *et al.* 2008 ; Konto-Ghiorgi *et al.* 2009).

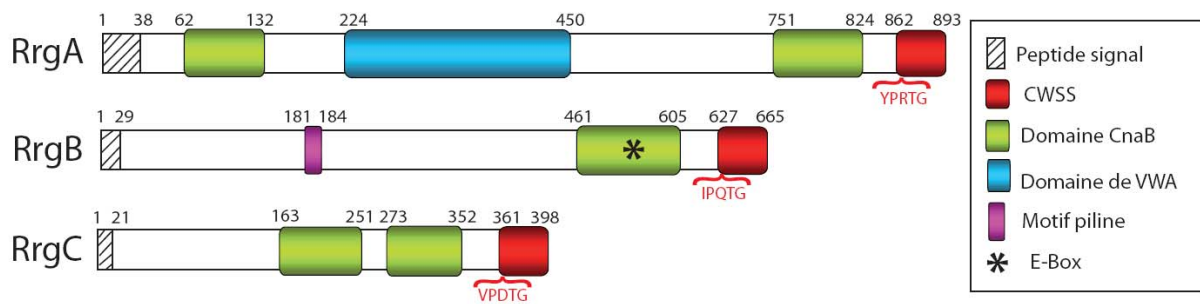
### III.5 Structure et assemblage du pilus chez *S. pneumoniae*

#### III.5.1 Les pilines du PI-1

Trois pilines sont présentes dans l'opéron PI-1 de *S. pneumoniae* TIGR4 : une piline majeure RrgB et deux pilines accessoires, RrgA et RrgC. Contrairement à la plupart des pilines identifiées jusqu'à lors, les pilines du pneumocoque présentent toutes des pentapeptides non conventionnels, à savoir YPRTG pour RrgA, IPQTG pour RrgB et VPDTG pour RrgC (Fig.21).

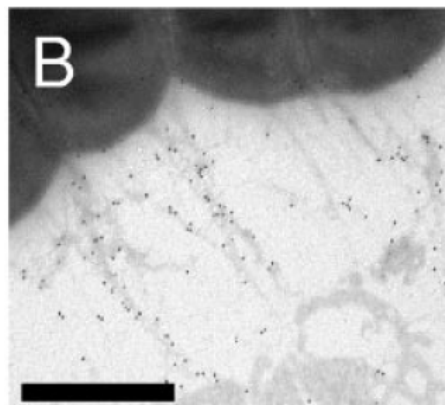
RrgB est la piline majeure. C'est une protéine de 665 résidus qui possède un motif piline YPKN (Lys183) et une E-Box, YYLEETKPAY (Glu577). Une analyse bioinformatique prédit la présence d'un domaine CnaB à l'extrémité C-terminale (Fig.21). L'observation des pili par microscopie électronique à transmission, couplée à l'immunomarquage avec des anticorps anti-RrgB, permet de localiser cette protéine tout au long de la fibre ce qui valide sa fonction de piline majeure (Fig.22) (LeMieux 2006).





**Figure 21: Représentation schématique de la topologie des protéines Rrg de *S. pneumoniae*.** Les domaines PFAM (CnaB et VonWillebrandt) correspondent aux prédictions obtenues à partir du site <http://smart.embl-heidelberg.de>. Le CWSS prend en compte le motif LPxTG, le domaine transmembranaire ainsi que la queue chargée positivement.

Les pilines RrgA et RrgC sont les pilines accessoires. RrgA porte essentiellement la fonction adhésive du pilus (Nelson 2007). Une souche mutante TIGR4 $\Delta$ rrgBC, qui n'exprime donc que RrgA, présente la même efficacité d'adhérence à des cellules A549 que la souche sauvage et la protéine recombinante purifiée RrgA démontre également des propriétés d'adhésion aux cellules humaines (Nelson 2007). L'analyse de la séquence de RrgA indique la présence d'un domaine de type VWA, déjà décrit comme impliqué dans la fonction d'adhésion de PilA chez *S. agalactiae*, et de deux domaines CnaB (Fig.21). La fonction de RrgC chez le pneumocoque n'est pas établie, cette protéine se compose de deux domaines CnaB consécutifs (Fig.21).

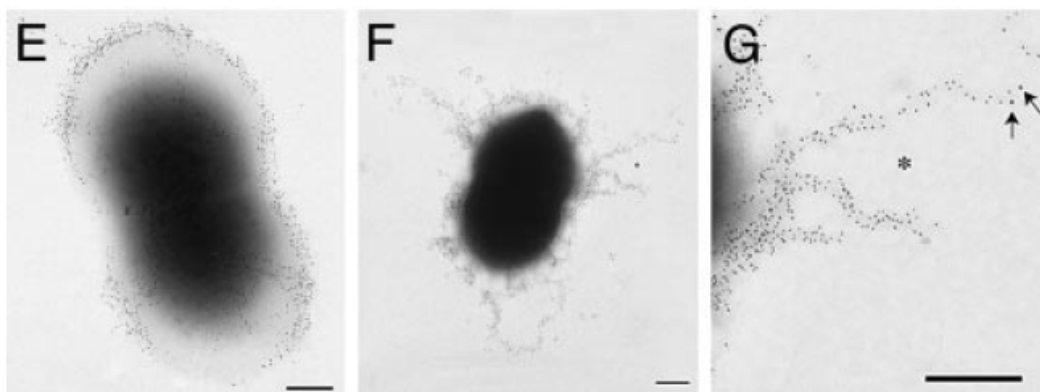


**Figure 22: Immuno-localisation de la piline majeure RrgB du PI-1 de *S. pneumoniae*.** Cliché de MET d'une bactérie *S. pneumoniae* TIGR4 immuno-marquée par des anticorps polyclonaux de souris  $\alpha$ -RrgB visualisés par des anticorps secondaires anti-souris couplés à des particules d'or de 18 nm. L'échelle représente une distance de 1 $\mu$ m (LeMieux *et al.* 2006)

### III.5.2 Modèles d'assemblage du pilus chez *S. pneumoniae*

Chez le pneumocoque, l'organisation générale du pilus reste controversée puisque quatre modèles ont été proposés depuis 2006, date de la première publication décrivant le pilus (Barocchi 2006). Les deux premiers modèles ont été conçus à partir de l'observation des pili en surface de la bactérie. Les deux suivants sont issus de l'analyse de pili natifs purifiés.

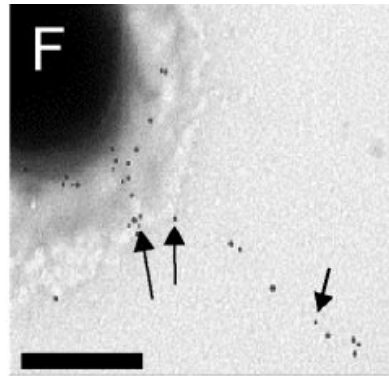
En 2006, Barocchi et ses collaborateurs proposent le premier modèle de cette structure à partir de clichés de microscopie électronique à transmission couplée à l'immunomarquage de chacune des pilines (Barocchi 2006). RrgA paraît localisée à la surface bactérienne suggérant que cette piline accessoire a une fonction d'ancrage du pilus. Les anticorps anti-RrgB décorent toute la longueur de la fibre, ce qui est en accord avec le rôle de piline majeure de la protéine. La piline mineure RrgC est localisée à l'extrémité distale du pilus (Fig.23 et Fig.27 modèle1).



**Figure 23: Immunolocalisation des pilines du PI-1 de *S. pneumoniae*.** E. Cliché de MET d'une bactérie *S. pneumoniae* TIGR4 immunomarkuée par des anticorps  $\alpha$ -RrgA visualisés à l'aide d'anticorps secondaires couplés à des particules d'or. F. Cliché de MET d'une bactérie *S. pneumoniae* TIGR4 immunomarkuée par des anticorps  $\alpha$ -RrgB (5 nm) et  $\alpha$ -RrgC (10 nm). G. Plus fort grossissement d'une bactérie *S. pneumoniae* TIGR4 immunomarkuée par des anticorps  $\alpha$ -RrgB (5 nm) et  $\alpha$ -RrgC (10 nm) (Les flèches indiquent le marquage anti-RrgC). L'échelle représente une distance de 200 nm (Barocchi *et al.* 2006)

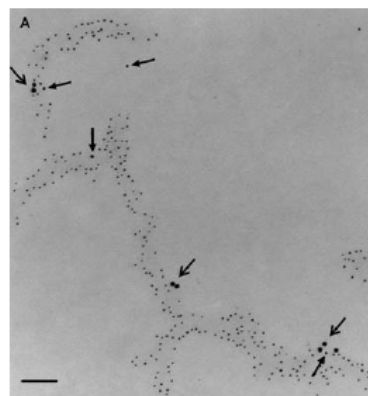
La même année, le groupe de Camilli réalise des expériences très similaires, mais présente des résultats divergents au modèle 1, notamment en ce qui concerne la localisation de RrgA (Fig.24 et Fig.27 modèle 2) (LeMieux 2006). Dans ce cas en effet, plusieurs molécules de RrgA sont regroupées, et ces organisations sont distribuées périodiquement le long de la fibre RrgB. De plus, sa localisation à la membrane n'est visualisée que chez un mutant TIGR4 $\Delta$ rrgBC, n'exprimant ni RrgB et ni RrgC (LeMieux 2006). Dans ce travail, la

localisation de RrgC n'a pas été étudiée et n'est donc pas incluse dans le modèle 2 présenté Figure 27.



**Figure 24: Immuno-localisation des pilines RrgA et RrgB du PI-1 de *S. pneumoniae*.** Cliché de MET d'une bactérie *S. pneumoniae* TIGR4 immunomarkée par des anticorps  $\alpha$ -RrgB révélés par des anticorps secondaires couplés à des billes d'or de 18 nm et par des anticorps  $\alpha$ -RrgA révélés par des anticorps secondaires couplés à des billes d'or de 12 nm. Les flèches indiquent le marquage RrgA. L'échelle représente une distance de 0.5 $\mu$ m (LeMieux *et al.* 2006)

En 2008, Hilleringmann et ses collaborateurs analysent l'assemblage des pilines à partir de pili extraits et purifiés (Hilleringmann 2008). L'observation de ces structures par microscopie électronique à transmission couplé à un triple immunomarquage montre l'association des fibres RrgB en protofilament de 3.5 nm de diamètre, sur lesquels sont associés des groupements de RrgA (Fig. 25 et Fig. 27, modèle 3). RrgC est également distribuée le long des fibres RrgB, associée ou non aux clusters de RrgA. Ce troisième modèle est en accord avec le modèle 2.

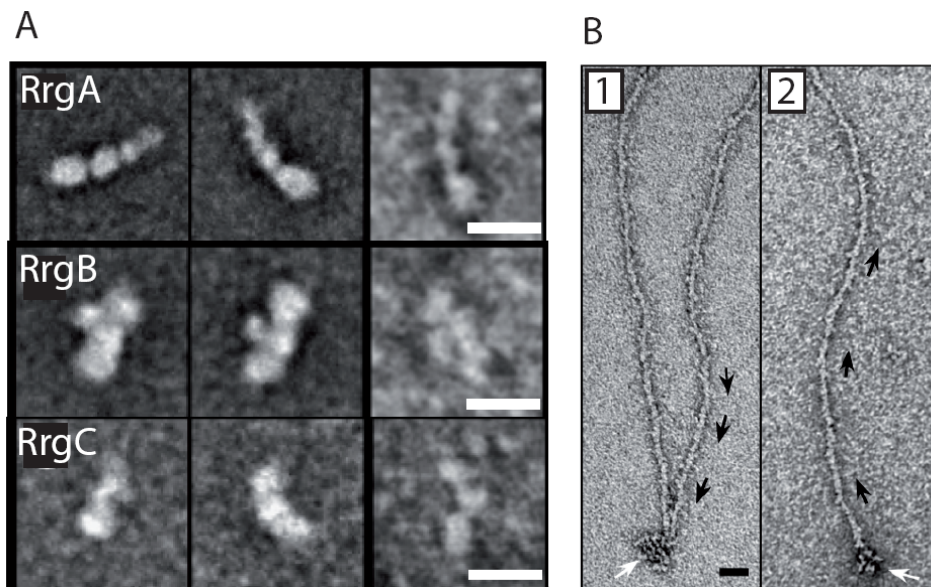


**Figure 25: Immuno-localisation des pilines RrgA, RrgB et RrgC sur des pili purifiés provenant de la souche *S. pneumoniae* TIGR4.** Les pili purifiés ont été incubés avec des anti-séras dirigés contre RrgA, RrgB et RrgC conjugués respectivement à des particules d'or de 15 nm, 5 nm et 10 nm. L'échelle représente une distance de 100 nm (Hilleringmann *et al.* 2008)

L'intérêt de cette approche réside dans la purification des pili et leur observation en microscopie électronique. L'analyse structurale de ces macromolécules montrent une super-organisation en « coiled-coil » des polymères de RrgB générant des fibres de diamètre

compris entre 7 et 9.5 nm, en accord avec les tailles des pili mesurés à la surface des bactéries (Barocchi 2006, LeMieux 2006, Hilleringmann 2008) (Fig. 27, modèle 3).

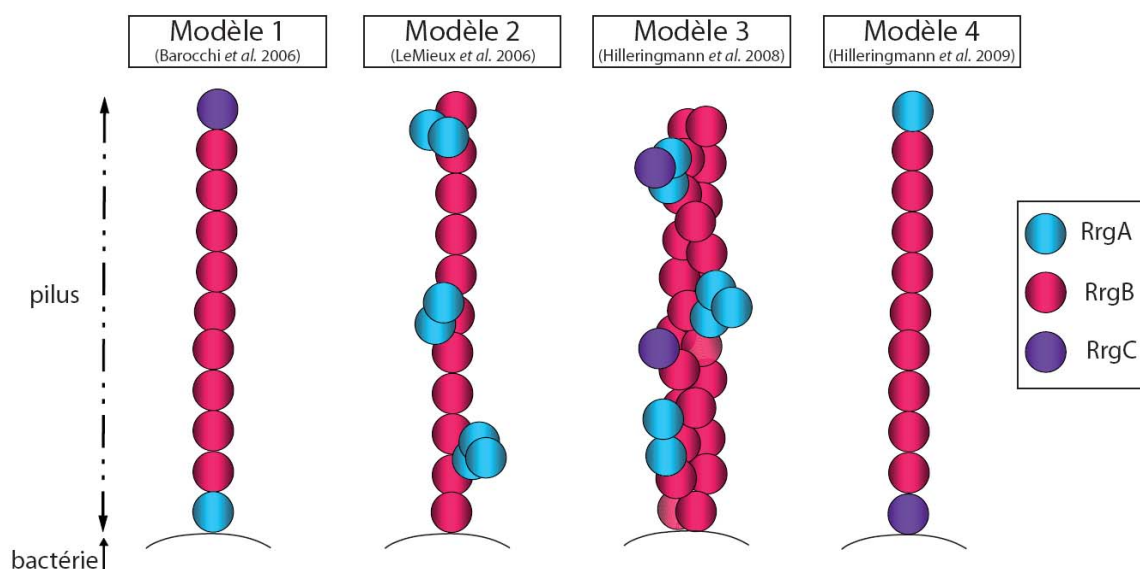
Enfin en 2009, la même équipe révisé son modèle en se basant sur une analyse des pili purifiés, et des sous-unités isolées, par STEM (Scanning Transmission Electron Microscopy) (Hilleringmann 2009) (Fig.26 et Fig.27, modèle 3).



**Figure 26: Immunolocalisation des pilines mineures RrgA et RrgC à partir de pili natifs purifiés de la souche TIGR4.** **A.** Analyse par ME en coloration négative des pilines recombinantes produites chez *E. coli*. La barre d'échelle représente une longueur de 10 nm. **B.** Immunolocalisation des pilines mineures RrgA (1) et RrgC (2). Les flèches noires permettent de visualiser la polarité des fibres grâce à la protrusion latérale des pilines RrgB. La flèche blanche indique les groupements d'anticorps marquant les pilines accessoires. La barre d'échelle représente une longueur de 20 nm (Hilleringmann *et al.* 2009).

L'association en super hélice des protofilaments observés lors de l'étude précédente est interprétée comme un artefact de la préparation, non homogène, des fibres de pili. L'optimisation des protocoles de préparation des échantillons bactériens, de purification des pili, et des méthodes de coloration négative ainsi que l'amélioration de la résolution en microscopie électronique permettent d'observer et d'analyser des pili individuels, non plus associés en « coiled-coil », aussi bien à la surface du pneumocoque qu'à partir de préparations de pili purifiés (Hilleringmann 2009). Des expériences d'immunomarquage avec des anticorps dirigés contre chacune des pilines sont réalisées sur ces fibres individuelles. La technique de microscopie électronique employée (STEM) permet d'observer directement les Ig : la haute résolution ainsi atteinte permet de montrer de façon non ambiguë que RrgB forme la fibre du pilus et que RrgA et RrgC sont chacune présentes à une extrémité de la fibre (Hilleringmann 2009) (Fig. 26, B). L'orientation de la fibre native est définie dans ces travaux par l'analyse

structurale des sous-unités recombinantes purifiées, en employant le même outil de microscopie électronique (Hilleringmann 2009) (Fig. 26, A). RrgA est formé d'un chapelet flexible de quatre petits domaines d'une longueur totale approximative de 18 nm. La protéine recombinante RrgB a la forme d'un haricot de 12 nm de long et présente une protrusion latérale. Enfin, RrgC la plus petite des trois pilines semble être composé de 2-3 domaines et présente une longueur approximative de 10 nm (Fig. 26, A). Ainsi, basé sur ces informations et sur l'analyse des pili natifs purifiés, Hilleringmann et ses collaborateurs proposent que la fibre du pilus est composée d'approximativement 150 monomères de RrgB qui se chevauchent légèrement et dont l'orientation est définie grâce à la protrusion latérale de la piline majeure. Cette observation permet également de montrer qu'un monomère de RrgA est localisé à l'extrémité distale de la fibre et un monomère de RrgC du côté proximal (Hilleringmann 2009) (Fig. 26, B). Ce quatrième modèle est plus en adéquation avec les données recueillies concernant les pilines accessoires. RrgA est l'adhésine du pilus, il paraît donc important qu'elle soit présente à l'extrémité distale de la fibre pour faciliter l'adhésion du pilus aux cellules hôtes, bien que sa localisation le long de la fibre puisse augmenter l'avidité d'une telle interaction. La position de RrgC à la base du pilus permet d'envisager une fonction dans l'ancrage du pilus à la paroi bactérienne, comme le suggère le phénotype de la souche mutante TIGR4 $\Delta$ rrgC, dans laquelle les pili se détachent plus facilement de la surface bactérienne et sont retrouvés en quantité plus importante dans le milieu de culture (Hilleringmann 2009).

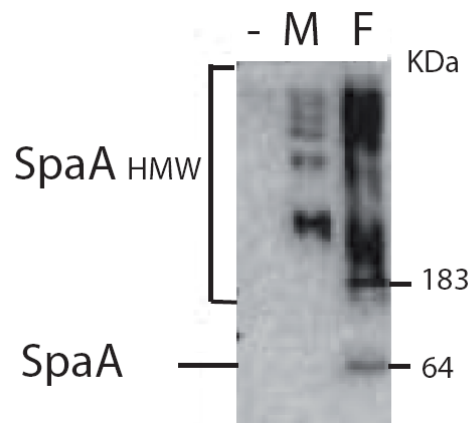


**Figure 27: Modèles de l'organisation générale des pili du pneumocoque.** Les modèles 1 et 2 sont basés sur des analyses par MET couplé à de l'immunomarquage des pili en surface de la souche *S. pneumoniae* TIGR4. Dans le modèle 2 la localisation de RrgC n'a pas été investiguée ce qui explique son absence. Les modèles 3 et 4 sont proposés à partir de l'analyse des pili natifs purifiés soit par MET couplé à de l'immunomarquage (modèle 3) soit par MET à balayage couplé à de l'immunomarquage (modèle 4).

### III.6 Rôle des sortases

#### II.6.1 *C. diphtheriae* : les premiers éléments de réponse

Les travaux pionniers de Ton-That en 2003 sur l'assemblage du pilus SpaABC de *C. diphtheriae* indiquent que les pili des bactéries Gram-positif sont liées à la paroi bactérienne et sont le résultat de l'assemblage covalent de différentes pilines. Les protocoles d'extraction de ces pili nécessitent l'usage d'agents dégradant le peptidoglycane, des enzymes (muramidase ou lysozyme) ou de l'acide formique. Ces extraits solubilisés de paroi bactérienne sont analysés par immunodétection après séparation sur SDS-PAGE et transfert sur membrane. Dans le cas du pilus SpaABC, de telles expériences réalisées avec un anticorps anti-SpaA permet de visualiser deux espèces différentes. Une première bande apparaît à une masse apparente proche de celle du monomère de SpaA. En revanche, une échelle de haut poids moléculaire (ou HMW pour High Molecular Weight) immunoréactive est également visible soit sous forme de bandes discrètes, dans le cas du traitement à la muramidase soit sous la forme d'une traînée après traitement à l'acide formique, montrant que les pili chez *C. diphtheriae* résultent de l'assemblage covalent de pilines SpaA, ce polymère étant lui-même associé de façon covalente à la paroi (Fig.28) (Ton-That 2003).



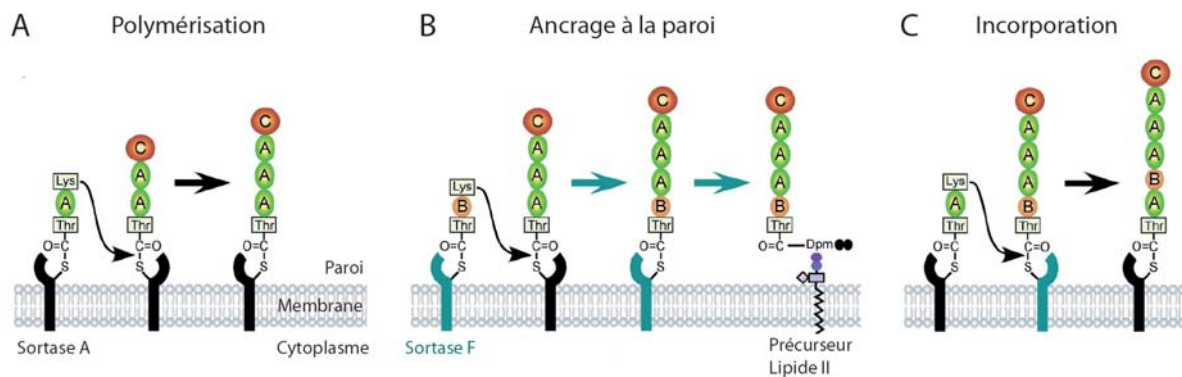
**Figure 28: Analyse du pilus SpaABC de *C. diphtheriae*.** Western Blot  $\alpha$ -SpaA des pili SpaABC traités (-) par simple chauffage en présence d'agent réducteur, (M) par de la muramidase, (F) par de l'acide formique (F) (Ton-That *et al.* 2003).

Les enzymes sortases, à l'origine de ces associations covalentes, reconnaissent comme substrats les motifs LPxTG des pilines et les transfèrent sur un groupement amine accepteur, catalysant ainsi une réaction de transpeptidation. Chez *C. diphtheriae*, les pilines SpaA, SpaB et SpaC sont codées en opéron avec la sortase A (qui est en réalité une sortase de classe C mais dont la nomenclature n'a pas été modifiée). La délétion de cette sortase conduit à

l'abolition des formes SpaA<sub>HMW</sub> suggérant le rôle de la sortase A dans la polymérisation de SpaA (Ton-That 2004). De plus, comme mentionné précédemment, la mutation du motif LPxTG de SpaA ou la mutation de la lysine du motif piline empêche la polymérisation de SpaA. Ainsi, basé sur le mécanisme de transpeptidation médié par les sortases ménagères, Ton-That propose que (i), la sortase reconnaît le LPxTG d'un monomère SpaA et clive la liaison Thr-Gly (ii), l'acyl-enzyme ainsi généré subit une attaque nucléophile par la lysine du motif piline d'une nouvelle sous-unité, ces étapes aboutissent à la formation d'une liaison peptidique entre deux pilines SpaA (Ton-That 2004). Cette hypothèse est validée en 2009 par l'équipe de Kang qui montre à partir de pili purifiés, digérés à la trypsine et analysés par spectrométrie de masse, que la liaison intermoléculaire entre deux pilines SpaA fait bien intervenir le résidu Thr du motif LPxTG et la lysine du motif piline (Kang 2009). Dans la souche NCTC13129 de *C. diphtheriae*, en plus du gène *srtA*, il existe 4 gènes codant pour des sortases de classe C et un gène codant pour la sortase ménagère SrtF. Afin de vérifier le rôle de chacun d'eux dans la polymérisation de SpaA, ces gènes ont été individuellement délétés et les souches mutantes analysées par western blot anti-SpaA. Aucune des délétions de sortases n'a d'effet sur la polymérisation de SpaA. Dans le cas du mutant  $\Delta srtF$ , la présence de SpaA<sub>HMW</sub> dans un échantillon simplement bouilli en présence de SDS traduit un défaut d'ancrage du polymère SpaA à la paroi. En conclusion, chez *C. diphtheriae*, la sortase ménagère SrtF semble jouer un rôle dans l'ancrage du pilus à la paroi, alors que la sortase A est nécessaire à l'assemblage du pilus. Des études ultérieures montrent que les sortases de classe C, trouvées dans les opérons SpaABC, SpaDEF et SpaGHI n'ont d'effet que sur les protéines LPxTG codées dans le même opéron (Ton-That 2003, Ton-That 2004, Ton-That 2005).

Qu'en est-il des deux pilines accessoires SpaB et SpaC ? En 2007, Swaminathan construit chez *C. diphtheriae*, une librairie de mutants n'exprimant qu'une seule des six sortases à la fois. L'analyse par western blot, avec des anticorps anti-SpaB et anti-SpaC, des pili produits par ces souches mutantes indique que seule la souche exprimant la SrtA permet la formation de SpaB<sub>HMW</sub> et SpaC<sub>HMW</sub> (Swaminathan 2007). L'année suivante, la piline mineure SpaB est identifiée comme étant le substrat spécifique de la sortase F, ancrant le pilus à la paroi (Mandlik 2008). Les données ainsi recueillies permettent de proposer un mécanisme d'assemblage du pilus SpaABC de *C. diphtheriae* dans lequel la sortase A joue un rôle prépondérant puisqu'elle est capable de polymériser SpaA et de lier SpaC à la fibre de SpaA. L'intervention de la sortase F permet à son tour la prise en charge de SpaB qui dans ce cas

précis peut-être soit simplement incorporé à la fibre, soit représenter le signal de fin d'élongation et d'ancrage à la paroi bactérienne (Fig.29).



**Figure 29: Modèle de l'assemblage du pilus SpaABC de *C. diphtheriae*.** A. représentation schématique de la polymérisation du pilus médié par SrtA (colorée en noir). Le résidu Lysine du motif piline et la Thréonine du LPxTG sont indiqués. B. Représentation schématique de l'ancrage du pilus polymérisé médié par la sortase ménagère SrtF (colorée en bleue). C. Modèle proposant comment SpaB peut se retrouver incorporé dans la fibre de SpaA. La piline majeure SpaA est colorée en verte, les pilines accessoires SpaB et SpaC sont colorées en orange (Mandlik *et al.* 2008)

### III.6.2 Pourquoi trois sortases chez *S.pneumoniae* ?

La souche TIGR4 de *S. pneumoniae* présente dans son génome quatre gènes codant pour des sortases : la sortase ménagère SrtA et trois sortases de classe C codées par le PI-1. De façon très intéressante, cet îlot de pathogénicité est le seul, identifié à ce jour, qui présente un nombre équivalent de gènes codant pour des sortases et de gènes codant pour les pilines. De cette observation, a émergé la proposition suivante : « une piline-une sortase », qui sous-entend que chaque sortase présenterait une spécificité stricte pour l'une des pilines. Mais après quelques années d'étude du PI-1, les mécanismes d'assemblage du pilus et les spécificités de substrats de chaque sortase paraissent bien plus complexes que ne le laissait supposer cette hypothèse un peu simpliste.

#### III.6.2.1 la polymérisation de RrgB

Contrairement aux autres bactéries Gram-positif, chez le pneumocoque, la sortase A n'est nécessaire ni à l'assemblage du pilus et ni à son ancrage à la paroi bactérienne. En effet, après extraction des pili par un traitement à base de mutanolysine (enzyme clivant les liaisons  $\beta$ 1-4 NAM-NAG) et de lysozyme, l'analyse par western blot anti-RrgB des pili d'une souche  $\Delta$ srtA indique que celle-ci est capable de produire RrgB<sub>HMW</sub> et que sa capacité à localiser les pili à la paroi est identique à celle d'une souche sauvage (LeMieux 2008). Les mêmes auteurs



montrent que chez une souche TIGR4 n'exprimant aucune sortase de classe C, aucune polymérisation de la piline majeure ne peut être détectée par western blot anti-RrgB, suggérant que l'une des sortases, au moins, est nécessaire à la polymérisation de RrgB (LeMieux 2008). Lorsque chacune des sortases est mutée individuellement, chaque souche mutante demeure capable de produire un polymère de RrgB et les pili produits sont tous localisés à la paroi bactérienne (LeMieux 2008). C'est la raison pour laquelle les auteurs concluent que les sortases C-1, C-2 et C-3 présentent une fonction redondante pour la polymérisation de RrgB (LeMieux 2008). L'inactivation de la SrtC-1 dans la souche TIGR4 réduit, dans nos conditions expérimentales, le taux de polymérisation de RrgB, et une accumulation de RrgB monomérique est détectable dans notre échantillon (Manzano 2008) (**Publication 1** Fig1.A). Dans la même étude, l'utilisation de protéines recombinantes solubles nous a permis de mettre en évidence que l'incubation à 37°C pendant 16h de RrgB avec la SrtC-1 dans un ratio 1 :2 conduit à la polymérisation de RrgB *in vitro*. Lorsque la sortase C-3 est utilisée dans les mêmes conditions, seul un début de fibre peut être visualisé par western blot anti-RrgB, indiquant que l'efficacité de catalyse de la SrtC-3 est moindre que celle de la SrtC-1. En revanche, aucune polymérisation de RrgB n'est détectée en présence de SrtC-2 (Manzano 2008). Ces résultats indiquent que dans nos conditions expérimentales la polymérisation de RrgB est principalement médiée par SrtC-1 et SrtC-3 peut en partie suppléer cette fonction contrairement à SrtC-2 qui en est incapable.

### III.6.2.2 l'incorporation des pilines accessoires

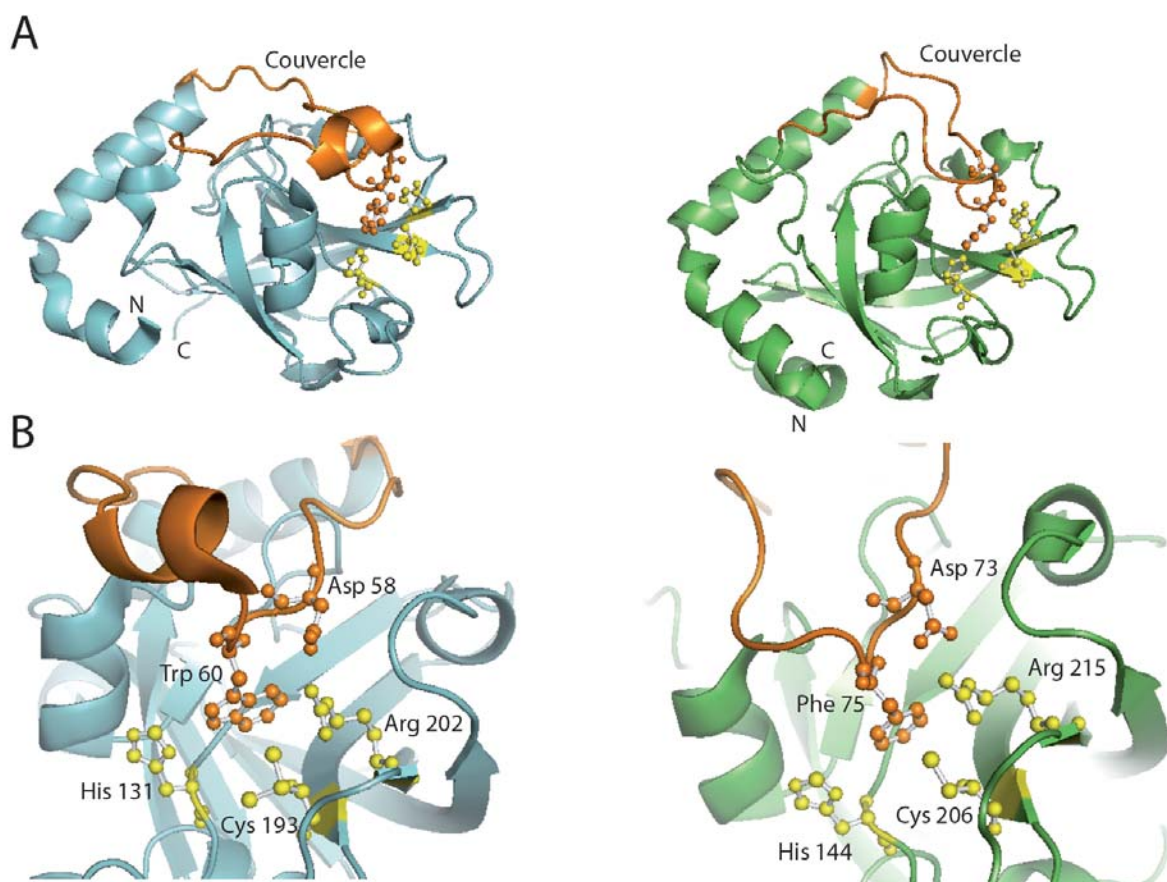
Selon les travaux de Fälker réalisés en 2008, les SrtC-1 et C-2 permettraient l'association de RrgA au pilus. En effet, parmi les souches de pneumocoque délétées des trois sortases de classe C (TIGR4 $\Delta$ C) et chez lesquelles chacune des sortases est individuellement surexprimée, seules les souches TIGR4 $\Delta$ C  $\nabla$ srtC-1 et TIGR4 $\Delta$ C  $\nabla$ srtC-2 sont capables d'incorporer RrgA dans les pili (Fälker 2008). La même année, l'équipe de Camilli montre également, mais par une approche de simple délétion, que RrgA peut être incorporé au pili par la SrtC-1 car chez un mutant de délétion *srtC-1*, aucune forme RrgA<sub>HMW</sub> n'est observée (LeMieux 2008). Par contre, ces auteurs indiquent que la SrtC-3 peut aussi intégrer RrgA dans le pilus, ce qui va à l'encontre des données obtenues par Fälker.

En ce qui concerne l'incorporation de la piline accessoire RrgC, les deux groupes présentent une nouvelle fois des résultats en contradiction. Dans un cas, seule la SrtC-1 est capable d'incorporer RrgC aux pili (Fälker 2008), dans l'autre cas, les deux sortases SrtC-1 et

C-3 sont capables d'accomplir cette fonction (LeMieux 2008). Une nouvelle fois la fonction de chacune des sortases dans l'incorporation des pilines accessoires reste controversée.

### III.6.2.3 les structures des sortases de classe C

En 2008, les structures cristallographiques de SrtC-1 et SrtC-3 ont été résolues par Clothilde Manzano à l'IBS dans l'équipe d'Andréa Dessen (Manzano 2008) (**Publication 1**). Bien que les SrtC-1 et C-3 ne présentent une identité de séquence que de 26.5%, ces deux protéines montrent une topologie similaire avec un cœur en tonneau  $\beta$  entouré d'hélices (Fig.30).



**Figure 30: Structures cristallographiques des sortases C-1 et C-3 de *S. pneumoniae*.** A. Structure globale de la sortase C-1 en bleu et de la sortase C-3 en vert. Les deux enzymes présentent une topologie similaire avec un tonneau  $\beta$  central, composé de 8 feuillets  $\beta$  anti-parallèles entouré d'hélices. Les résidus de la triade catalytique sont représentés en jaune et le couvercle qui masque l'entrée au site catalytique est représenté en orange. B. Zoom sur le site actif et sur les interactions entre les résidus de la triade catalytique et ceux du couvercle.

Ces deux sortases présentent chacune un site actif composé d'une triade catalytique (en jaune sur les structures). Pour la sortase C-1, les résidus catalytiques sont la cystéine 193, l'arginine 202 et l'histidine 131 alors que pour la sortase C-2 la triade catalytique est formée par les résidus cystéine 206, arginine 215 et histidine 144. Contrairement à la sortase ménagère SrtA

de *S. aureus*, l'accès au site actif de ces sortases est masqué par un couvercle (représenté en orange sur les structures) qui est maintenu en position fermée. En effet, certains acides aminés du couvercle interagissent soit directement avec des résidus catalytiques (Asp58 – Arg202 dans SrtC-1 ; Asp73 – Arg215 dans SrtC-2), soit sont positionnés juste en face (Trp60 – Cys193 dans SrtC-1 ; Phe75 – Cys206 dans SrtC-2) (Fig.30). Ces données suggèrent donc que l'accès au site actif va dépendre de l'ouverture du couvercle et que cette dernière intervient probablement suite à la reconnaissance spécifique du substrat.

La structure cristallographique de la troisième sortase, SrtC-3 a aussi été résolue (Neiers 2009). Cette dernière présente une identité de séquence de 58% avec SrtC-1 et de 27% avec SrtC-2, et une topologie similaire aux deux précédentes.

### III.7 Conclusion

L'adhésion des bactéries aux cellules hôtes est un processus essentiel dans les étapes initiales de colonisation des tissus. Ces mécanismes nécessitent le concours de facteurs de virulence divers et variés trouvés en surface de la bactérie.

Des structures fibrillaires nommées pili ont été récemment identifiées à la surface des bactéries Gram-positif et démontrés comme ayant un rôle dans les étapes initiales de colonisation des tissus hôtes. D'abord observés par microscopie électronique, l'avènement des séquençages complets de génomes bactériens a permis l'identification des éléments génétiques associés à ces structures. De façon intéressante, les gènes codant pour les sous-unités structurales des pili sont codés dans le même opéron que les éléments nécessaires à leur assemblage: des enzymes appelées sortases. En effet, les premières études moléculaires réalisées sur les pili des bactéries Gram-positif attestent que ces structures sont le résultat d'un assemblage covalent de différentes pilines catalysé par des sortases ce qui expliquerait leur caractère particulièrement résistant.

En fonction de la bactérie considérée le nombre de piline et de sortase est variable. Néanmoins, les études pionnières réalisées chez *C. diphtheriae* ont permis de déterminer l'existence de pilines dites majeurs et dont la polymérisation représente le squelette du pilus et de pilines dites accessoires qui décoorent la fibre et qui ont un rôle soit d'adhésine soit d'ancrage. Des éléments conservés ont été identifiés au sein des pilines majeurs : le motif piline et le motif E-Box. Alors que le premier élément semble indispensable à la polymérisation des pilines majeurs, le second n'a pas de rôle clairement attribué.

La résolution cristallographique de la piline majeure Spy128 de *S. pyogenes* a mis en évidence l'existence de ponts intramoléculaires au sein des pilines dont la fonction stabilisante a été prouvée. Des liaisons covalentes de même type sont présente chez d'autres pilines de bactéries Gram-positif ce qui semble en faire une de leur caractéristique commune. Quant est il des pilines du pneumocoque ? Présentent-elles des ponts intramoléculaires ?

Chez certaines bactéries, telle que *C. diphtheriae*, les mécanismes de biogenèse du pilus commencent à être élucidés et font l'unanimité. Chez le pneumocoque plusieurs modèles divergents de formation du pilus ont été proposés suite à des études de délétions génétiques. La résolution cristallographique des trois sortases de classe C chez *S. pneumoniae* indique la présence d'un couvercle qui bloque l'accès au site actif de l'enzyme. Ces informations suggèrent donc que l'activité enzymatique de ces sortases nécessite l'ouverture de ce couvercle et est probablement lié à une spécificité de reconnaissance du substrat. Néanmoins, aucune donnée biochimique permettant d'expliquer précisément ces mécanismes de reconnaissance au niveau moléculaire n'est encore disponible.



# **Matériels & Méthodes**



## MATERIELS & METHODES

### I. Souches, vecteurs et milieux de culture

#### I.1 Les souches

Trois souches bactériennes d' *Escherichia coli* ont été utilisées pour cette étude

- *E. coli* DH5 $\alpha$  (*supE44*,  $\Delta$ *lacU169*, ( $\phi$ 80  $\Delta$ *lacZ* $\Delta$ *M15*), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*)
- *E. coli* BL21 (DE3) RIL (*F- ompT hsdS(rB- mB-) dcm+ Tetr E. coli gal  $\lambda$  (DE3) endA Hte [argU ileY leuW Camr]*)
- La souche *E. coli* (DE3) STAR (*F- ompT hsdSB (rB-mB-) gal dcm rne131 (DE3)*)

#### I.2 Les vecteurs

##### Le Vecteur pLIM01

Le vecteur pLIM01 est dérivé du vecteur d'expression pLX02 de PX Therapeutics (Grenoble). Il présente une cassette de résistance à l'ampicilline et un site de clonage permettant de fusionner six résidus histidine du côté N-terminal de la protéine d'intérêt. Le site de clonage est précédé d'un promoteur T5, d'un opérateur lactose (non représenté sur la carte) ainsi que d'un RBS (Ribosome Binding Site). Ce plasmide porte le gène *lacIq* qui permet de réprimer la transcription du gène par fixation du répresseur *lacI* au niveau de l'opérateur lactose. En présence d'IPTG celle-ci fixe le répresseur, l'ADN polymérase T5 d'*E.coli* peut alors transcrire le gène d'intérêt.

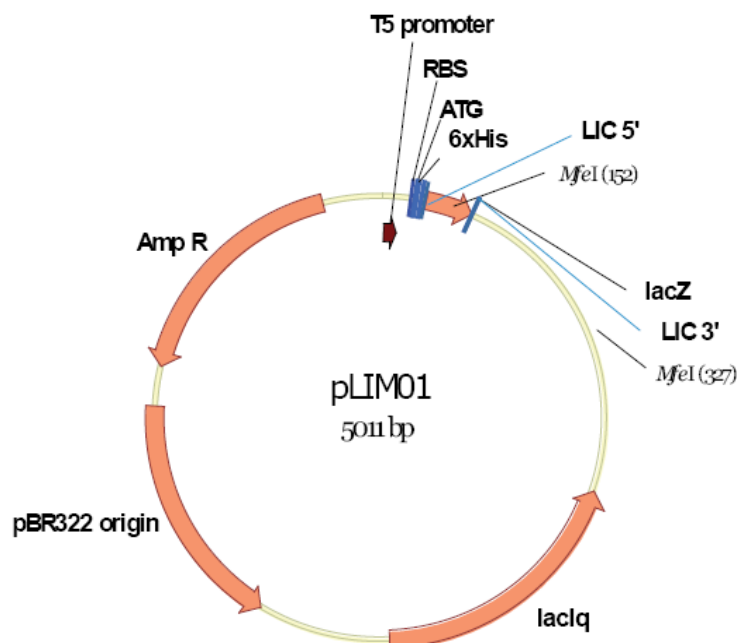


Figure 31: Organisation schématique du vecteur pLIM01



## Les vecteurs bi-cistronique pETDuet®, pACYCDuet® et pCDFDuet®

Les trois vecteurs présentent respectivement des résistances à l'ampicilline, au chloramphénicol et à la streptomycine et deux sites de clonage nommés MCS1 et MCS2 (pour Multiple Cloning Sites). Chacun d'eux est précédé d'un promoteur T7, d'un opérateur lac et d'un rbs. Pour chacun des MCS1, une séquence codant pour une étiquette poly-histidine peut-être utilisée en amont du gène à cloner. Pour chacun des MCS2, une séquence codant pour un S-Tag peut-être utilisée en aval du gène à cloner, celui-ci étant suivi d'un codon stop.

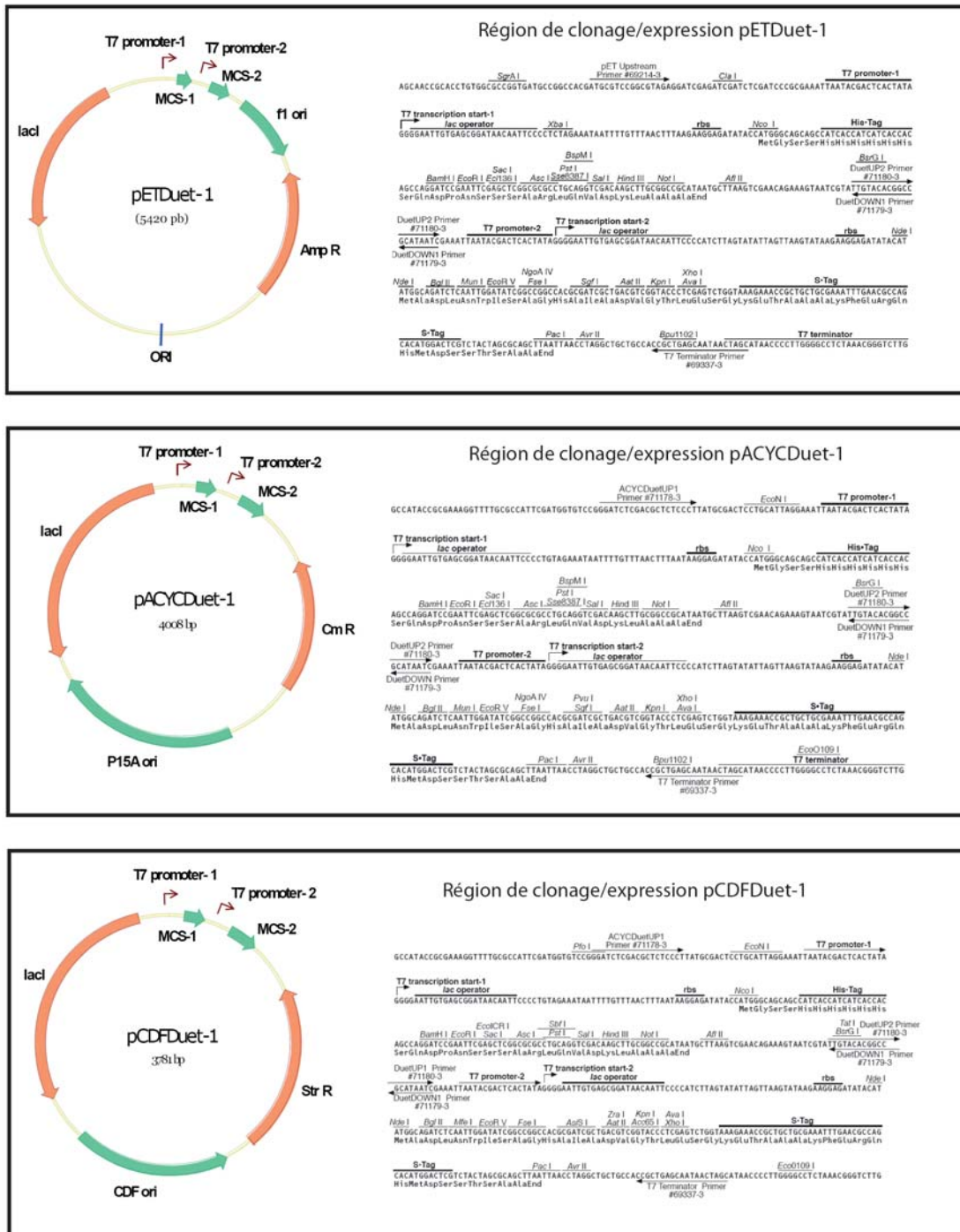


Figure 32: Organisation schématique des vecteurs pETDuet®, pACYCDuet® et pCDFDuet®.

### I.3 Les constructions

Deux types de constructions ont été utilisées lors de cette étude : Les constructions initiales ont été réalisées dans le vecteur pLIM01 (par Anne Marie Di Guilmi) et les constructions nécessaires au système de co-expression ont été réalisés dans les vecteurs bi-cistroniques pETDuet®, pACYCDuet® et pCDFDuet® (Novagen).

#### I.3.1 Constructions dans le vecteur pLIM01

Les régions codantes des pilines SP0462 (RrgA), SP0463 (RrgB) et SP0464 (RrgC) chez *S. pneumoniae* TIGR4 ont été utilisées pour les clonages dans le plasmide pLIM01. La région codant pour le peptide signal à l'extrémité N-terminale a été déléetée et remplacée par une étiquette poly-histidine clivable, grâce à l'ajout d'un site spécifique reconnu par la protéase Tev. A l'extrémité C-terminale, le CWSS a également été déléeté permettant ainsi d'obtenir des protéines recombinantes solubles. Néanmoins, pour chacune des pilines les motifs LPxTG spécifiques ont été conservés (Fig.33, Tableau 4).

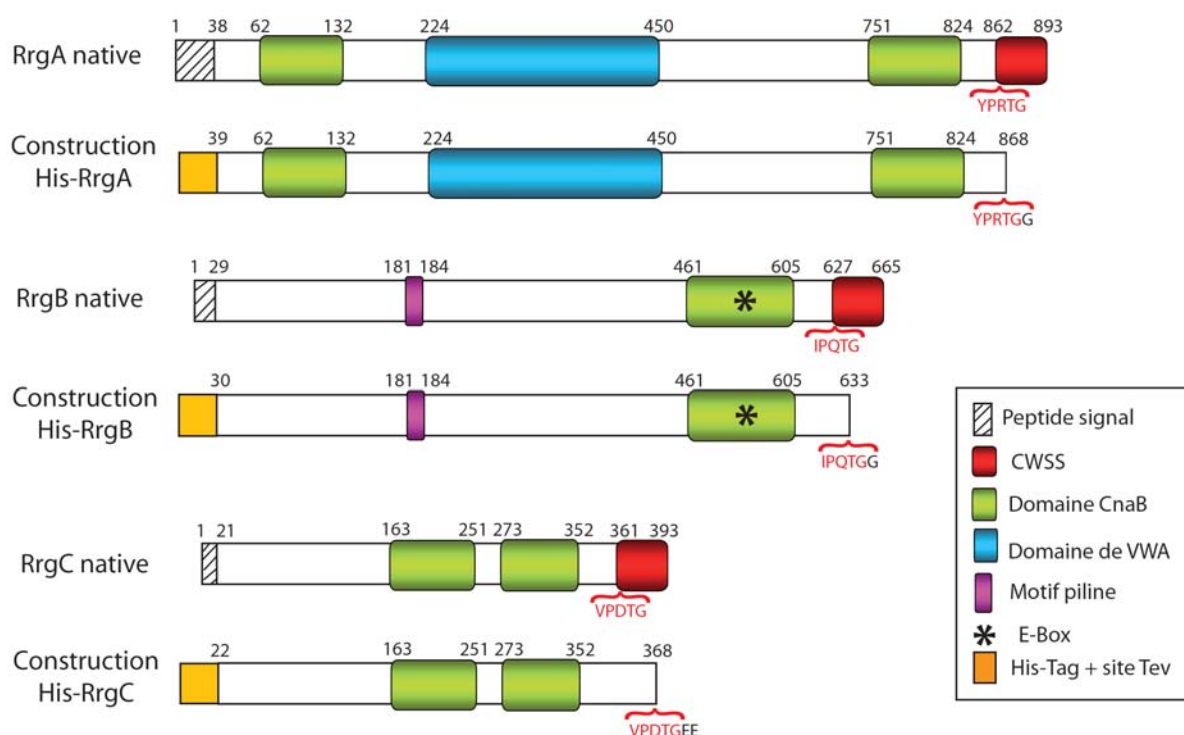


Figure 33: Représentation schématique des protéines natives et des constructions des pilines dans les vecteurs pLIM01.

**Tableau 4: Informations relatives aux protéines natives et aux protéines recombinantes provenant des constructions réalisées dans le vecteur pLIM01.**

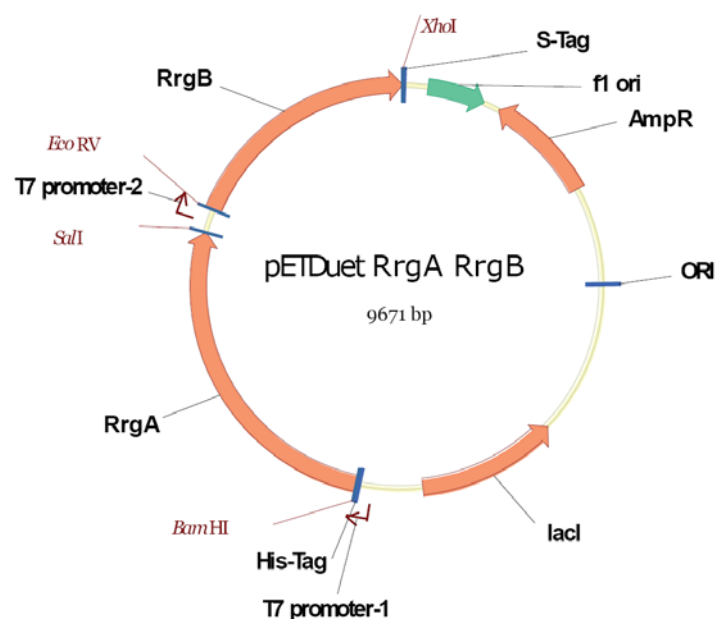
Protéine	Nombre acides aminés	Masse théorique (Da) / Pi	Masse théorique (Da) / Pi (sans Méthionine)
RrgA native	893	99 730.73 / 5.84	99 599.54 / 5.84
His*-RrgA	851	95 302.84 / 5.63	95 171.65 / 5.63
RrgA clivée His-Tag	831	92 773.09 / 5.35	
RrgB native	665	71 364.09 / 5.01	71 232.90 / 5.01
His*-RrgB	625	67 584.30 / 5.25	67 453.11 / 5.25
RrgB clivée His-Tag	605	65 054.56 / 4.95	
RrgC native	393	44 712.00 / 5.71	44 580.80 / 5.72
His*-RrgC	368	42 088.40 / 5.82	41 957.21 / 5.82
RrgC clivée His-Tag	348	39 558.66 / 5.34	

\* : Etiquette poly-histidine clivable par la protéase Tev

### I.3.2 Constructions dans les vecteurs pETDuet®, pACYCDuet® et pCDFDuet®

Les régions codantes de RrgA (39-868), RrgB (30-633), RrgC (22-368), SrtC-1 (17-228), SrtC-2 (9-220), et SrtC-3 (32-254) chez *S. pneumoniae* TIGR4 ont été utilisées pour les clonages dans les vecteurs bi-cistroniques pETDuet®, pACYCDuet® et pCDFDuet®.

Les séquences codantes de RrgA et RrgB ont été respectivement clonées dans le MCS1 et le MCS2 du vecteur pETDuet de telle sorte que la protéine RrgA produite à partir de ce vecteur présente une étiquette poly-Histidine non clivable à son extrémité N-terminale alors que RrgB présente une étiquette S-Tag à son extrémité C-terminale (Fig.34, tableau 5 et 6).



**Figure 34: Carte du vecteur pETDuet RrgA RrgB**

Les séquences codantes de RrgC et SrtC-1 ont été respectivement clonées dans le MCS1 et le MCS2 du vecteur pACYCDuet de telle sorte que la protéine RrgC produite à partir de ce vecteur présente une étiquette poly-Histidine non clivable à son extrémité N-terminale alors que SrtC-1 ne porte aucune étiquette puisqu'un codon stop a été inséré à la fin de sa séquence codante (Fig.35, tableau 5 et 6).

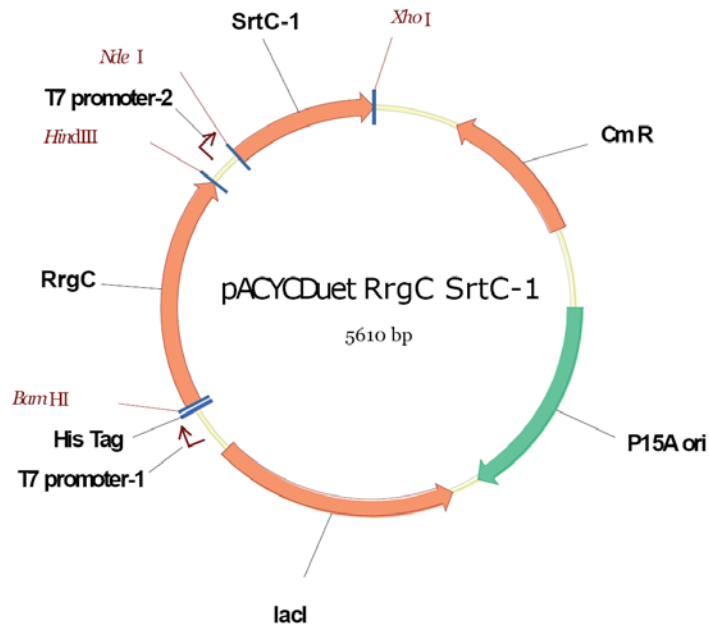


Figure 35: Carte du vecteur pACYCDuet RrgC SrtC-1

Enfin, les séquences codantes de SrtC-2 et SrtC-3 ont été respectivement clonées dans le MCS1 et le MCS2 du vecteur pCDFDuet de telle sorte que les deux protéines produites à partir de ce vecteur ne portent aucune étiquette (Fig.36, tableau 5 et 6).

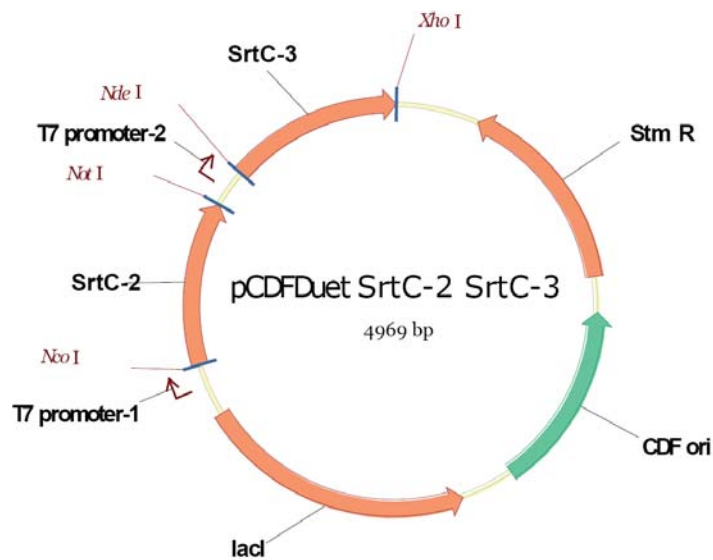


Figure 36: Carte du vecteur pCDFDuet SrtC-2 SrtC-3

**Tableau 5: Liste des amorces utilisés pour les clonages destinés aux co-expressions.**

Amorces	Sequence
RrgA_F62Bam	5' CGC G GAT CCG GAA ACG CCT GAA ACC AGT CCA GCG 3'
RrgA_R62Sal	5' GCG GTC GAC TCA ACC ACC AGT TCG AGG ATA TTC TCT C 3'
RrgB_F63Eco	5' CGC G ATA TCG GCT GGG ACG ACA ACA ACA TCT GTT ACC 3'
RrgB_R63Xho	5' GCG CTC GAG ACC ACC CGT TTG TGG GAT AGT G 3'
RrgC_F64Bam	5' CGC G GAT CCG GTC CAA GCG CAA GAA GAT CAC ACG 3'
RrgC_R64Hind	5' GCG AAG CTT TCA TTC TTC CCC TGT ATC TGG CAC 5'
SrtB_F66Nde	5' CGC CAT ATG GAA TCA AAT CAA CAA ATT GCT GAC 3'
SrtB_R66Xho	5' GCG CTC GAG TCA ATG ACT GAG TTT GTT TGC TGC AAT AAA 3'
SrtC_F67Nco	5' CGC CC ATG GGC GAG TCA AAC GAG GTT ATT AAA GAG 3'
SrtC_R67Not	5' GCG GCGGCCGC TCA TTG CCC ACG CTC TCT CAC CGC TCG 3'
SrtD_F68Nde	5' CGC CAT ATG CAG TCT CTT GGA CAG GTG AAA GGT C 3'
SrtD_R68Xho	5' GCG CTC GAG TCA AGA GGT TGC AAC ACG CGA TAC AG 3'

**Tableau 6: Informations relatives aux protéines recombinantes obtenue à partir des constructions réalisées dans le vecteur pETDuet (pour RrgA et RrgB) pACYCDuet (pour RrgC et SrtC-1) et pCDFDuet (pour SrtC-2 et SrtC-3).**

Protéine	Nombre acides aminés	Masse théorique (Da) / Pi	Masse théorique (Da) / Pi (sans Méthionine)
His_RrgA	844	94 328.70 / 5.52	94 197.51 / 5.52
RrgB_STag	636	68 463.32 / 4.95	68 332.13 / 4.95
His_RrgC	361	41 114.27 / 5.66	40 983.07 / 5.66
SrtC-1	213	23 857.10 / 5.06	23 725.91 / 5.06
SrtC-2	214	24 183.51 / 5.05	24 052.32 / 5.05
SrtC-3	224	25 088.31 / 5.09	24 957.11 / 5.09

### I.3.4 Récapitulatif des vecteurs construits pour cette étude

La totalité des vecteurs construits pour l'étude ainsi que les oligonucléotides utilisés pour les réalisés sont listés en annexe (Annexes 1 et 2)

### I.3.5 Récapitulatif des souches d'expression réalisées pour cette étude

Les premières souches utilisées notamment pour l'étude des pilines individuelles ont été réalisées dans des souches d'expression *E. coli* BL21 (DE3) RIL par Anne Marie Di Guilmi avant mon arrivée au laboratoire (cf. Annexe 2 A).

Les 32 combinaisons nécessaires au système de co-expression ont été réalisées dans des souches d'expression *E.coli* STAR car contrairement aux souches *E. coli* BL21 (DE3) RIL, celles-ci ne présentent pas de résistance naturelle au chloramphénicol, résistance portée par le vecteur pACYCDuet (cf. Annexe 2B).

Les 57 souches *E. coli* STAR exprimant des vecteurs mutés nécessaires à l'analyse complète des co-expressions sont listées dans l'annexe 3 C.

## **I.4 Milieux de culture et antibiotiques**

### **I.4.1 Milieux de culture**

Les souches *E. coli* sont cultivées en milieu riche :

- Liquide Luria Bertina (LB) : contenant 10g de Bacto-Tryptone, 5g de Bacto-Yeast extract et 10 g de NaCl par litre (milieu destiné aux précultures)
- Liquide Terrific Broth (TB) : contenant 12g de Bacto-Tryptone, 24g de Bacto-Yeast Extract et 4 ml de glycérol par litre (milieu destiné aux cultures)
- Liquide SOB : contenant 20g de Bacto-Tryptone, 5g de Bacto-Yeast Extract, 0.5g de NaCl et 2.5 ml de KCl 1M par litre auxquels sont ajoutés 10 ml de MgCl<sub>2</sub> 1M (milieu destiné à la préparation de cellules compétentes)
- Liquide SOC : contenant du SOB supplémenté de 20 ml de glucose 1M stérile (milieu utilisé pour les cultures bactérienne fraîchement transformées)
- Solide : LB supplémenté de 15g d'agar

### **I.4.2 Antibiotiques**

L'ampicilline (Euromedex), le chloramphénicol (Euromedex) et la streptomycine (Euromedex) sont ajoutés aux milieux de culture (liquides ou solides) à une concentration finale de 100µg/ml, 34µg/ml et 50µg/ml respectivement.

## II. Techniques de biologie moléculaire

### II.1 Clonage

#### *Amplification de fragment d'ADN par PCR (Polymerase Chain Reaction)*

L'ADN polymérase utilisée est l'enzyme Phusion (Finnzymes)

Mélange réactionnel :

100 ng d'ADN génomique TIGR4

0.5  $\mu$ M de chaque oligonucléotide

200  $\mu$ M de chaque dNTP

10  $\mu$ l Tampon de réaction 5X

x  $\mu$ l H<sub>2</sub>O (qsp 50 $\mu$ l)

0.5 $\mu$ l Phusion (2U/ $\mu$ l)

La réaction d'amplification PCR débute par une dénaturation initiale à 98°C puis nécessite 35 cycles des étapes suivantes:

- Dénaturation de l'ADN à 98°C durant 10 secondes
- 30 secondes d'hybridation spécifique des oligonucléotides à une température fonction de la composition en AT et GC de ces derniers
- Elongation à 72°C avec l'ADN polymérase à raison de 30 secondes par kilobase d'ADN

La réaction nécessite une élongation terminale d'une durée de 10 minutes à 72°C.

Les produits PCR sont ensuite purifiés à l'aide du kit Nucléospin Extract II® (Macherey Nagel) et élués dans 50  $\mu$ l d'eau.

#### *Digestion par des endonucléases de restriction*

Les réactions sont réalisées comme suit :

15 $\mu$ l de produit PCR (insert) ou 10 $\mu$ g de vecteur

Tampon de restriction approprié 10 X

1 $\mu$ l endonucléase 1

1 $\mu$ l endonucléase 2

Les digestions ont généralement lieu à 37°C durant 2 heures. L'ADN digéré est analysé sur gel d'agarose, extrait et purifié à l'aide du kit Nucléospin Extract II® (Macherey Nagel) et élué dans 30 $\mu$ l d'eau.

### Ligation

La ligation des inserts et des vecteurs digérés est réalisée par l'ADN ligase du phage T4 (Fermentas). Les réactions sont réalisées 1 heure à température ambiante avec toujours deux ratios vecteur/insert :

Ratio vecteur/insert	1/2	1/3
Ligase	1µl	1µl
Tampon ligase	1µl	1µl
H2O q.s.p.	10µl	10µl

Les produits de ligation peuvent ensuite servir à transformer les bactéries *E. coli* DH5α compétentes

## II.2 Mutagenèse

Les mutations ont été introduites dans les différents vecteurs soit par mutagenèse dirigée basée sur la PCR soit par clonage lorsqu'un vecteur intermédiaire muté était déjà disponible.

### Mutagenèse dirigée basée sur la PCR

Les mutations ponctuelles ont été réalisées par PCR à l'aide du kit QuickChange® II XL commercialisé par Stratagene. La polymérase du kit est l'enzyme Pfu Ultra.

Mélange réactionnel:

- 10 ng de matrice ADN (vecteur X)
- 125 ng oligonucléotide D
- 125ng oligonucléotide R
- 1 µl dNTP mix
- 5 µl Tampon de réaction 10X
- 3 µl Quick solution
- 34.5 µl H<sub>2</sub>O
- + 1µl Pfu Ultra (2.5U/µl)

La mutagenèse nécessite 18 cycles PCR des étapes suivantes:

- Dénaturation de la matrice à 95°C durant 50 secondes
- 50 secondes d'hybridation spécifique des oligonucléotides à une température fonction de la composition en AT et GC de ces derniers
- Elongation à 68°C avec l'ADN polymérase à raison de 1 min par kilobase d'ADN



Le produit PCR est ensuite digéré 1 h à 37°C par l'enzyme DpnI qui a pour particularité de digérer l'ADN méthylé donc la matrice ADN. Le produit PCR (1µl) traité à la DpnI peut ensuite servir à transformer les bactéries *E. coli* DH5α compétentes.

### II.3 Préparation des cellules « Ultra-compétentes » *E. coli* STAR

Le protocole utilisé est issu de la « méthode Inoue » pour la préparation et la transformation de cellules *E. coli* « Ultra-compétentes » (Inoue, 1990).

Tampons de transformation Inoue

-MnCl <sub>2</sub> .4H <sub>2</sub> O	55 mM
-CaCl <sub>2</sub> .2H <sub>2</sub> O	15 mM
-KCl	250 mM
-PIPES (0,5M, pH6,7)	10 mM

(PIPES: piperazine-1,2bis(2-ethanesulfonique acid))

5ml de LB sont inoculés avec un stock congelé de la souche *E. coli* STAR (DO 0.3) et cultivé 16h à 37°C sous agitation. Cette préculture est inoculée dans 250 ml de milieu SOB incubé à 37°C sous agitation jusqu'à atteindre une DO<sub>600</sub> de 0,6. La culture est alors arrêtée par centrifugation à 2500g durant 10 minutes. Le surnageant est éliminé, puis le culot bactérien est repris dans 80ml de tampon Inoue froid. Une nouvelle centrifugation est réalisée et le culot bactérien ainsi obtenu est repris à froid dans 20ml d'Inoue. 1,5 ml de DMSO sont ajoutés à la suspension bactérienne qui repose 10 minutes dans la glace avant d'être utilisable pour transformation ou pour congélation à l'azote liquide. Les aliquots congelés peuvent être conservés 2 à 3 mois sans diminution notable de la compétence.

### II.4 Transformation des cellules compétentes

1µl de PCR de mutagenèse ou 10 µl de ligation ou 1µl d'ADN plasmidique sont incubés en présence de 50µl de cellules compétentes pendant 30 minutes dans la glace, puis soumis à un choc thermique à 42°C pendant 45 secondes. Les cellules sont ensuite placées dans la glace durant 2 minutes puis incubées durant 1 heure à 37°C sous agitation dans 450µl de milieu SOC. Les cultures bactériennes sont centrifugées puis le culot bactérien est repris dans 50µl de SOC pour être finalement étalé sur boîte de pétri contenant du LB agar supplémenté de ou des antibiotiques de sélection. Les boîtes sont ensuite incubées à 37°C toute la nuit.

## **II.5 Mini préparation plasmidique et séquençage**

Pour chaque vecteur réalisé, deux colonies ont été sélectionnées aléatoirement et cultivées 16h à 37°C dans 4 ml de LB contenant l'antibiotique de sélection.

1 ml de cette culture permet de constituer un stock glycérolé conservé à -80°C et à partir des 3 ml restants, l'ADN plasmidique est extrait à l'aide du kit Nucleospin Plasmid QuikPure (Stratagene). Chacun des vecteurs réalisé lors de cette étude a été vérifié par séquençage (Beckman Coulter).

### **III. Techniques biochimiques**

#### **III.1 Production des protéines recombinantes en grand volume**

50 ml de LB contenant l'antibiotique approprié sont inoculés à partir des stocks glycérolés de *E. coli* BL21(DE3)RIL ou STAR et cultivés 16h à 37°C afin de servir de préculture. Celle-ci est ensuite transférée dans 1L de TB contenant l'antibiotique approprié et incubée à 37°C jusqu'à une DO<sub>600nm</sub> approximative de 0,5. La surexpression des protéines d'intérêt est induite par l'ajout d'IPTG 0.5 mM pendant 3 heures à 37°C. Les cultures sont ensuite centrifugées 10 minutes à 4000 rpm. Les culots bactériens sont repris dans 50mL de tampon A (50 mM HEPES pH 7,5 / 200 mM NaCl / 20 mM Imidazole) contenant 1 pastille de Complete EDTA-free (Roche) et lysés par sonication. Les lysats sont ensuite centrifugés 20 minutes à 18000 rpm et le surnageant conservé pour purification.

#### **III.2 Production des protéines recombinantes en mini-cultures**

*Toutes les protéines recombinantes destinées aux analyses des co-expressions ont été purifiées en parallèle sur des plaques multi-puits.*

4 ml de LB contenant l'antibiotique approprié sont inoculés à partir des stocks glycérolés de *E. coli* STAR et cultivés 16h à 37°C afin de servir de précultures. 100µl de chacune des précultures sont ensuite transférées toujours en plaque multi-puits dans 4ml de TB contenant l'antibiotique approprié et incubés à 37°C durant 2 heures. La surexpression des protéines d'intérêt est induite par l'ajout d'IPTG 0.5 mM pendant 3 heures. Les cultures sont ensuite centrifugées 10 minutes à 4000 rpm. Les culots bactériens sont lysés chimiquement durant 20 minutes par 1 ml de BugBuster® 1X (Novagen) dilué dans du tampon A (50 mM HEPES pH7,5 / 200 mM NaCl / 20 mM Imidazole) et supplémenté de 2µl de r-lysozyme (Novagen) et 0.8 µl de benzonase nucléase (Novagen). Les lysats bactériens sont clarifiés par une nouvelle centrifugation puis les surnageants conservés pour purification.

#### **III.3 Purification des protéines recombinantes**

Les purifications des protéines, pour leur étude individuelle, ont été réalisées par chromatographie d'affinité au nickel sur une colonne HisTrap de 1 ml (Amersham), le lavage est effectué avec 45ml de Tampon A (50 mM HEPES pH7,5 / 200 mM NaCl / 20 mM Imidazole) . L'élution est réalisée par un gradient croissant d'Imidazole à l'aide d'un Tampon B (50 mM HEPES pH7,5 / 200 mM NaCl / 500 mM Imidazole). Les éluions de purification

sont analysées par électrophorèse sur gel polyacrylamide en conditions dénaturantes (pourcentage du gel polyacrylamide variable de 7.5% à 15% selon la protéine considérée). Les fractions contenant les protéines d'intérêt sont rassemblés et dialysés une nuit contre un Tampon C (50 mM HEPES pH7.5 / 50 mM NaCl).

En fonction des analyses réalisées sur les protéines recombinantes l'étiquette poly-histidine peut-être clivée par l'utilisation de protéase Tev (Tobacco Etch Virus) produite au laboratoire. Le clivage à lieu durant 2 heures à un ratio 3 :1 (protéine :Tev) et les protéines d'intérêts sont séparées de la Tev par une nouvelle chromatographie d'affinité Nickel.

La dernière étape de purification est réalisée par chromatographie d'exclusion (Filtration sur gel préparative S200 Superdex HR de 120 ml).

### ***Cas particuliers : purifications des complexes nécessitant des étapes supplémentaires***

#### **- Purification du complexe RrgB-RrgC catalysé par SrtC-1**

Après chromatographie d'affinité au nickel sur une colonne HisTrap de 5 ml (Amersham), les fractions contenant le complexe RrgB-RrgC sont rassemblées, dialysées contre un tampon D (50mM MES pH6.0 / 50 mM NaCl) et soumis à une chromatographie sur une colonne Ressource-Q 6 ml échangeuse d'anions (GE-Healthcare). L'élution est réalisée par un gradient de sels (NaCl 50-250 mM). Une dernière étape de chromatographie d'exclusion (S200 Superdex, GE Healthcare) est réalisée dans un tampon E (50mM MES pH6.0, 150mM NaCl). La quantité finale de complexe purifié obtenue est de l'ordre de 500 µg par litre de culture.

#### **- Purification du complexe acyl-enzyme RrgC\*SrtC-3**

Après chromatographie d'affinité au nickel sur une colonne HisTrap de 5 ml (Amersham), les fractions contenant le complexe RrgC\*SrtC-3 sont rassemblées, dialysées contre un tampon F (50mM HEPES pH7.0 / 50 mM NaCl) et soumis à une chromatographie sur une colonne Ressource-Q 6 ml échangeuse d'anions (GE-Healthcare). L'élution est réalisée par un gradient de sels (NaCl 50-250 mM). Une dernière étape de chromatographie d'exclusion (S200 Superdex, GE Healthcare) est réalisée dans un tampon G (50mM HEPES pH7.0, 150mM NaCl). La quantité finale de complexe purifié obtenue est de l'ordre de 1 mg par litre de culture.

### **III.4 Minipurification des surnageants de co-expressions**

Les surnageants provenant des mini-cultures des co-expressions, soit environ 750µl, sont purifiés sur le système His-MultiTrap® HP-columns (GE Healthcare). Ce système en plaque présente 96 mini-colonnes contenant chacune une résine Nickel sepharose haute performance, permettant ainsi de réaliser les mini-purifications en parallèle. Les colonnes sont lavées 2 fois par du tampon A avant le dépôt des surnageants bactériens qui est suivi de trois étapes de lavage par du tampon A. Les protéines d'intérêt sont éluées en une seule étape par ajout de 200 µl de tampon B.

### **III.5 Analyses des protéines**

#### **III.5.1 Analyse des protéines par électrophorèse en conditions dénaturantes (SDS-PAGE)**

Les échantillons protéiques sont incubés avec une solution de bleu Laemmli de concentration finale 1X (Composition 1X : 62.5mM Tris pH6.8 / 2% SDS / 650 mM β-mercaptoéthanol / 10% glycérol / 0.1% bleu de bromophénol) puis chauffés à 100°C durant 10 minutes. Les échantillons sont ensuite chargés sur des gels de polyacrylamide dont le pourcentage en Acrylamide/Bisacrylamide est variable en fonction de la protéine analysée : 7.5 à 15% ou sur des gels commerciaux Criterion XT precast gel (Biorad) qui présentent un gradient Bis-Tris de 4-12% permettant de conserver une bonne résolution tant pour les complexes de hauts poids moléculaires que pour les petites protéines. La migration est réalisée à 200V soit dans un tampon d'électrophorèse 1X réalisé au laboratoire (Tris 25mM / SDS 0.1% / Glycine 192mM) soit dans le tampon MOPS commercialisé par Biorad.

#### **III.5.2 Coloration des protéines au bleu de Coomassie**

Le gel de protéine est plongé durant 30 minutes dans une solution de coloration de composition suivante: éthanol 50%, acide acétique 8%, 0.25% Coomassie R250. Celui-ci est ensuite rincé à l'eau et décoloré par une solution de décoloration composé par de l'éthanol 5% et de l'acide acétique 7.5%.

#### **III.5.3 Transfert des protéines sur membrane de nitrocellulose**

Les protéines séparées par SDS-PAGE sont electro-transférées sur une membrane de nitrocellulose (BioRad) à un voltage de 100V durant 30 minutes. Le transfert est réalisé à froid en présence d'un tampon TransBlot 1X (Tris-HCl 25 mM / Glycine 190mM)

### III.5.4 Immunorévélation des protéines sur membrane de nitrocellulose

Les membranes de nitrocellulose sont saturées dans une solution de PBS Tween 0.3% contenant 5% lait durant une heure.

- *Pour une détection His-Tag* : La membrane de nitrocellulose est incubée durant 1 heure avec des anticorps anti-His-Tag couplés à la HRP (Horse Radish Peroxydase commercialisé par SIGMA) utilisés au  $1/50000^{\text{ème}}$  dans une solution de PBS Tween 0.3% contenant 5% lait.
- *Pour une détection anti-RrgA, anti-RrgB ou anti-RrgC* : La membrane de nitrocellulose est incubée durant 1h30 avec des anticorps primaires polyclonaux anti-RrgA, Anti-RrgB ou anti-RrgC qui ont été produits chez des souris. Ces anticorps sont utilisés au  $1/5000^{\text{ème}}$  dans une solution de PBS Tween 0.3% contenant 5% lait. Après 3 lavages au PBS Tween 0.3%, la membrane est incubée durant 1h avec des anticorps secondaires anti-souris couplés à la HRP utilisés au  $1/120000^{\text{ème}}$  dans une solution de PBS Tween 0.3% contenant 5% lait.
- *Pour une détection anti-SrtC-1, anti-SrtC-2 ou anti-SrtC-3* : La membrane de nitrocellulose est incubée durant 1h30 avec des anticorps primaires polyclonaux anti-SrtC-1, Anti-SrtC-2 ou anti-SrtC-3 qui ont été produits chez des lapins. Ces anticorps sont utilisés au  $1/2000^{\text{ème}}$  dans une solution de PBS Tween 0.3% contenant 5% lait. Après 3 lavages au PBS Tween 0.3%, la membrane est incubée durant 1h avec des anticorps secondaires anti-lapin couplés à la HRP utilisés au  $1/10000^{\text{ème}}$  dans une solution de PBS Tween 0.3% contenant 5% lait.

Les membranes de nitrocellulose sont ensuite lavées 3 fois par une solution de PBS Tween 0.3%, puis révélées par chimioluminescence (ECL, Amersham).

### III.5.7 Tests de stabilité par protéolyse à la trypsine et séquençage N-terminal

20µg de chaque protéine purifiée est mélangé à 50 ng de Trypsine (Sigma) et la digestion enzymatique à lieu à température ambiante durant 16 heures. Les échantillons sont ensuite chauffés à 100°C durant 10 minutes en présence de bleu laemmli 1x puis analysés par SDS-PAGE. Les protéines sont soit colorées par une solution de bleu de Coomassie soit électrotransférés sur une membrane Problott® (applied Biosystems) puis colorées selon les instructions du fournisseur. Les bandes à analyser ont été excisées des membranes Problott® et le séquençage N-terminal a été réalisé au Laboratoire d'enzymologie Moléculaire de l'IBS par Jean-Pierre Andrieu.

### **III.5.8 Tests de stabilité par Thermal Shift Assay (TSA)**

Les expériences de TSA sont réalisées à l'aide d'un instrument commercialisé par BioRad qui se compose d'un module IQ5 couplé à un thermocycleur. Cet instrument permet de suivre en temps réel l'évolution de fluorescence d'échantillons, protéiques dans notre cas, dans un format de plaque 96 puits (le principe de la technique est décrit ci-dessous).

25ng de chacune des protéines purifiées RrgA, RrgB et RrgC ainsi que leurs variants sont mélangés à 2  $\mu$ l de Sypro Orange 100x (sonde moléculaire) dans un volume final de 25 $\mu$ l. Chaque expérience est réalisée en triplicat. Les échantillons sont ensuite dénaturés à raison de 1°C par minute (20-95°C) et la fluorescence émise par la sonde fluorescente est mesurée tout au long de l'expérience.

## IV Principes

### IV.1 Principe du Thermal Shift Assay

Le thermal Shift Assay (TSA) est une méthode qui permet d'étudier la stabilité thermique d'une protéine purifiée. Le principe de la méthode consiste à mélanger l'échantillon protéique avec un fluorophore, le Sypro Orange, qui se fixe et ne fluoresce que dans un environnement hydrophobe. L'échantillon est chauffé progressivement, à raison de 1 °C / min. A basse température, la protéine est repliée, ne présentant à sa surface que des résidus hydrophiles. Lorsque la température de fusion  $T_m$  de la protéine (pour melting temperature) est atteinte, la protéine se dénature, exposant ses résidus hydrophobes sur lesquels le Sypro Orange se fixe et fluoresce. Le  $T_m$  est défini comme le point d'inflexion de la courbe schématisée figure 37. Afin d'extraire facilement le  $T_m$  la dérivée première de la courbe est calculée. Au  $T_m$ , la courbe obtenue présente un minimum. Si plusieurs espèces se trouvent en solution ou si la protéine est multi-domaines, il est possible d'observer plusieurs  $T_m$  différents (Pantoliano 2001, Geerlof 2006).

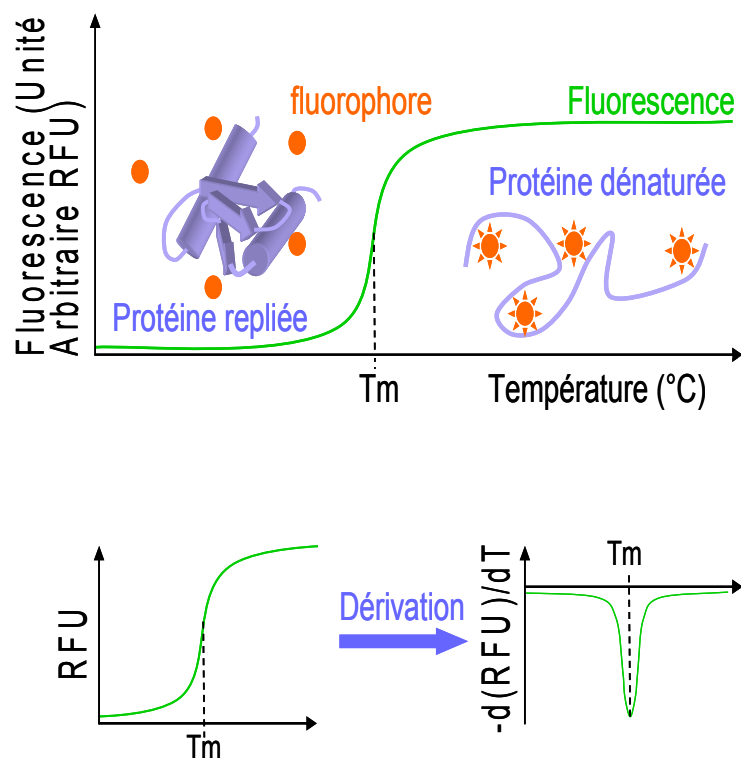


Figure 37: Schéma du principe de la méthode TSA (en haut). Calcul mathématique effectué par le logiciel IQ5 donnant un accès rapide aux  $T_m$  (en bas)



## **IV.2 Principe de la plateforme de co-expression chez *E. coli***

Un système de co-expression permettant de tester toutes les combinaisons des protéines impliquées dans la formation du pilus a été mis en place afin de comprendre le rôle de chacune des sortases dans l'association des pilines accessoires (RrgA et RrgC) avec la piline majeure (RrgB). Ce système a mené à l'obtention de souches *E. coli* STAR exprimant une, deux ou les trois protéines de surface du pilus associées ou non à une, deux ou les trois sortases. Dans ce système, seules les protéines RrgA et RrgC possèdent une étiquette poly-histidine, en N-terminal, permettant leurs rétentions sur une colonne d'affinité Nickel, la protéine RrgB possède un S-tag en C-terminal et les trois sortases ne portent pas d'étiquette. Ainsi, seules les protéines interagissant de manière forte avec RrgA ou RrgC sont présentes dans les fractions d'élution d'une purification sur colonne Nickel. C'est la raison pour laquelle les fractions d'élution sont analysées par Western Blot anti-RrgB ou anti-sortase.

32 souches *E. coli* STAR ont été nécessaires pour identifier le rôle de chacune des sortases. L'expression des protéines à partir de ces souches a été systématiquement testée avant d'être analysée plus avant. 57 souches supplémentaires exprimant des pilines et ou des sortases mutées ont été réalisées car nécessaires à cette étude (ponts intramoléculaires, délétions et inversion des motifs LPxTG).

# **Présentation des travaux**



## PRESENTATION DES TRAVAUX

Mes travaux de thèse s'intègrent dans la thématique « hôte-pathogène » développée au laboratoire par Anne Marie Di Guilmi et qui porte plus particulièrement sur les facteurs de virulence du pneumocoque. Mes travaux concernant l'étude du pilus PI-1 de *S. pneumoniae* ont porté sur (i) la caractérisation biochimique de chacune des protéines impliquées dans la formation du pilus (ii) la mise en place d'une plateforme de co-expression chez *E. coli* permettant d'élucider, étape par étape, les mécanismes moléculaires de la biogenèse du pilus.

Comme décrit précédemment dans l'introduction, le pilus est une structure importante dans la virulence du pneumocoque puisqu'il intervient dans la colonisation, dans l'établissement de maladies invasives et dans l'induction de la réponse inflammatoire chez l'hôte (Barocchi 2006). Plusieurs études menées sur le rôle des trois pilines formant le pilus restent incomplètes (Hava 2002, Munoz-Elias 2008, Nelson 2007, Hilleringmann 2008). D'autre part, la biogenèse de la structure même reste encore largement incomprise et fait l'objet de nombreuses controverses (Barocchi 2006, LeMieux 2006, Hilleringmann 2008, LeMieux 2008, Fälker 2008, Hilleringmann 2009). C'est dans le but d'apporter des réponses à ces questions que nous avons choisi d'utiliser des approches biochimiques et structurales.

Dans un premier temps, nous avons exprimé de façon soluble chacune des pilines et des sortases, afin de les étudier individuellement d'un point de vue structural dans le but d'élucider ou de mieux comprendre leur fonction.

Dans un second temps, nous avons cherché à comprendre le rôle de chacune des sortases dans la formation du pilus. Les travaux publiés dans ce domaine consistent à identifier la fonction individuelle de chaque sortase en analysant le phénotype de la souche mutante dont le gène de la sortase étudié a été inactivé. Ces études ne permettent pas d'apporter de réponse très claire quant au rôle spécifique de chacune des sortases dans la biogenèse du pilus. C'est la raison pour laquelle, nous avons fait le choix d'une stratégie inverse qui consiste à reconstruire *in vitro* le pilus afin d'élucider, pas à pas, les mécanismes moléculaires de sa biogenèse. Nous avons ainsi développé une plateforme de co-expression chez *E. coli* permettant de produire chacune des trois pilines et des trois sortases, dans toutes les combinaisons possibles. Le postulat de départ, sur lequel un consensus est établi, est que RrgB, la piline majeure, forme le squelette du pilus et que les pilines accessoires RrgA et

RrgC y sont associées. Les questions posées dans ce travail sont: quelle(s) sortase(s) catalyse la polymérisation de RrgB? Comment les pilines accessoires RrgA et RrgC sont-elles incluses dans la fibre de RrgB? Quelle(s) sortase(s) sont impliquées dans ce processus? Dans notre système de co-expression, seules les pilines accessoires portent des étiquettes poly-histidines permettant leur purification sur des colonnes de Nickel. Toutes les combinaisons de co-expression de pilines et de sortases ont été réalisées et soumises à une purification sur colonne Nickel et chaque fraction d'élution est analysée pour identifier RrgB par immuno-détection. Ce système permet donc d'identifier la formation de complexes covalents faisant intervenir chacune des trois pilines. Dans ces travaux, chaque sortase a été étudiée séparément, par conséquent, leur rôle dans la formation des complexes peut être proposé de façon non ambiguë pour chacune d'entre elles.

# Résultats



## RESULTATS

### I. Analyse structurale des pilines de *S. pneumoniae*

Une étape préliminaire à l'étude des mécanismes moléculaires de biogénèse du pilus est l'analyse structurale des protéines formant la fibre, les pilines, et les enzymes qui catalysent leur association covalente, les sortases. Au cours de ces travaux de thèse, nous avons eu recours à l'utilisation combinée de la cristallographie des rayons X, qui permet d'accéder aux structures de protéines à résolution atomique et à d'outils biochimiques et biophysiques qui apportent des informations concernant le repliement et la stabilité des protéines. L'ensemble des données obtenues concernant les pilines ont mis en évidence : (1) l'organisation modulaire de chacune d'entre elles, (2) le rôle de certains résidus clés dans les interactions inter-pilines et (3) la présence d'une modification post-traductionnelle particulière interne à chaque piline.

Les pilines du pneumocoque ont été exprimées dans le système hétérologue *E. coli*. Pour chacune d'elles, la délétion du peptide signal et d'une partie de la région du CWSS en aval du motif LPxTG, l'hélice transmembranaire et la région cytoplasmique chargée positivement, ont permis de produire les pilines sous forme soluble dans le cytoplasme de souches d'*E. coli*. L'étude individuelle des pilines a été effectuée soit à partir de constructions déjà disponibles au laboratoire à mon arrivée (vecteur pLIM01) soit par des constructions effectuées au cours de ma thèse (vecteurs pETDuet et pACYCDuet). Les structures cristallographiques ont été réalisées en collaboration avec l'équipe d'Andréa Dessen à l'IBS par Clothilde Manzano, Thierry Izoré et Carlos Contreras-Martel.

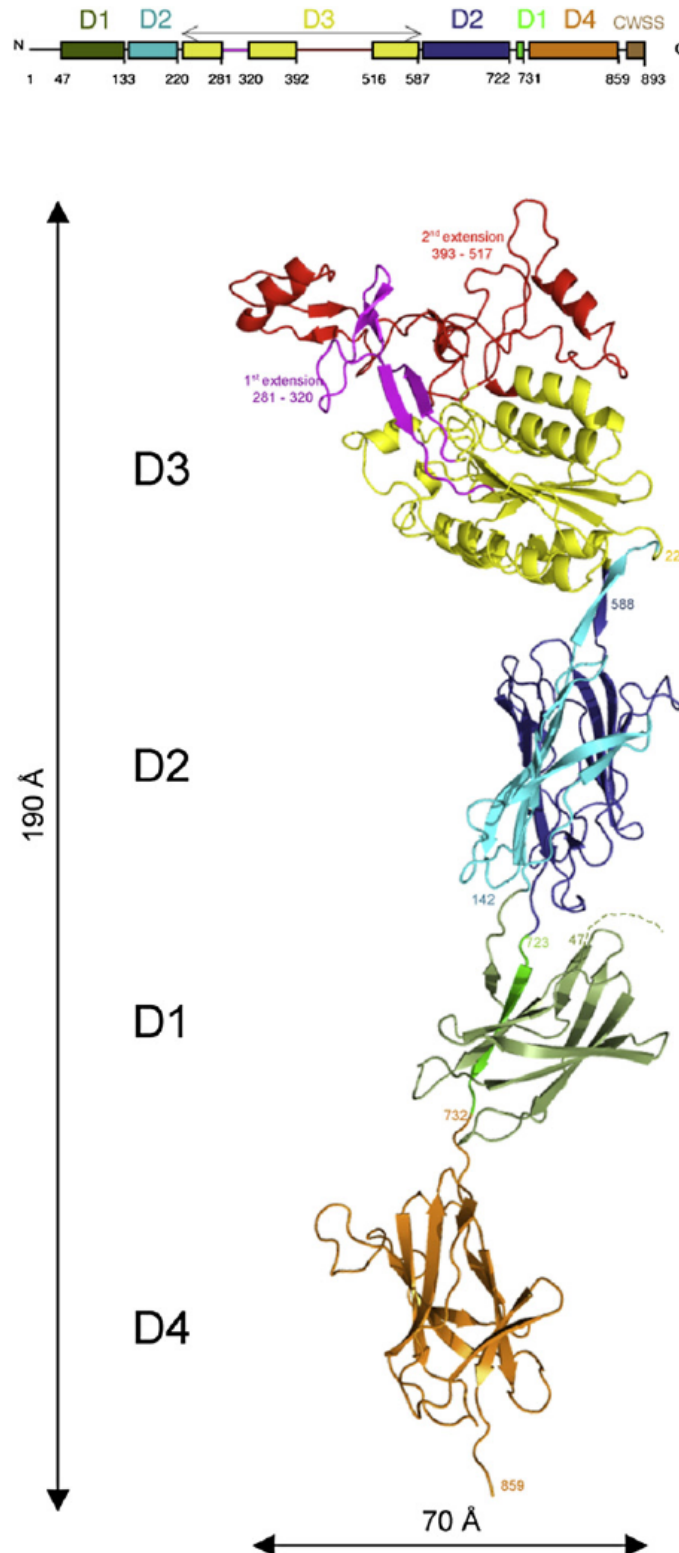


## **I.1 Structure de la piline accessoire RrgA**

Chez *S. pneumoniae*, la fonction d'adhésion du pilus aux tissus de l'hôte est portée par la piline RrgA. La délétion du gène *rrgA* dans la souche TIGR4 conduit à des défauts de virulence et de colonisation de la souche mutante dans des modèles murins d'infection par le pneumocoque (Hava, 2002). Des résultats similaires ont été obtenus dans un modèle cellulaire *in vitro* : la souche bactérienne n'exprimant plus RrgA a un pouvoir d'adhésion réduit d'un facteur huit sur des cellules épithéliales pulmonaires A549 (Nelson, 2007). Enfin, des tests d'interaction en phase solide montrent que la forme recombinante de RrgA se lie à la fibronectine, au collagène de type I et à la laminine (Hilleringman, 2008). Au vu de l'importance du rôle de RrgA, l'étude structurale de cette piline a été menée dans le but de comprendre les bases moléculaires de la fonction d'adhésion associée à RrgA. La résolution de la structure cristallographique a été effectuée par Thierry Izoré, Carlos Contreras-Martel et Andréa Dessen à l'IBS (Izoré, 2009). RrgA est une molécule allongée, composée de quatre domaines (D1, D2, D3 et D4), associés semi linéairement en formant un collier de perle (Fig.38).

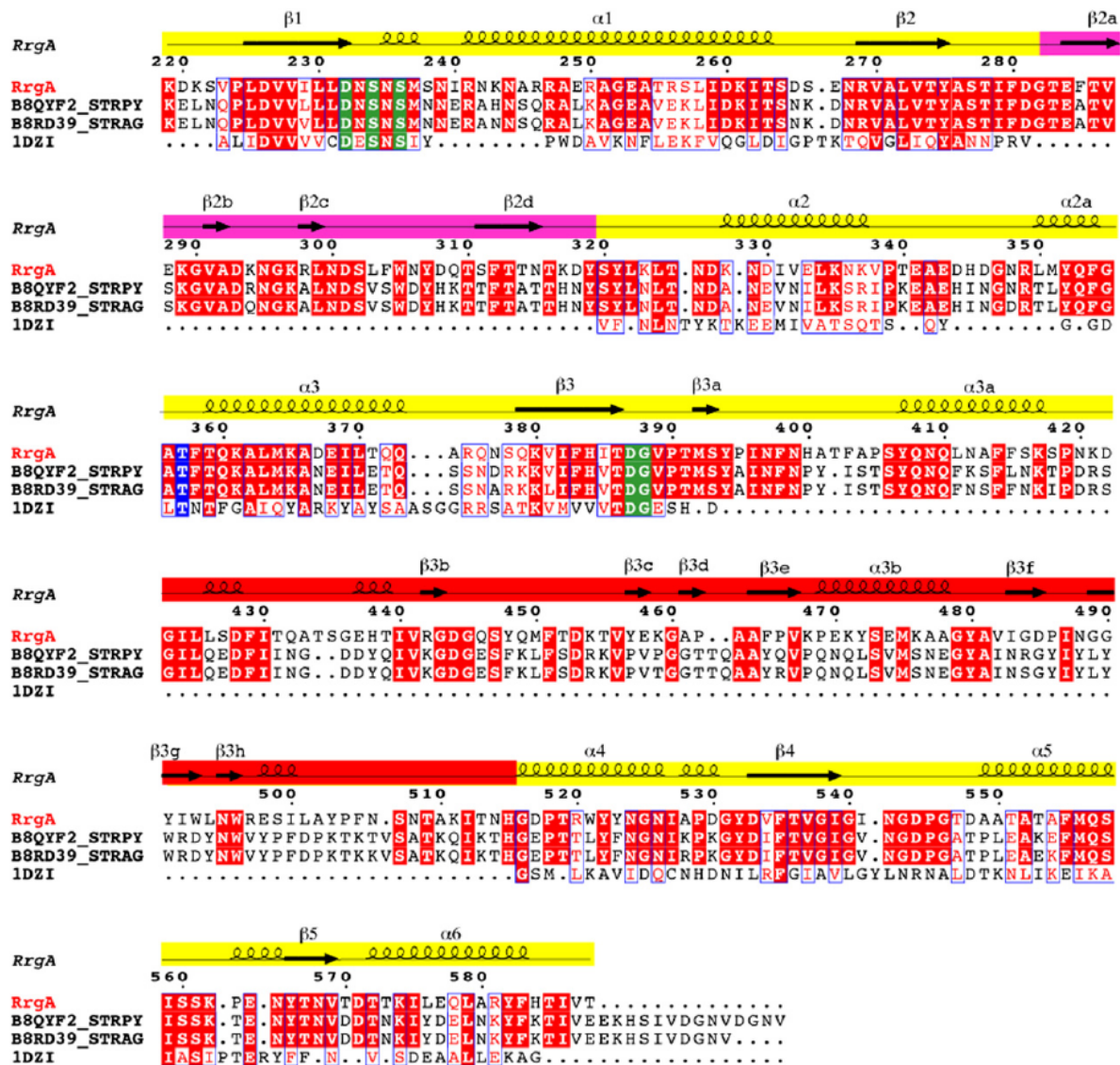
### **I.1.1. Bases structurales de la fonction d'adhésion de RrgA**

Le domaine D3, localisé à une des extrémités de la protéine, est formé par un domaine similaire au domaine A3 du facteur de von Willebrandt (VWA) humain et au domaine I présent dans plusieurs familles d'intégrine (en jaune dans la figure 38). Ces protéines eucaryotes sont très souvent impliquées dans des interactions avec d'autres protéines, via notamment un site MIDAS (Métal Ion Dépendent Adhésion Site) également présent dans RrgA (Izoré, 2009). Il est donc envisageable que les fonctions de reconnaissance de RrgA pour les récepteurs cellulaires, dont les protéines de la matrice extracellulaire mentionnées ci-dessus, puissent être médiées par un ou plusieurs sites portés par ce domaine d'origine eucaryote. De façon intéressante, ce domaine globulaire est décoré par des régions structurées (en rouge et magenta dans la figure 38), dont les séquences sont insérées dans celle codant pour le domaine de von Willebrandt.



**Figure 38: Arrangement schématique de la structure tertiaire de RrgA.** (En haut) RrgA est composée de quatre domaines, où D3 est inséré dans D2, et D2/D3 sont insérés dans D1. (En bas) En représentation tridimensionnelle, les quatre domaines de RrgA sont alignés semi-linéairement en formant un collier de perle; D1 est représenté en vert foncé (région N-terminale) et vert clair (feuillet C-terminale); le domaine D2 de type Cna-B est formé de résidus provenant de la région N-terminale (bleu clair) et de la région C-terminale (bleu foncé) de la molécule. Le domaine de type intégrine est représenté en jaune, avec deux bras étendus en rouge et magenta. L'extrémité C-terminale de RrgA est composée du domaine D4 (orange) (Izoré *et al.* 2010).

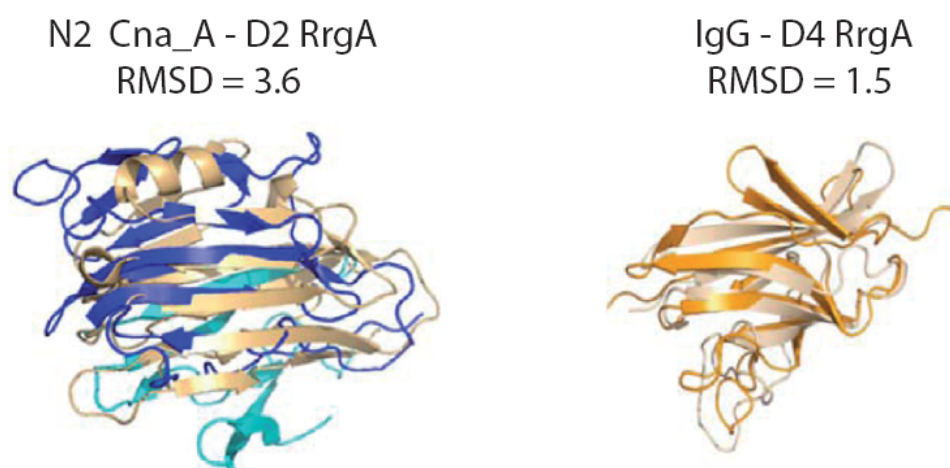
Sur la base de ces données structurales, une analyse de séquences des adhésines de pili chez les Gram-positifs montre que ces séquences insérées, dans le domaine de von Willebrandt, sont exclusivement présentes dans les adhésines de Streptocoques (Fig.39). La détermination de la structure tridimensionnelle de RrgA a donc mis en évidence la topologie et le repliement originaux de ce domaine D3 ainsi que son origine composite, eucaryote et procaryote.



**Figure 39: L'Adhésine du pilus de *S. pneumoniae* montre des caractéristiques structurales d'origine multiple.** Alignement de séquence du domaine D3 de RrgA avec des domaines analogues d'adhésines de pili (homologues de RrgA) de *S. agalactiae* (B8RD39) et *S. pyogenes* (B8QYF2) de même qu'avec le domaine I de l'intégrine eucaryote  $\alpha 2\beta 1$ . Les résidus identiques sont indiqués en blanc sur fond rouge. Les résidus impliqués dans le motif MIDAS sont surlignés en vert. La Thr357, qui pourrait potentiellement lier un ion métallique directement via un ligand induisant un changement de conformation, est surlignée en bleu. Les éléments de structure secondaire correspondant aux bras d'insertion sont indiqués en magenta et rouge; Notez leur absence dans la séquence de l'intégrine  $\alpha 2\beta 1$ . La région du bras d'insertion rouge montre le niveau le plus élevé de divergence (Izoré *et al.* 2010)

### I.1.2. Repléments de type IgG et CNA

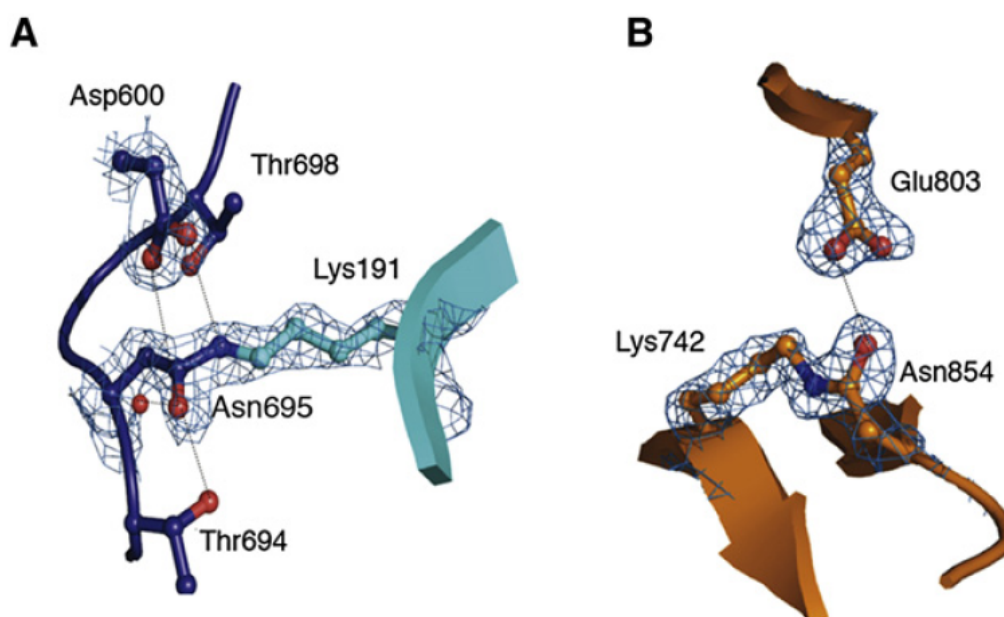
Le repliement des domaines D1 et D4 est similaire à celui des immunoglobulines (Fig.40). Le domaine D2 quant à lui est formé par l'association de deux séquences protéiques codées par des séquences éloignées l'une de l'autre (Fig.38) et montre une grande similarité de structure avec le domaine N2 de la région A (Cna\_A) présente dans l'adhésine Cna de *S. aureus* (Fig.40) (Deivanayagam 2000, Zong 2005). Par ailleurs, ce dernier domaine a été montré comme liant le collagène de type I, il est donc envisageable qu'une partie de la fonction de reconnaissance de RrgA pour les récepteurs cellulaires puisse être médiée par ce domaine D2.



**Figure 40: Repliement de type CNA et IgG.** Superposition du domaine D2 de RrgA (en bleu foncé et bleu clair) avec le domaine N2 de la région Cna\_A (en beige) de *S. aureus* (2F68). Superposition du domaine D4 de RrgA (en orange) avec le domaine IgG (en beige) de la protéine Gbs52 de *S. agalactiae* (2PZ4). Les déviations standards (RMSD : Root Mean Square deviation) sont indiquées. (Zong *et al.* 2005 ; Deivanayagam *et al.* 2000). (Izoré *et al.* 2010)

### I.1.3. Identification de deux ponts intramoléculeux dans RrgA

Deux ponts intramoléculeux covalents ont été identifiés dans la structure de RrgA (Fig.41). L'un, présent dans le domaine D4, est formé par la Lys742 et l'Asn854, cette liaison covalente étant surmontée du résidu Glu803. L'autre pont est inclus dans le domaine D2 et fait intervenir des résidus de chacune des deux régions le composant : Lys191, Asn695 et Asp600.



**Figure 41: Les ponts intramoléculaires de RrgA.** **A.** Le domaine D2 présente une liaison covalente entre les chaînes latérales des résidus Lys191 (bleu clair) et Asn695 (bleu foncé) qui est stabilisée par l'Asp600 et les groupements hydroxyles de deux Thréonines environnantes. **B.** Le pont intramoléculaire du domaine D4 est stabilisé par le Glu803 qui surmonte la liaison Lys742-Asn854. (Izoré *et al.* 2010)

Le rôle de ces résidus dans la formation des ponts covalents a été vérifié par mutagenèse (Tableau 12). La mesure des masses moléculaires des protéines mutantes par spectrométrie de masse a validée l'absence de formation des ponts. En effet, la formation de la liaison covalente entre une lysine et une asparagine se traduit par une masse inférieure de 17Da par rapport à la masse théorique, résultant de la perte d'un ion ammonium. Nous avons ainsi pu vérifié l'absence de formation d'un des deux ponts dans chacun des simples mutants et l'absence des deux ponts dans le double mutant (Tableau 7).

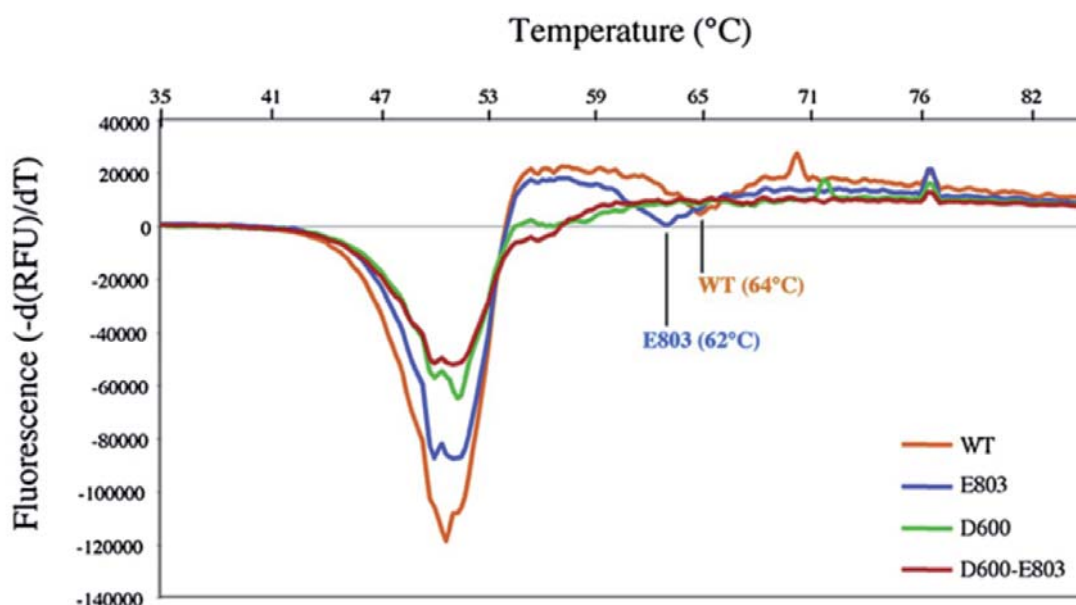
**Tableau 7 : Mesure des masses de RrgA et de ses variants par ESI-MS (ElectroSpray Ionization Mass Spectrometry Time-of-Flight) (Izoré *et al.* 2010)**

	CalculatedMass (Da)	MeasuredMass (Da)	Mass Difference (Calculated – Measured)	Number of Isopeptide Bonds Present
WT RrgA	94 197.5	94,163.7	33.8	2
E803A	94 139.5	94,123.8	15.7	1
K191A	94 140.4	94,124.8	15.6	1
N695A	94 154.5	94,139.3	15.2	1
D600A	94 153.5	94,139.3	14.2	1
D600A/E803A	94 095.5	94,096.0	-0.5	0

The presence of one isopeptide bond corresponds to the loss of one NH<sub>3</sub> group (17Da).

### I.1.4. Rôle de stabilisation des ponts intramoléculaires

Des mesures de stabilité thermique des protéines RrgA de type sauvage et des formes mutantes E803A, D600A et E803A/D600A ont été effectuées par Thermal Shift Assay ou TSA (Fig.42).



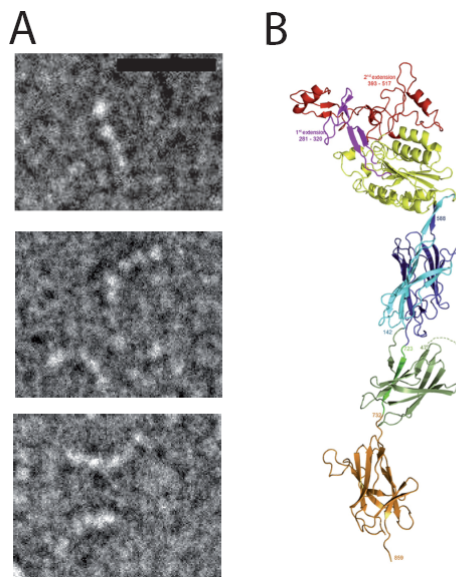
**Figure 42: Les Domaines D2 and D4 sont stabilisés par les ponts intramoléculaires.** Analyse de la thermostabilité de RrgA de type sauvage (en orange) et des formes mutantes E803A (en bleu), D600A (en vert) et E803A/D600A (en rouge) par TSA (Izoré *et al.* 2010).

La technique de TSA permet d'évaluer la stabilité thermique d'une protéine. La dénaturation de la chaîne polypeptidique par la chaleur conduit à l'exposition de zones hydrophobes sur lesquelles se fixe une sonde qui devient alors fluorescente. La courbe dérivée du niveau de fluorescence en fonction de la température permet de déterminer une valeur de température de fusion ( $T_m$ ) dont la valeur élevée (à partir environ de  $50^\circ\text{C}$ ) est un gage de la haute stabilité de la protéine considérée. Cette analyse révèle que la protéine RrgA de type sauvage présente deux maxima : le premier de  $51^\circ\text{C}$  ( $\pm 0.5^\circ\text{C}$ ), qui est également trouvé chez tous les mutants de RrgA, et un second pic, moins prononcé, dont la valeur de  $T_m$  est de  $64^\circ\text{C}$  ( $\pm 0.2^\circ\text{C}$ ). Etant donné que le pic majeur est présent dans les expériences avec toutes les formes de RrgA indépendamment de la présence ou de l'absence des ponts intramoléculaires, celui-ci pourrait représenter la dénaturation du domaine D3. Pour le variant RrgA E803A, qui ne présente plus de pont dans son domaine D4, la valeur du  $T_m$  du second pic est abaissé de  $2^\circ\text{C}$  par rapport à celle de la protéine RrgA sauvage. Cette observation suggère que le pic mineur correspond à la dénaturation des domaines plus petits (D1, D2 et D4). Dans les cas des variants RrgA

D600A et RrgA D600A/E803A, le second pic n'est pas observé. L'absence de ce second point de transition suggère que l'absence du pont intramoléculaire du domaine D2 engendre une déstabilisation de l'axe de la molécule RrgA formé par les domaines D1/D2/D4. Ces résultats indiquent donc que les ponts intramoléculaires identifiés chez RrgA ont un rôle important dans la stabilisation de la piline.

### I.1.5. Analyse de la protéine recombinante RrgA par microscopie électronique à transmission

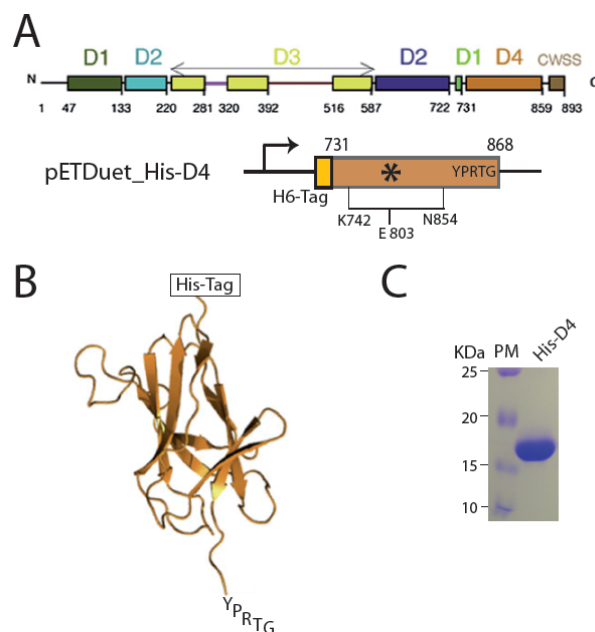
La protéine recombinante RrgA a été analysée par Microscopie Electronique à Transmission (MET) en coloration négative (Fig.43 A). La protéine se présente sous la forme d'un chapelet d'une longueur de 20nm et d'une épaisseur de 5 nm (Fig.43 A). Ces mesures sont en accord avec les données cristallographiques, soit, 19 nm de longueur et 7 nm d'épaisseur. Le chapelet visible sur les clichés de MET est formé par l'association de quatre petites perles correspondant aux quatre domaines composant RrgA (Fig.43 A et B). Enfin, l'adoption de conformations plus ou moins courbées suggère une certaine flexibilité inhérente à la structure modulaire de RrgA (Fig.43 A : clichés 2 et 3)



**Figure 43: Analyse de la protéine recombinante RrgA.** A. Clichés de microscopie électronique à transmission de la protéine RrgA monomérique en coloration négative. La barre d'échelle représente une distance de 20 nm. B. Structure cristallographique de RrgA.

### I.1.6. Production du domaine D4

Des travaux menés au cours de cette thèse dans le but d'étudier l'incorporation de RrgA dans le pilus ont nécessité la production d'une forme minimale de RrgA contenant le motif YPRTG. C'est la raison pour laquelle le domaine D4 a été produit indépendamment du reste de la protéine RrgA. La construction génétique a été conçue sur la base des données structurales. Le gène a été cloné dans le plasmide pETDuet de telle sorte à ce que la protéine résultante présente une étiquette poly-histidine à son extrémité N-terminale (Fig.44). La protéine recombinante His-D4 s'exprime de façon soluble. Son analyse par spectrométrie de masse révèle la présence du pont intramoléculaire K742-N854. En effet, sa masse moléculaire mesurée de 16 992 Da est en accord avec la perte d'un ion NH<sub>3</sub>, par rapport à sa masse théorique (17 009 Da).



**Figure 44: Production du domaine D4 recombinant.** **A.** Topologie de la protéine RrgA présentant ses différents domaines (en haut) ; Représentation schématique de la construction du domaine D4 dans le vecteur bi-cistronique pETDuet (en bas). La position de l'étiquette poly-histidine ainsi que les résidus impliqués dans la formation du pont intramoléculaire sont indiqués. **B.** Structure du domaine His-D4 recombinant. **C.** Protéine His-D4 purifiée sur colonne His-Muti Trap après avoir été surexprimée chez *E. coli* (SDS-PAGE, coloration au bleu de coomassie).



## **I.2 Etude de la piline accessoire RrgC**

### **I.2.1. Topologie de RrgC**

Une analyse bioinformatique de la séquence de la piline RrgC basée sur la recherche de séquences similaires (BLAST), d'alignement de séquences (CLUSTAL) et de recherche de domaines (Pfam) a révélée la présence de deux régions prédites en domaines CnaB (PF05738) dans la séquence de RrgC (Fig. 45).

REMARQUE : Il est important d'indiquer à ce stade du manuscrit que les domaines annotés CnaB et qui sont prédits par la banque de données Pfam présentent des topologies de type Immunoglobuline. A titre indicatif, nous annoterons dans la suite du manuscrit les domaines CnaB prédits : CnaB (PF05738). La structure cristallographique de RrgA a montré qu'un pont intramoléculaire est localisé dans le domaine D4 de la protéine, dont la structure est de type IgG. Or, l'analyse Pfam de RrgA présente ce domaine IgG comme un domaine CnaB (PF05738). Conscient de cette incohérence, nous avons néanmoins utilisé les prédictions Pfam CnaB (PF05738) pour réaliser l'analyse des alignements de séquences présentés dans la figure 45.

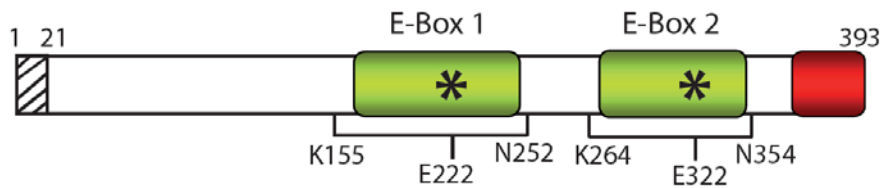
Celle-ci indique la présence de résidus Lysine et Glutamate hautement conservés, qui dans les structures cristallographique de Gbs52 et RrgA, sont associés à la formation de liaisons intramoléculaires. Les résidus Lysine et Glutamate sont retrouvés dans des motifs KVD et GxYxLxExAPxGY (E-Box) respectivement, alors que le résidu Asparagine est localisé 25 à 30 résidus en aval du Glutamate. La première information importante qui est mise en avant par cet alignement est l'existence de motifs E-Box dans les séquences de pilines mineures, qui était jusqu'à lors utilisés comme des éléments de reconnaissance des pilines majeures et dont la fonction était inconnue. La seconde information est que le Glutamate hautement conservé de ces motifs E-box est associé à la formation des ponts intramoléculaire comme identifiés pour RrgA et Gbs52.



L'ensemble de ces informations permet de proposer la présence de deux ponts intramoléculaires dans RrgC, localisés dans chacun des domaines CnaB (PF05738) : CnaB1 et CnaB2 (Fig. 46) chacun associé à une E-Box : E-Box 1 et E-Box 2. Les résidus potentiellement impliqués dans la formation de ces ponts sont les suivants :

CnaB1 : Lys155, Glu222 et Asn252

CnaB2 : Lys264, Glu322 et Asn354



**Figure 46: Les ponts intramoléculaires de RrgC.** Topologie de la protéine RrgC indiquant la position des ponts intramoléculaires. (El Mortaji *et al.* 2010)

## I.2.2. Mise en évidence des ponts intramoléculaires de RrgC

La prédiction de la présence de deux liaisons covalentes dans RrgC et la proposition des résidus impliqués dans la formation de ces ponts nécessitait une démonstration expérimentale. Une forme recombinante de RrgC (résidus 22-368) qui présente une étiquette poly-histidine non clivable à son extrémité N-Terminale (pACYCDuet\_H6-RrgC) a été produite et son analyse par ESI-MS a confirmé la présence de deux ponts intramoléculaires (Tableau 8). La masse moléculaire observée (40 948 Da) est en accord avec une perte de deux ions ammonium (par comparaison avec la masse théorique de 40 983 Da). Chacun des résidus identifiés comme jouant un rôle dans la formation des ponts a été muté en alanine ; les mutants simples ainsi que le double mutant E222A/E322A ont été produits et purifiés de la même manière que la protéine native. Les mesures obtenues par spectrométrie de masse confirment la perte d'un seul NH<sub>3</sub> pour chacun des mutants individuels, alors que le double mutant ne présente plus de différence de masse en comparaison à sa masse théorique. Ces données confirment que chacun des résidus identifiés, Lys155, Glu222, N252, Lys264, Glu322 et N354, joue un rôle dans la formation des ponts intramoléculaires de la piline accessoire RrgC.

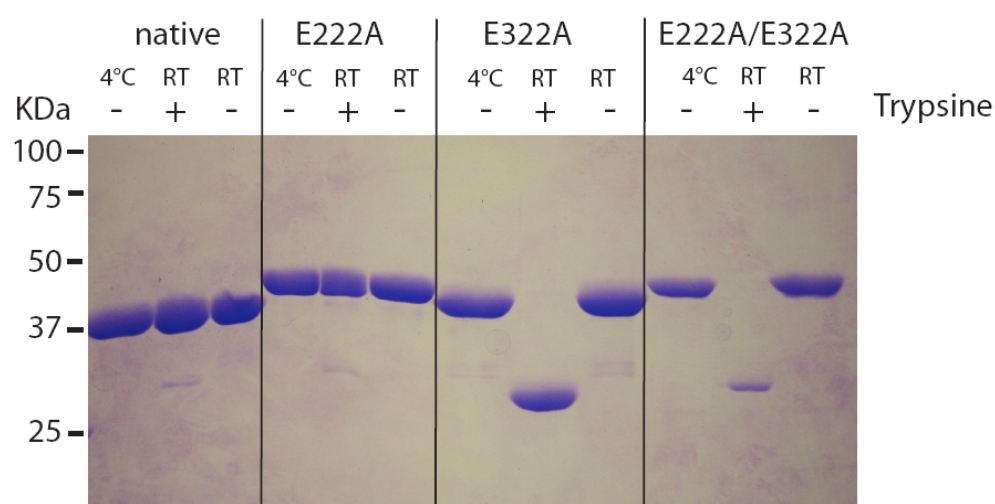
**Tableau 8: Analyse par ESI-MS de la protéine recombinante native et de ses mutants.** (El Mortaji *et al.* 2010)

His_RrgC	Masse moléculaire moyenne (Da)		$\Delta (M_{théo} - M_{obs})$ (Da)	Nombre de ponts intramoléculaires
	Théorique	Observée par ESI-TOF		
native	40 983	40 948	35	2
K155A	40 926	40 908	18	1
E222A	40 925	40 907	18	1
N252A	40 940	40 922	18	1
K264A	40 926	40 908	18	1
E322A	40 925	40 907	18	1
N354A	40 940	40 922	18	1
E222A/E322A	40 867	40 867	0	0

### I.2.3. Les deux ponts intramoléculaires stabilisent RrgC

#### I.2.3.1 Sensibilité à la digestion protéolytique

Les formes natives et mutantes de RrgC ont subi des tests de digestion à la trypsine afin d'analyser leur résistance à la protéolyse (Fig. 47). Ni la forme native, ni le mutant E222A ne semblent affectés par la digestion tryptique, alors que le mutant E322A et le double mutant E222A/E322A sont dégradés en un produit de masse apparente d'environ 30kDa (Fig. 47, ligne 8 et 11). Il est intéressant de noter qu'en absence de trypsine, les trois variants, auxquels il manque un ou deux ponts intramoléculaires, migrent sur SDS-PAGE un peu plus lentement que la protéine native. Ceci suggère que les ponts intramoléculaires peuvent aider à la formation d'une structure plus compacte de la protéine (Fig. 47).



**Figure 47: Stabilité protéolytique de RrgC.** Produits de digestion tryptique séparés par SDS-PAGE 12,5%. Le marqueur de poids moléculaire est indiqué à gauche du gel. Les conditions d'ajout ou non de trypsine sont indiquées au dessus du gel ainsi que les températures de réaction (RT : Température ambiante). (El Mortaji *et al.* 2010)

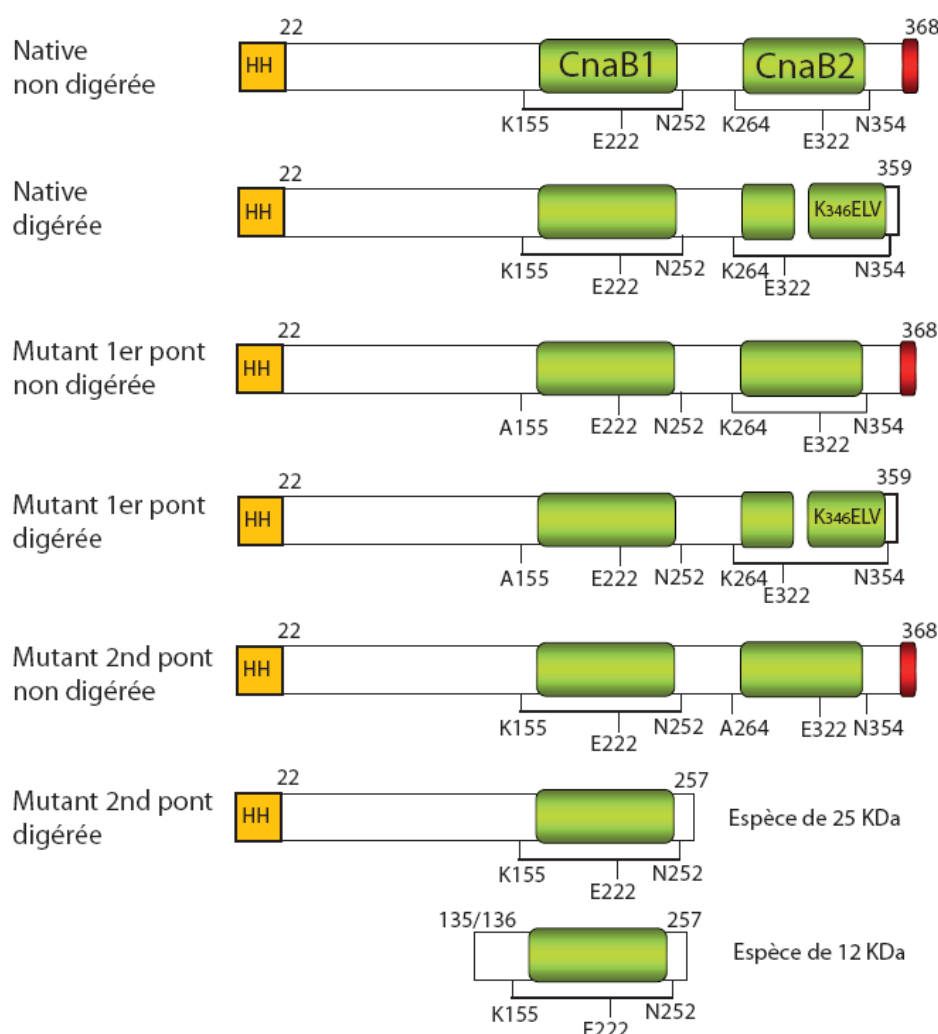
Une caractérisation plus précise des produits de digestion de RrgC a été réalisée par spectrométrie de masse et séquençage N-terminal (Tableau 9).

**Tableau 9: Masses moléculaires et séquences N-terminales des variants RrgC natif et mutants digérés à la trypsine.** Les masses calculées prennent en compte la présence d'un ou deux ponts intramoléculaires en fonction du variant considéré. (El Mortaji *et al.* 2010)

	Non digéré			Digéré	
	Masses calculées	Masses mesurées	Séquences N-terminales	Masses mesurées	Séquences N-terminales
<b>native</b>	40 949	40 948	GSSHHHH	40 010	GSSHHHH KELVTVV
<b>K155A</b>	40 909	40 908	GSSHHHH	39 970	GSSHHHH KELVTVV
<b>E222A</b>	40 908	40 907	GSSHHHH	39 969	GSSHHHH KELVTVV VEPLAGY ELVTVVK
<b>N252A</b>	40 923	40 922	GSSHHHH	39 984	GSSHHHH KELVTVV
<b>K264A</b>	40 909	40 908	GSSHHHH	28 551 12 893 12 765	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>
<b>E322A</b>	40 908	40 907	GSSHHHH	28 551 12 893 12 765	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>
<b>N354A</b>	40 923	40 922	GSSHHHH	28 551 12 893 12 765	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>

<sup>a</sup>ND, non déterminé

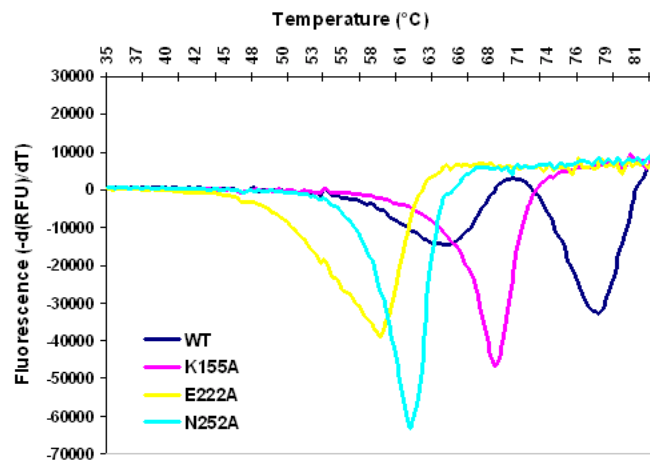
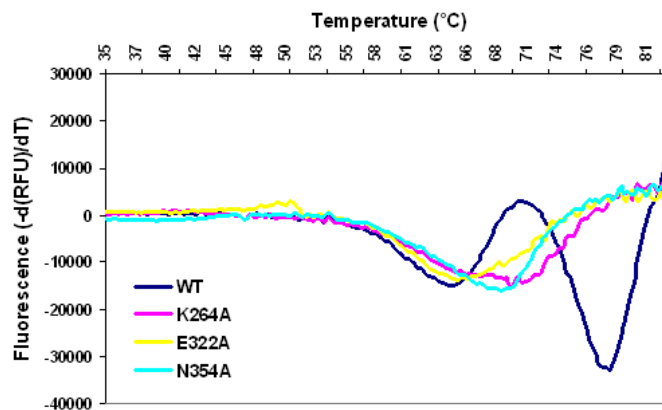
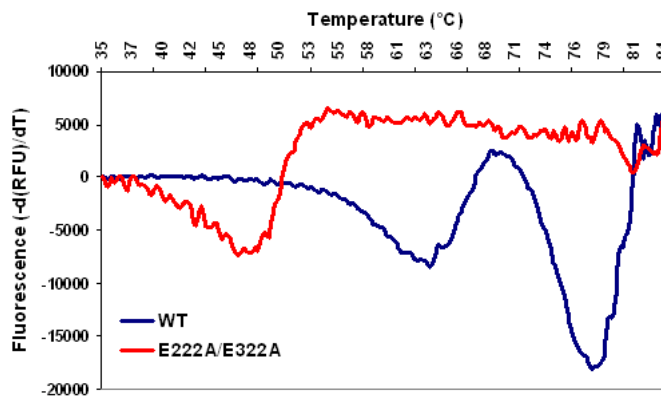
L'interprétation de ces données est schématisée dans la figure 48. Pour la protéine native RrgC et les trois variants du premier pont intramoléculaire, un clivage par la trypsine a lieu à l'intérieur du domaine CnaB2, à la position <sup>346</sup>KELVT, sans que les deux chaînes polypeptidiques ne soient dissociées (Fig. 48), indiquant que le pont intramoléculaire associé au CnaB2 est capable de maintenir les deux fragments (Fig. 48). Chez le mutant du second pont, la protéolyse induit la digestion totale du domaine CnaB2 (espèce de 25 kDa) et de la région N-terminale de RrgC (espèce 12 kDa). Ces résultats suggèrent donc que le second pont intramoléculaire joue un rôle clé dans la stabilisation du domaine CnaB2, mais également qu'il influence la stabilité globale de RrgC (Fig. 48)



**Figure 48:** Schéma récapitulatif des données obtenues lors des expériences de digestions par la trypsine. (El Mortaji *et al.* 2010)

### I.2.3.2 Stabilité thermique

Afin de caractériser d'avantage le rôle des ponts intramoléculaires dans la stabilité de RrgC, nous avons procédé à des analyses par Thermal Shift Assay (TSA) des différents variants de RrgC (Fig. 49). La protéine native RrgC présente deux valeurs de  $T_m$  distinctes :  $63.7^\circ\text{C}$  ( $\pm 0.6$ ) et  $77.5^\circ\text{C}$  ( $\pm 0.5$ ). En revanche, toutes les formes mutantes portant une simple mutation, et donc dépourvues d'un des deux ponts, ne montrent qu'une seule valeur de  $T_m$  : K155A ( $68.3^\circ\text{C}$  ;  $\pm 0.6$ ), E222A ( $57.4^\circ\text{C}$  ;  $\pm 1.3$ ), N252A ( $61^\circ\text{C}$  ;  $\pm 1$ ), K246A ( $66^\circ\text{C}$  ;  $\pm 2.4$ ), E322A ( $65^\circ\text{C}$  ;  $\pm 0.8$ ), N354A ( $67^\circ\text{C}$  ;  $\pm 2.7$ ) (Fig. 49 A et B). Le double mutant E222A/E322A qui ne présente plus aucun pont intramoléculaire montre une température de fusion dramatiquement diminuée d'une valeur de  $47.9^\circ\text{C}$  ( $\pm 2.1$ ) (Fig. 49 C). En conclusion ces données indiquent que les ponts intramoléculaires influencent la stabilité thermique de RrgC.

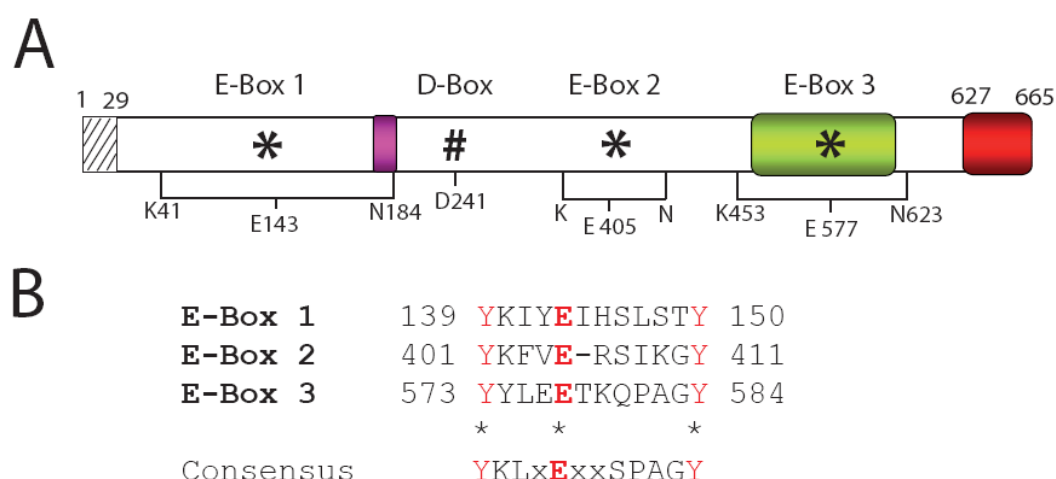
**A****B****C**

**Figure 49: Thermostabilité de RrgC et de ses variants.** Les analyses ont été réalisées par TSA. Les expériences ont toutes été réalisées au moins trois fois de façon indépendante. Les graphiques proviennent d'une expérience représentative. La courbe correspondant à la protéine native a été incluse dans chacune des figures afin de faciliter la comparaison avec les différents variants de RrgC. **A.** Thermostabilité de la protéine native et des mutants du premier pont. **B.** Thermostabilité de la protéine native et des mutants du second pont. **C.** Thermostabilité de la protéine native et du double mutant Glutamate. (El Mortaji *et al.* 2010)

## I.3 Etude de la piline majeure RrgB

### I.3.1. Topologie de RrgB

L'analyse par ESI-MS (ElectroSpray Ionization Mass Spectrometry) de la protéine recombinante His-RrgB (produite à partir de la construction pLIM01\_RrgB) indique une masse de 67 400.51 Da soit 52.6 Daltons de moins que sa masse théorique (67 453.11 Da). Ce résultat laisse donc penser que la protéine renfermerait trois ponts intramoléculaires, la formation de chacun d'eux entraînant une perte de masse de 17 Da. L'identification des résidus formant les ponts dans RrgB s'est basée sur les données obtenues avec RrgA et RrgC. En effet, dans ces pilines accessoires, le glutamate impliqué dans la formation et/ou la stabilisation des ponts intramoléculaires correspond à celui identifié dans la séquence conservée E-Box, GxYxLxExAPxGY, le résidu lysine est inclus dans la séquence KVD et l'asparagine est localisée 25 à 30 résidus en aval du glutamate. Nous avons donc recherché la présence de motifs E-Box dans la séquence de la piline majeure RrgB. L'analyse Pfam de la protéine RrgB a mis en évidence la présence d'un domaine CnaB (PF05738) dans lequel est trouvé un motif E-Box (E-box 3) (Fig.50 A). Par ailleurs, en analysant « manuellement » la séquence en acides aminés de RrgB, deux motifs E-Box supplémentaires ont été identifiés, E-Box 1 et E-Box 2 (Fig.50 B). L'alignement de la séquence de RrgB avec d'autres domaines CnaB (PF05738) de pilines mineures ainsi qu'avec la piline majeure Spy128 a permis d'identifier les résidus K41-N184 et K453-N623 potentiellement impliqués dans les ponts intramoléculaires associés à respectivement aux E-Box 1 et E-Box 3 (Fig.50 A).



**Figure 50: Les ponts intramoléculaires de RrgB.** **A.** Topologie de la protéine RrgB indiquant la position des ponts intramoléculaires. **B.** Alignement des motifs E-box identifiés chez RrgB. (El Mortaji *et al.* 2010)



### I.3.2. Mise en évidence des ponts intramoléculaires de RrgB

Les résidus glutamate de chacun des trois motifs E-Box identifiés ont été mutés en alanine, les protéines mutées ont été produites, purifiées et analysées par ESI-MS (Tableau 10). Les résultats obtenus indiquent que chacun des mutants individuels E405A, E577A et N623A présentent une perte de masse équivalente à un ion NH<sub>3</sub> validant ainsi leur implication dans un pont intramoléculaire. En revanche, la mutation des résidus K41A, E143A et N184A n'a aucun effet puisque ces mutants possèdent, comme la protéine native, trois ponts intramoléculaires. Ce résultat est toutefois surprenant si l'on tient compte de la conservation de ces résidus dans la séquence de RrgB. Nous considérons néanmoins qu'un pont est potentiellement établi dans la E-box 1 mettant en jeu les résidus K41A, E143A et N184A, mais que pour une raison encore inconnue, cette liaison n'est pas formée dans la protéine RrgB recombinante en solution. Il n'en reste pas moins qu'il demeure dans RrgB un pont intramoléculaire non identifié. Nous avons donc recherché la présence d'un aspartate, résidu qui joue un rôle similaire dans la formation du pont intramoléculaire du domaine D2 de RrgA. L'alignement des séquences des pilines RrgA, RrgB et de leurs homologues chez d'autres bactéries Gram-positif a permis de définir une séquence consensus qualifiée de D-Box (Fig.51). Dans RrgB, le résidu D241 est ainsi proposé comme impliqué dans la formation d'un pont intramoléculaire. Afin de valider l'implication de ce résidu, la protéine mutée RrgB<sub>D241A</sub> a été produite et purifiée, son analyse par ESI-MS a confirmé le rôle de ce résidu dans la formation d'un pont intramoléculaire (Tableau 10). L'analyse ESI-MS du triple mutant D241A/E405A/E577A indique que cette forme variante ne présente plus aucun lien intramoléculaire, confirmant définitivement l'identification de tous les résidus impliqués dans la formation des liaisons isopeptidiques de RrgB.

**Tableau 10: Analyse des spectres de masse ESI-TOF de la protéine recombinante native et de ses mutants.** (El Mortaji et al. 2010)

	Masse moléculaire moyenne (Da)		$\Delta (M_{\text{théo}} - M_{\text{obs}})$ (Da)	Nombre de ponts intramoléculaires
	Théorique	Observée par ESI-TOF		
His-RrgB native	67 453.11	67 400.51	52.60	3
K41A	67 396.00	67 343.52	52.48	3
E143A	67 395.07	67 342.74	52.33	3
N184A	67 410.00	67 357.53	52.47	3
D241A	67 409.10	67 375.51	33.59	2
E405A	67 395.07	67 361.38	33.69	2
E577A	67 395.07	67 361.84	33.23	2
N623A	67 410.00	67 374.38	35.62	2
D241A/E405A/E577A	67 293.03	67 293.78	-0.75	0

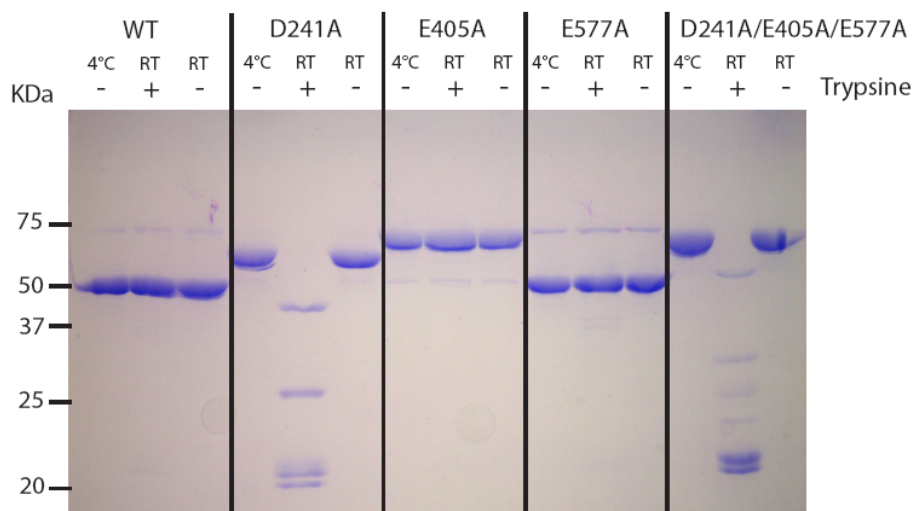
<i>S. pyogenes</i>	Q1J8U5	266	QYQTLRWED	MTKGLTYKAG	285
<i>S. suis</i>	B5UAE3	235	AYKTFPWE	TMLAGLDFKLN	254
<b><i>S. pneumoniae</i></b>	<b>RrgB</b>	233	NYATANWSD	RMTEGLAFNKG	252
<i>F. magna</i>	C2HF05	293	KLKTAYWSD	EMTEGLQYNN	312
<i>S. agalactiae</i>	B9UR22	239	DYKCLIWTD	SMTKGLTFNND	250
<b><i>S. pneumoniae</i></b>	<b>RrgA</b>	592	SIENGTITD	PMGELIDLQLG	611
<i>C. perfringens</i>	B1RRE2	437	SINNGTVID	PMGDMVDI IKS	456
<i>S. pyogenes</i>	Q1JIZ5	589	SIVDGNVTD	PMGEMIEFQLK	608
<i>S. agalactiae</i>	B8RD39	589	SIVDGNVTD	PMGEMIEFQLK	608
<i>E. faecalis</i>	C7VQW6	383	TIHEGLIID	PMGENILLQGA	402
			* * :		
Consensus			SIKTGNWTD	PMGEGLDFQLG	

**Figure 51: Identification de la « D-Box ».** Alignement partiel des pilines RrgA, RrgB et de leurs homologues chez d'autres bactéries Gram-positif montrant la D-Box. (El Mortaji *et al.* 2010)

### I.3.3. Les trois ponts intramoléculaires stabilisent RrgB

#### I.3.3.1 Sensibilité à la digestion protéolytique

Nous avons évalué le rôle des trois ponts intramoléculaires dans la stabilisation de RrgB en analysant la susceptibilité protéolytique des différents variants de RrgB (Fig.52). Ni la protéine native, ni les mutants E405A et E577A ne sont affectés dans leur résistance à la digestion tryptique, alors que le mutant D241A et le triple mutant D241A/E405A/E577A sont clivés en de multiples fragments allant de 20 à 50 kDa (Fig.52, ligne 5 et 14). Ces résultats indiquent donc que la seule mutation D241A induit une déstabilisation majeure de RrgB. Il est intéressant de noter qu'en absence de protéase, les variants D241A, E405A et le triple mutant migrent sur SDS-PAGE un peu plus lentement que la protéine native indiquant que, comme pour RrgC, les ponts isoéptidiques augmentent le degré de compaction de RrgB (Fig.52).

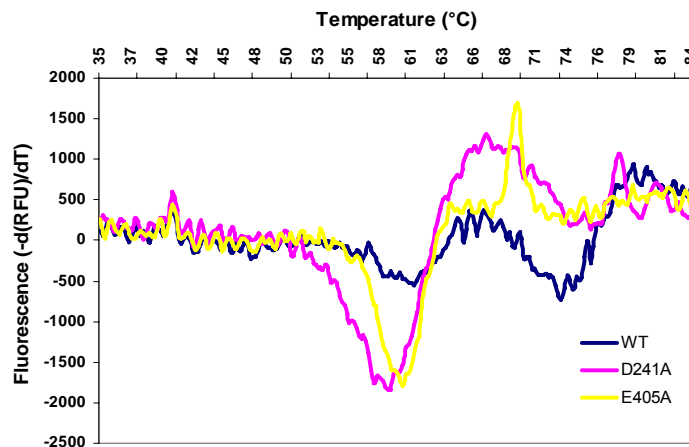


**Figure 52: Stabilité protéolytique de RrgB.** Les produits de digestion tryptique ont été séparés par SDS-PAGE 12,5%. Le marqueur de poids moléculaire est indiqué à gauche du gel. Les conditions d'ajout ou non de trypsine sont indiquées au dessus du gel ainsi que les températures de réaction (RT : Température ambiante) (El Mortaji *et al.* 2010)

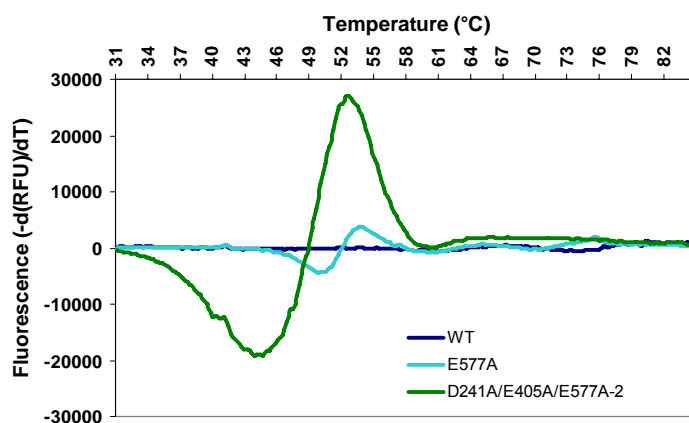
### I.3.3.2 Stabilité thermique

Les analyses réalisées par TSA montrent qu'aucune dénaturation thermique significative n'est détectable dans la forme native de RrgB, preuve de la nature extrêmement stable de la protéine (Fig.53). Au contraire, les mutants D241A et E405A présentent des valeurs respectives de  $T_m$  de  $58.5^{\circ}\text{C}$  ( $\pm 0.5$ ) et  $58.9^{\circ}\text{C}$  ( $\pm 2.5$ ), (Fig.53 A). Le mutant E577A présente un défaut de stabilité prononcé se traduisant par un profil de dénaturation en trois états, caractérisés par des valeurs de  $T_m$  de  $48.6^{\circ}\text{C}$  ( $\pm 2.1$ ),  $59.2^{\circ}\text{C}$  ( $\pm 2.0$ ) et  $68.8^{\circ}\text{C}$  ( $\pm 1.9$ ) (Fig.53 B). Enfin, le triple mutant RrgB qui ne présente plus aucun pont intramoléculaire montre une stabilité thermique fortement diminuée qui se traduit par une valeur de température de fusion de  $43.1^{\circ}\text{C}$  ( $\pm 1.1$ ) (Fig. 53 B).

**A**



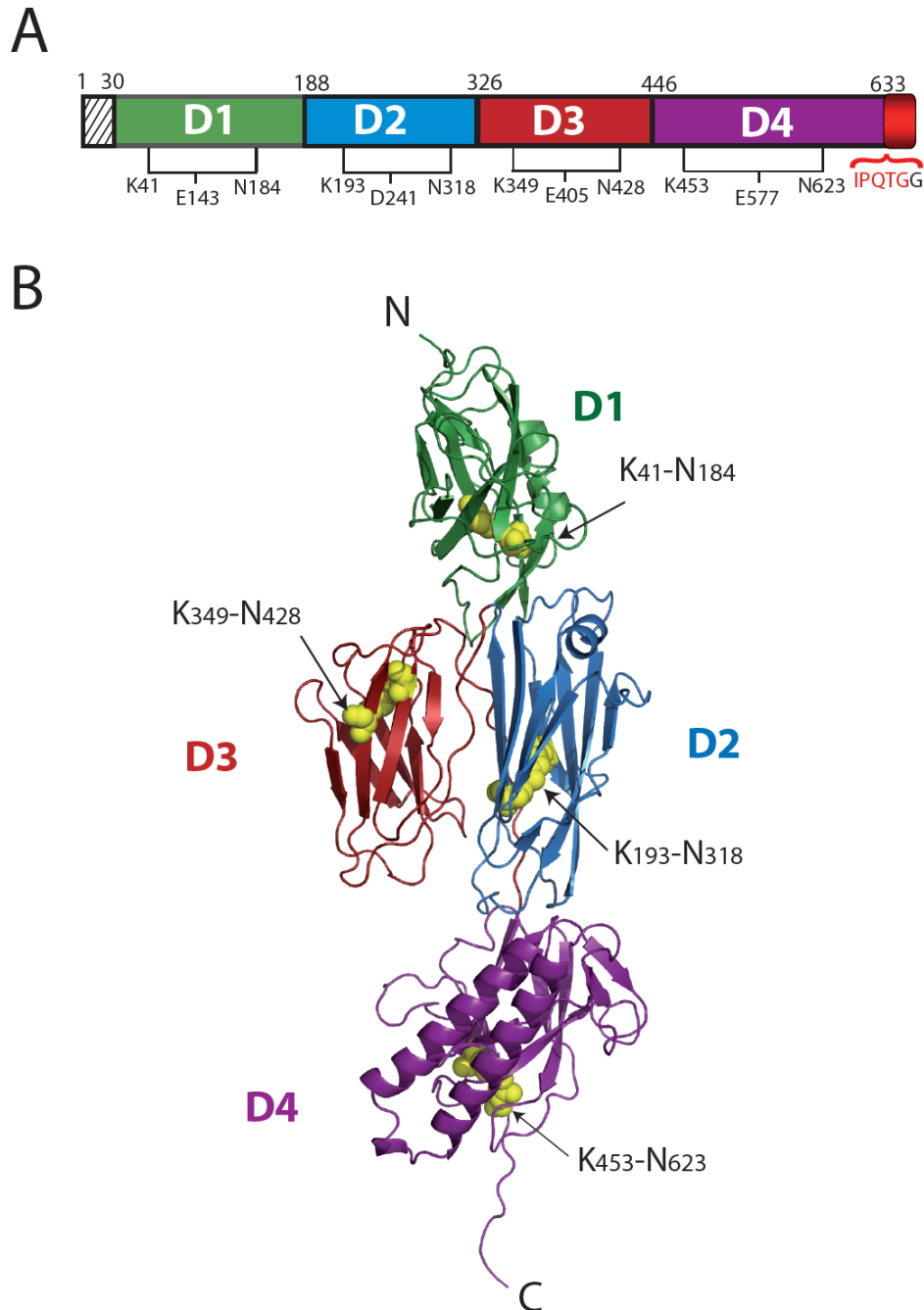
**B**



**Figure 53: Thermostabilité de RrgB et de ses variants.** Les analyses ont été réalisées par TSA. Les expériences ont toutes été réalisées au moins trois fois de façon indépendante. Les graphiques proviennent d'une expérience représentative La courbe correspondant à la protéine native a été incluse dans chacune des figures afin de faciliter la comparaison avec les différents variants de ponts de RrgB. **A.** Thermostabilité de la protéine native et des mutants D241A et E405A. **B.** Thermostabilité de la protéine native, du mutant E577A et du triple mutant D241A/E405A/E577A. (El Mortaji *et al.* 2010)

### I.3.4. Structure cristallographique de la piline majeure RrgB

La structure cristallographique de la forme complète de RrgB a été résolue à 2.9Å par Carlos Contreras-Martel par remplacement moléculaire à partir des coordonnées d'une structure de RrgB tronquée (Fig. 54) (Spraggon 2010).

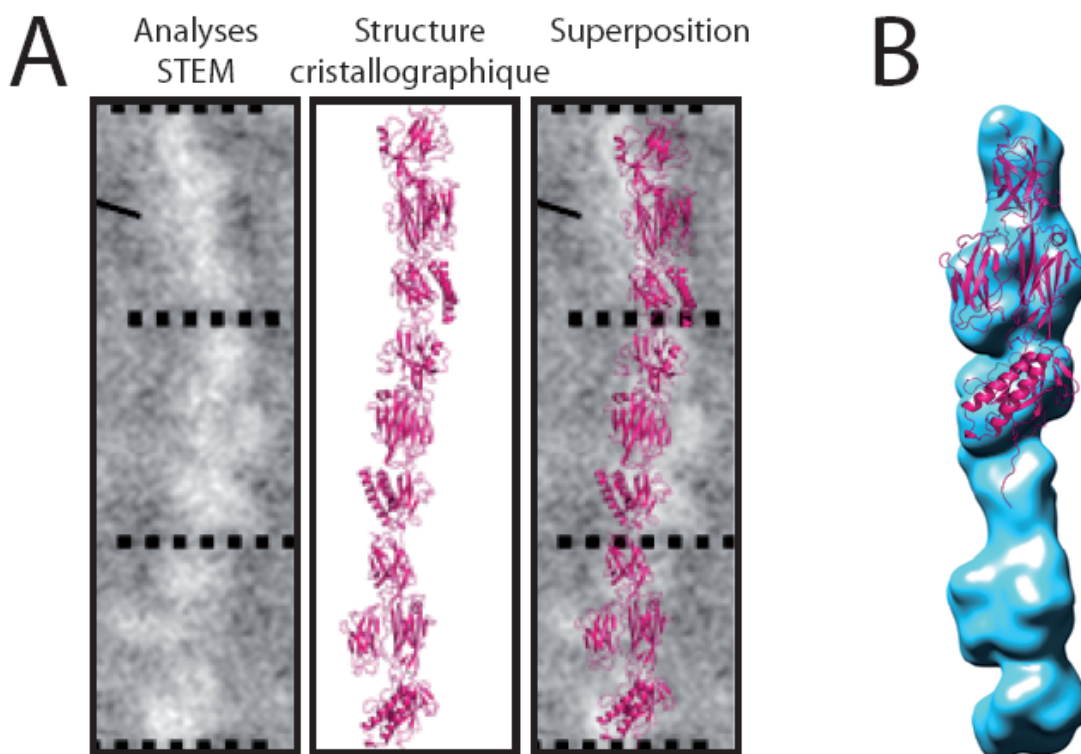


**Figure 54: Structure cristallographique de RrgB.** **A.** Représentation schématique de la protéine RrgB. Les différents domaines de la protéine sont représentés (D1 en vert, D2 en bleu, D3 en rouge et D4 en violet). Les positions du peptide signal, des ponts intramoléculaires et du CWSS sont indiqués. **B.** Représentation en ruban de la structure complète de RrgB. les différents domaines sont représentés suivant le même code couleur que **A.** Les ponts intramoléculaires sont représentés par des sphères en jaune.

La protéine RrgB est une molécule composée par quatre domaines, D1, D2, D3 et D4 qui présentent tous une topologie de type IgG (Fig.54 A et B). Les domaines D1, D2 et D4 sont alignés sur un même axe vertical alors que le domaine D3 est localisé sur le même axe horizontal que le domaine D2 (Fig.54 B). Le domaine D4 présente à son extrémité C-terminale un segment linéaire qui porte le motif IPQTG.

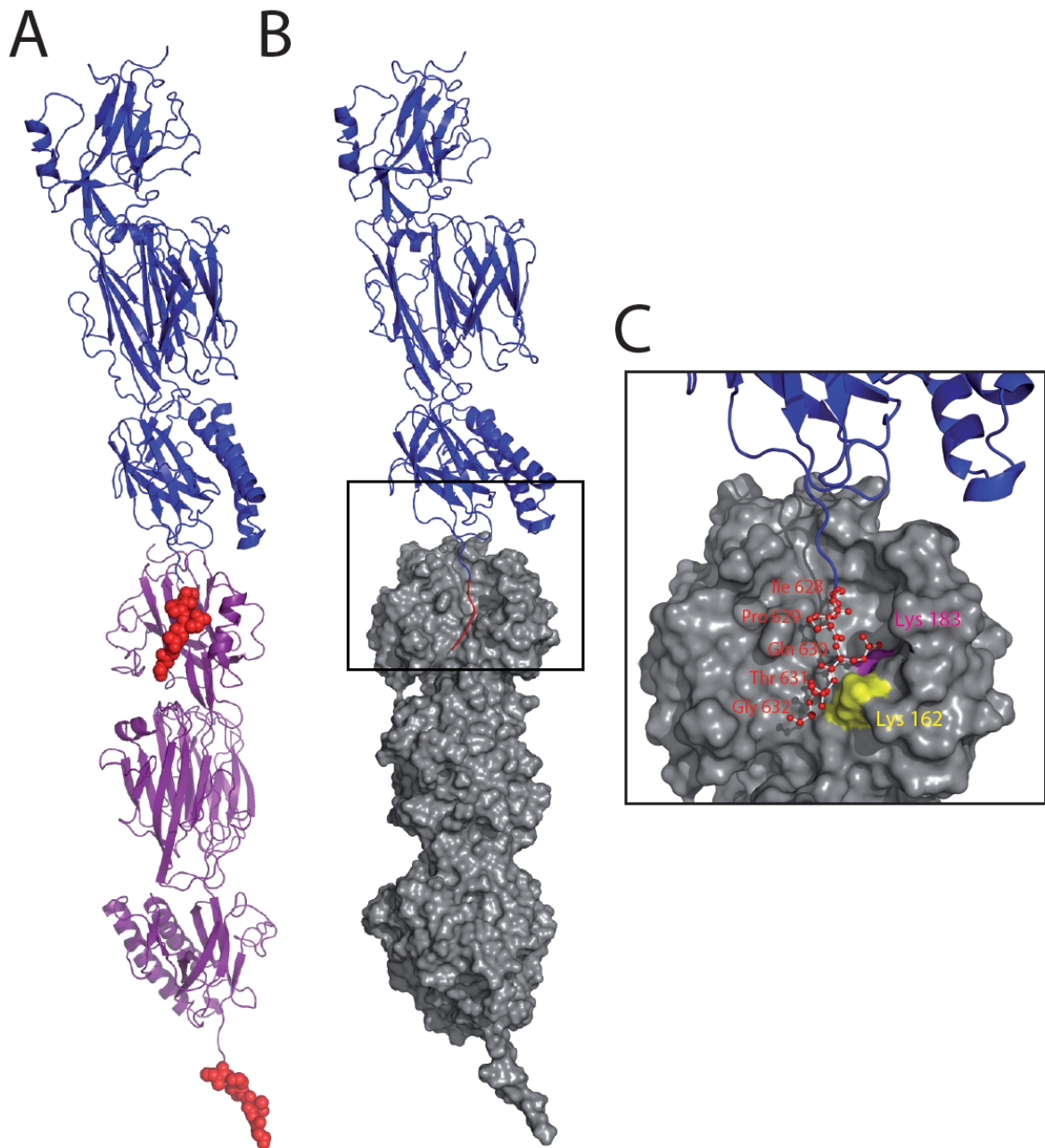
La structure cristallographique de RrgB confirme que chacun des domaines D2, D3 et D4 est stabilisé par un pont intramoléculaire qui fait intervenir les résidus précédemment identifiés par alignement de séquences (Fig.54 A et B). En effet, dans le domaine D2, les résidus Lys193 et Asn318 sont covalentement liés l'un à l'autre et le résidu Asp241 (D-Box) est positionné juste en face de cette liaison. Les domaines D3 et D4 sont stabilisés respectivement par les ponts intramoléculaires Lys349-N428 et Lys453-Asn623, associés aux glutamates Glu405 (E-Box 2) et Glu577 (E-Box 3). De façon très surprenante, et contrairement aux données obtenues au cours de l'étude de la protéine en solution, un pont intramoléculaire supplémentaire est parfaitement visible au niveau du domaine D1. Celui-ci fait intervenir les résidus qui avaient été identifiés auparavant par alignement, à savoir, la Lys41 et l'Asn184 ainsi que le Glu143 de la E-Box 1. La raison pour laquelle ce pont n'est visible que dans la forme cristalline et pas dans la protéine en solution reste encore non déterminée.

L'unité asymétrique du cristal contient trois monomères de RrgB. Dans l'empilement cristallin, chacune de ces molécules est incorporée dans une fibre artificielle qui pourraient représenter un modèle de l'assemblage du polymère de RrgB (Fig.55 A). En effet, chaque molécule est empilée en « tête à queue » avec la molécule suivante formant un angle de rotation approximatif de 120° entre deux monomères successifs. En se référant aux clichés de STEM (Scanning Transmission Electron Microscopy) obtenus par Hillerigmann en 2009 à partir de pili natifs purifiés, la fibre artificielle formée dans l'empilement cristallin se superpose assez bien avec son homologue natif. Néanmoins, en utilisant la reconstruction tridimensionnelle des pili natifs, seule la structure cristallographique d'un monomère peut être incorporée dans la surface du modèle reconstruit car la rotation entre deux molécules successives est de 20° environ contre les 120° mesurés à partir de la fibre artificielle trouvée dans l'empilement cristallin (Fig.55 B).



**Figure 55: Comparaison des fibres natives et artificielles de RrgB.** **A.** Superposition des fibres natives et artificielles. Le premier panneau correspond aux analyses STEM (Scanning Transmission Electron Microscopy) des pili natifs purifiés (issu de Hilleringmann *et al.* 2009). Les pointillés permettent de séparer chaque monomère de RrgB. Le deuxième panneau correspond à la structure cristallographique d'une fibre artificielle de RrgB trouvée dans l'empilement cristallin. Le troisième panneau correspond à la superposition manuelle des deux panneaux précédents. **B.** Semi-concordance de la structure cristallographique de RrgB dans la reconstitution tridimensionnelle des pili de *S. pneumoniae*. La reconstruction 3D du pilus est représentée en bleu (issu de Spraggon *et al.* 2010) et la structure cristallographique de RrgB est en rose.

L'analyse de l'empilement cristallin montre que le motif de reconnaissance IPQTG d'une première molécule RrgB se situe dans une gorge du domaine D1 d'une autre molécule RrgB (Fig.56 B et C). Selon le mécanisme de transpeptidation, la thréonine du motif IPQTG du premier monomère est liée covalentement à une lysine du monomère suivant. Dans la gorge du domaine D1 dans laquelle plonge le motif IPQTG, deux résidus lysines sont présents à proximité du résidu thréonine, les Lys162 et Lys183, distants du résidu thréonine respectivement de 3.6Å et 9.8Å. Même si la Lys183 fait partie du motif piline, et donc potentiellement impliquée dans la liaison intermoléculaire entre deux molécules RrgB, les données cristallographiques permettent de proposer la Lys162 comme groupement nucléophile le plus favorablement situé pour être impliqué dans la liaison peptidique (Fig.56 C).



**Figure 56: Assemblage de RrgB dans la fibre artificielle.** **A.** Assemblage de deux molécules successives de RrgB dans l'empilement cristallin. La première molécule de RrgB (RrgB1) est représentée en bleu et la molécule suivante (RrgB2) en violet. Dans les deux cas, le motif IPQTG est représenté par des sphères rouges. **B.** Contacts intermoléculaires de deux molécules successives de RrgB dans l'empilement cristallin. RrgB1 est colorée en bleu et son motif IPQTG est en rouge. RrgB2 est en représentation de surface. **C.** Vue élargie de la zone de contact intermoléculaire. Les résidus du motif IPQTG sont représentés en balles et bâtons. Les Lysines accessibles de RrgB2 sont indiquées en jaune et rose.

## I.4 Conclusions concernant l'étude individuelle des pilines

La structure cristallographique de RrgA nous a apporté des éléments clés dans la compréhension de sa fonction d'adhésine et également dans la compréhension de sa stabilité. Cette molécule de près de 900 acides aminés présente un repliement tout à fait original qui lui permet de se structurer sous la forme de quatre domaines modulaires interconnectés lui conférant ainsi une flexibilité certaine. Le domaine D3, présent à l'une des extrémités de RrgA, se compose d'éléments d'origine eucaryote (domaine de VWA avec un site MIDAS et un tripeptide RGD) connus pour médier l'adhérence cellulaire, dans lesquels sont insérés des éléments d'origine procaryote. Le domaine D2 connecté au domaine D3 est, quant à lui, le résultat de l'assemblage de deux demi-domaines N et C-terminaux qui sont physiquement verrouillés dans cette conformation grâce à la présence d'un pont intramoléculaire. La structure globale de D2, ainsi obtenue, présente une homologie structurale avec le domaine N2 d'une protéine connue comme interagissant avec le collagène 1: la région A de la protéine Cna de *S. aureus*. Enfin, les deux derniers domaines de RrgA, D1 et D4, ont des topologies de type IgG avec une particularité supplémentaire pour le domaine D4: un pont intramoléculaire. En conclusion, RrgA présente à l'une de ses extrémités deux domaines, D3 et D2, qui pourraient potentiellement interagir avec des éléments de l'hôte et donc représenter les éléments adhésifs de la protéine, ces deux domaines étant exposés dans des conditions optimales par la formation d'une « tige » formée par les domaines D1 et D4. De plus, la présence de deux ponts intramoléculaires dans la structure de RrgA explique le caractère particulièrement résistant de cette piline comme le prouvent les analyses de thermostabilité menées sur RrgA.

La structure cristallographique de la piline accessoire RrgC n'a pas été obtenue durant cette étude. Néanmoins, les expérimentations menées sur la piline recombinante HisRrgC, notamment par spectrométrie de masse, ont indiqué la présence de deux ponts intramoléculaires au sein de la protéine. Ainsi, une analyse complémentaire, basée sur des prédictions de domaines et des alignements de séquences, a permis d'identifier les résidus impliqués dans la formation de ces ponts. Nous avons ensuite validé le rôle de chacun des résidus identifiés dans la formation des ponts et avons ainsi mis en évidence un rôle pour le motif « E-Box », connu depuis de nombreuses années mais dont la fonction n'avait pas encore été identifiée. En effet, le glutamate hautement conservé des motifs « E-Box » est impliqué dans la formation de ponts intramoléculaires. Enfin, la fonction stabilisante de ces ponts, au



sein des pilines, a été prouvée par des tests de thermostabilité et de digestion protéolytique menées sur RrgC et ses variants dépourvus de ponts.

L'analyse individuelle de la piline majeure RrgB a permis d'identifier la présence de ponts intramoléculaires au sein de la protéine. Le rôle des motifs « E-Box », déterminé par l'étude menée sur RrgC, a permis d'identifier trois ponts intramoléculaires chez la piline majeure. Néanmoins, les analyses par spectrométrie de masse des variants de ces ponts ont indiqué que l'un d'entre eux n'est pas formé dans la protéine recombinante en solution. Une analyse supplémentaire des séquences de RrgA, RrgB et de leurs homologues a permis l'identification d'une « D-box » dont le résidu Aspartate hautement conservé est impliqué dans la formation du dernier pont intramoléculaire. La fonction des ponts intramoléculaires au sein de la piline majeure a été prouvée par des analyses de thermostabilité et de digestion protéolytique réalisées sur RrgB et ses variants de ponts. Enfin, la structure cristallographique de RrgB a été résolue indiquant que la protéine se compose de quatre domaines présentant chacun une topologie de type IgG. Contrairement à la protéine en solution, la forme cristalline montre la présence de quatre ponts intramoléculaires localisés dans chaque domaine de la protéine. Enfin, l'organisation des monomères de la piline majeure dans l'empilement cristallin semble mimer la fibre RrgB native apportant ainsi des informations clés sur les résidus potentiellement impliqués dans les liaisons intermoléculaires et qui seront utilisés dans la seconde partie de l'étude.

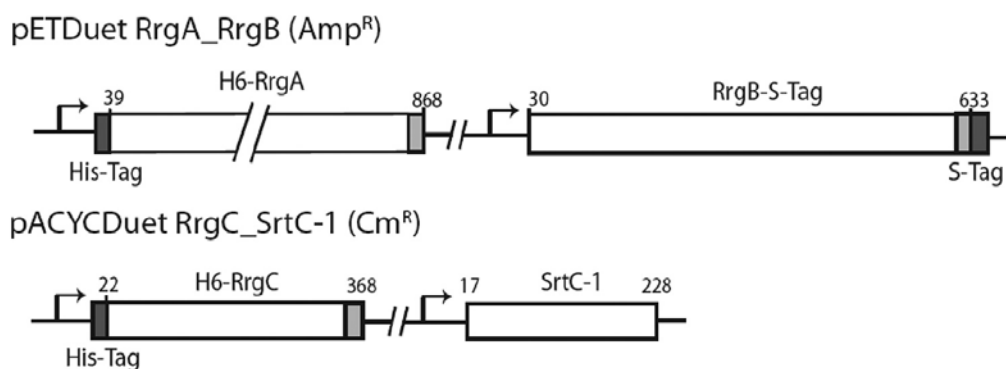
## II. Détermination du rôle des sortases

Afin d'élucider le mécanisme moléculaire de biogenèse du pilus, nous nous sommes proposé de mimer, par des approches biochimiques, les réactions d'assemblage catalysées par les sortases. Dans ce but, nous avons développé chez *Escherichia coli*, une plateforme de co-expression des pilines et des sortases recombinantes, qui dans différentes combinaisons, permettent de reconstituer chacune des étapes de l'assemblage du pilus de *S. pneumoniae*. Des indications moléculaires précises quant à la fonction de chacune des sortases ont été obtenues et sont ici présentées.

### II.1 Fonction de la sortase C-1

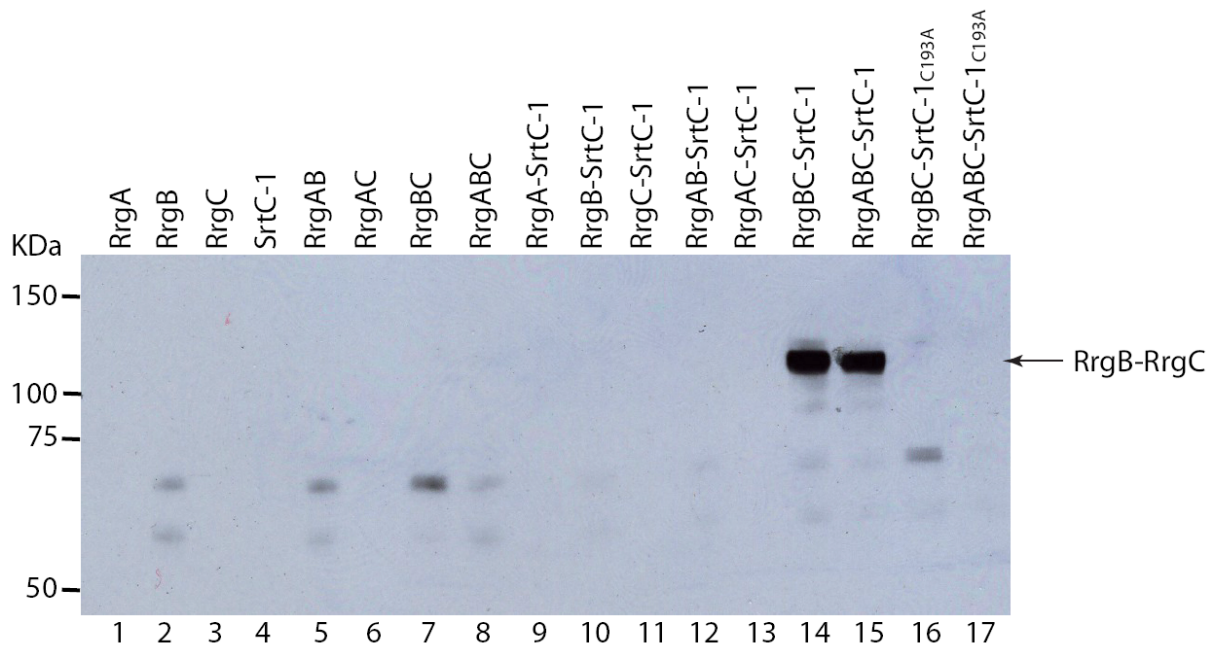
#### II.1.1 La sortase C-1 associe covalentement RrgB et RrgC

Nous avons montré dans une étude précédente (**Publication 1**) que la SrtC-1 est la principale enzyme responsable de la polymérisation de RrgB (Manzano 2008). Afin de déterminer si cette sortase a pour unique substrat RrgB, ou si elle est également capable de reconnaître RrgA et RrgC, nous avons mis au point un système de co-expression qui permet de détecter les interactions établies entre les pilines et les sortases mais également d'identifier les produits de réactions, c'est-à-dire les complexes covalents entre pilines catalysés par les sortases (Cf. partie Matériel et Méthode). Un schéma récapitulatif des constructions génétiques réalisées pour cette étude est représenté dans la figure 57.



**Figure 57: Constructions utilisées dans la plateforme de co-expression *E. coli*.** Représentation schématique des constructions pETDuet RrgA RrgB et pACYCDuet RrgC SrtC-1 (El Mortaji *et al.* 2010)

Seules les protéines RrgA et RrgC sont fusionnées à une étiquette 6xHis permettant leur purification sur résine de nickel et différentes combinaisons de co-expression piline-sortases ont été testées. Tous les produits issus des co-expressions ont été purifiés en parallèle sur un système 96-puits His-Trap™ HP permettant la rétention des protéines His-RrgA et His-RrgC. Ces protéines sont ensuite éluées et séparées par électrophorèse sur un gel de polyacrylamide 4-12%. Les protéines sont ensuite transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps polyclonaux anti-RrgB (Fig.58). La protéine RrgB ne possédant pas l'étiquette 6xHis, sa présence dans les fractions éluées résulte de son association avec soit RrgA, soit RrgC (Fig.57).



**Figure 58: Catalyse par SrtC-1 du complexe covalent RrgB-RrgC.** Les co-expressions de RrgA, RrgB, RrgC et SrtC-1 ont été réalisées chez *E. coli*. Les purifications ont toutes été réalisées en parallèle sur le système His-Trap™ HP columns et seules les pilines accessoires RrgA et RrgC présentent une étiquette poly-histidine à leur extrémité N-terminale. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps anti-RrgB polyclonaux (El Mortaji *et al.* 2010).

Comme attendu, la protéine RrgB n'est pas immunodétectée dans les fractions d'éluion des souches exprimant seulement RrgA, RrgC et SrtC-1 (Fig.58, lignes 1, 3, 4). Cependant, malgré l'absence d'étiquette poly-histidine sur la protéine recombinante RrgB, deux bandes de faible intensité (60 et 70 kDa) sont révélées dans les fractions d'éluion issues des co-expressions contenant RrgB (Fig.58, lignes 2, 5, 7, 8 et 16). Cette observation signifie qu'une faible proportion de RrgB se lie de façon non spécifique à la résine His-Trap. Le résultat le plus marquant concerne les conditions de co-expressions de SrtC-1, RrgB et RrgC dans lesquelles est présente une espèce protéique migrant à une masse moléculaire apparente de

120 kDa, reconnue par les anticorps anti-RrgB (Fig.58, ligne 14) et anti-RrgC (non montré). Ce résultat suggère que la SrtC-1 est capable de lier covalentement RrgB et RrgC. Le même produit est identifié lorsque RrgA est également co-exprimée, bien que ne jouant aucun rôle dans la formation du complexe RrgB-RrgC (Fig.58, ligne 15). La co-expression de la SrtC-1 avec RrgA et RrgB ne donne lieu à la formation d'aucun complexe entre ces deux pilines (Fig.58, ligne 12) démontrant ainsi que la sortase C-1 présente une spécificité de substrat pour l'association de RrgB et RrgC. Afin de vérifier la nature catalytique de la formation du complexe RrgB-RrgC, nous avons muté la cystéine 193 du site actif de la SrtC-1 en alanine et avons réalisé la même expérience que précédemment décrite. Aucun complexe RrgB-RrgC n'a pu être identifié, indiquant que l'association entre RrgB et RrgC est spécifiquement catalysée par SrtC-1 (Fig.58, lignes 16 et 17).

### II.1.2 La sortase C-1 reconnaît le motif IPQTG de RrgB

Afin de vérifier la composition du complexe catalysé par la SrtC-1, celui-ci a été purifié et analysé par spectrométrie de masse en tandem ou MS/MS (Tableau 11). Les données obtenues indiquent que RrgB et RrgC constituent effectivement ce complexe. Cependant, l'analyse des peptides issus de la protéolyse du complexe n'a pas permis d'identifier la liaison covalente établie entre RrgB et RrgC (Tableau 11).

**Tableau 11: Résumé des peptides identifiés lors de l'étude du complexe RrgB-RrgC par MS/MS**  
(El mortaji *et al.* 2010)

Observed	Mr (expt)	Mr (calc)	Delta	Start	Stop	Sequence	Identified protein
521,75	1041,48	1041,48	0,00	243	251	MTEGLAFNK	RrgB
523,79	1045,56	1045,56	0,00	129	138	FNTANLPAAK	RrgB
555,75	1109,49	1109,49	0,00	42	51	LLATDGDMDK	RrgB
556,79	1111,56	1111,56	0,00	199	208	ANPDTPRVVK	RrgB
563,29	1124,57	1124,57	0,00	300	309	ITYSATLNDK	RrgB
571,27	1140,52	1140,52	0,00	109	119	LSGAMPATAMK	RrgB
588,82	1175,63	1175,63	0,00	580	590	QPAGYALLTSR	RrgB
602,84	1203,67	1203,67	0,00	377	387	VVQTVTLTDDK	RrgB
635,31	1268,61	1268,61	0,00	109	120	LSGAMPATAMKK	RrgB
679,89	1357,76	1357,76	0,00	491	502	QLVVTTKDALDR	RrgB
714,36	1426,71	1426,71	0,00	243	255	MTEGLAFNKGTVK	RrgB
740,37	1478,73	1478,73	0,01	431	443	DENPKPLDPTPEK	RrgB
744,92	1487,82	1487,82	0,00	485	497	VSQEEKQLVVTTK	RrgB
747,37	1492,72	1492,72	0,00	52	65	IANELETGNYAGNK	RrgB
831,92	1661,82	1661,82	0,00	228	242	IPALANYATANWSDR	RrgB
843,94	1685,86	1685,86	0,01	281	296	LTDAGLAKVNDQNAEK	RrgB
923,49	1844,97	1844,96	0,00	121	138	LTEAEGAKFNTANLPAAK	RrgB
958,47	1914,93	1914,92	0,00	410	427	GYSADYQEITTAGIIVK	RrgB
644,37	1930,08	1930,07	0,00	481	497	KADKVSQEEKQLVVTTK	RrgB
982,00	1961,98	1961,98	0,00	503	520	AVAAYNALTAQQQTQKEK	RrgB
987,54	1973,06	1973,06	0,00	120	138	KLTEAEGAKFNTANLPAAK	RrgB
687,04	2058,10	2058,10	0,00	485	502	VSQEEKQLVVTTKDALDR	RrgB
1032,53	2063,05	2063,04	0,01	461	480	LAGAIEFVIANADNAGQYLAR	RrgB
1071,54	2141,07	2141,06	0,00	209	227	DTPVNHQVGDVVEYIIVTK	RrgB
1110,57	2219,12	2219,12	0,00	503	522	AVAAYNALTAQQQTQKEKEK	RrgB
1166,14	2330,26	2330,25	0,01	163	183	AVPIEIELPLNDVDAHVYPK	RrgB

1171,09	2340,16	2340,15	0,01	141	162	IYEIHSLSYVGEDGATLTGSK	RrgB
781,75	2342,23	2342,22	0,01	52	74	IANELETGNYAGNKVGLPANAK	RrgB
828,76	2483,26	2483,25	0,00	206	227	VDKDTPVNHQVGDVVVEYIVTK	RrgB
845,10	2532,27	2532,26	0,01	498	520	DALDRAVAAYNALTAQQQTQKEK	RrgB
854,78	2561,31	2561,31	0,00	503	525	AVAAYNALTAQQQTQKEKVDK	RrgB
861,09	2580,24	2580,23	0,01	591	615	QKFEVTATSYATGGQIEYTAGSGK	RrgB
862,41	2584,20	2584,20	0,00	42	65	LLATDGDMDKIANELETGNYAGNK	RrgB
1306,66	2611,30	2611,29	0,01	256	280	VTVDDVALEAGDYALTEVATGFDLK	RrgB
1316,66	2631,31	2631,31	0,01	139	162	YKIYEIHSLSYVGEDGATLTGSK	RrgB
1037,83	3110,48	3110,47	0,01	591	620	QKFEVTATSYATGGQIEYTAGSGKDDATK	RrgB
1079,22	3234,62	3234,62	0,01	199	227	ANPDTPRVDKDTPVNHQVGDVVVEYIVTK	RrgB
1127,92	3380,74	3380,72	0,02	256	288	VTVDDVALEAGDYALTEVATGFDLKLTDAGLAK	RrgB
1145,58	3433,71	3433,70	0,01	42	74	LLATDGDMDKIANELETGNYAGNKVGLPANAK	RrgB
983,77	3931,05	3931,04	0,01	555	590	LVSDAQRFEITGLLAGTYYLEETKQPAGYALLTSR	RrgB
Observed	Mr (expt)	Mr (calc)	Delta	Start	Stop	Sequence	Identified protein
470,73	939,44	939,44	0,00	190	198	YSSSGQVGR	RrgC
538,29	1074,57	1074,57	0,00	181	189	EVPLIGEYR	RrgC
572,26	1142,50	1142,50	0,00	72	80	DLHSWDENK	RrgC
573,81	1145,61	1145,60	0,00	297	307	EVVVTSGKDGR	RrgC
607,30	1212,58	1212,58	0,00	269	279	TNTSLQGAMFK	RrgC
624,75	1247,50	1247,49	0,00	57	66	LDDSYSYDDR	RrgC
779,88	1557,75	1557,74	0,00	283	296	EESGHYTPVLQNGK	RrgC
817,43	1632,84	1632,84	0,00	176	189	DVSEKEVPLIGEYR	RrgC
853,42	1704,82	1704,81	0,01	72	85	DLHSWDENKLSFFK	RrgC
853,95	1705,88	1705,88	0,00	205	219	NGEIFVTNLPLGNRYR	RrgC
611,98	1832,91	1832,91	0,00	72	86	DLHSWDENKLSFFK	RrgC
615,30	1842,88	1842,87	0,00	57	71	LDDSYSYDDRQIVR	RrgC
974,98	1947,94	1947,94	0,00	280	296	VMKEESGHYTPVLQNGK	RrgC
786,74	2357,19	2357,19	0,00	283	304	EESGHYTPVLQNGKEVVVTSGK	RrgC
1214,63	2427,25	2427,24	0,01	199	219	TLYTDKNGEIFVTNLPLGNRYR	RrgC
896,12	2685,34	2685,34	0,00	283	307	EESGHYTPVLQNGKEVVVTSGKDGR	RrgC
916,80	2747,38	2747,38	0,00	280	304	VMKEESGHYTPVLQNGKEVVVTSGK	RrgC
755,62	3018,45	3018,44	0,01	87	111	TSFEMTFLENQIEVSHIPNGLYYVR	RrgC
1026,19	3075,53	3075,53	0,01	280	307	VMKEESGHYTPVLQNGKEVVVTSGKDGR	RrgC

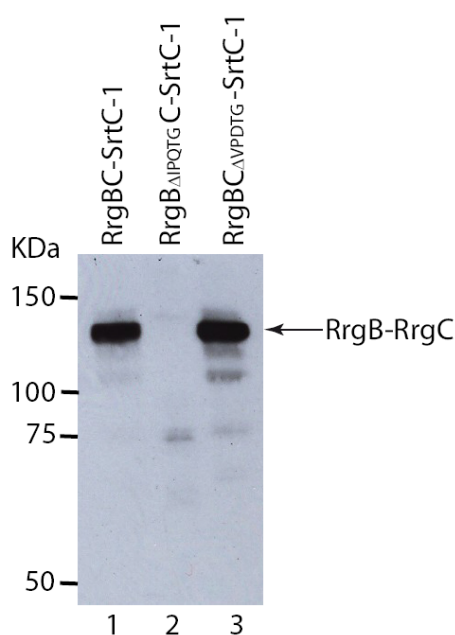
Une autre analyse par spectrométrie de masse a été effectuée sur le complexe RrgB-RrgC purifié dans le but de déterminer sa masse. La valeur mesurée est de 106 565.51 Da. Cette valeur correspond à la liaison covalente d'une molécule His-RrgC associée à une sous-unité RrgB<sub>IPQT</sub>, dont le motif IPQTG a été clivé entre les résidus thréonine et glycine et qui a donc perdu le S-Tag C-terminal (Tableau 12). Les 2.16 Da de différence de masse observés sont compris dans la précision de mesure de la technique (1 pour 10 000 Da). Ce résultat confirme le mode d'action proposé pour la SrtC-1 : la reconnaissance du motif IPQTG de RrgB et la catalyse de la liaison peptidique avec RrgC.

**Tableau 12: Récapitulatif des masses des protéines recombinantes seules ou en complexe.**

Protéine	Masse théorique (Da)	Masses réelles <sup>(*)</sup> (Da)
His_RrgC	40 983.07	40 949.07
RrgB_STag	68 332.13	68 281.13
RrgB <sub>IPQT</sub>	65 683.28	65 632.28
RrgB_STag-His_RrgC	109 315.20	109 230.20
Liaison covalente (-18 Da)	109 297.20	109 212.20
RrgB <sub>IPQT</sub> -His_RrgC	106 666.35	106 581.35
Liaison covalente (-18 Da)	106 648.35	<b>106 563.35</b>

<sup>(\*)</sup> Prends en compte la présence des ponts intramoléculeaires

Dans le but de confirmer la spécificité de reconnaissance du motif IPQTG de RrgB par la SrtC-1, les motifs IPQTG de RrgB et VPDTG de RrgC ont été délétés et les protéines mutantes résultantes ont été analysées dans le système de co-expression (Fig.59). Le complexe RrgB-RrgC est produit même lorsque le motif VPDTG de RrgC est absent (Fig.59, ligne3). Au contraire, la production du complexe est totalement abolie lorsque le motif IPQTG de RrgB est délété (Fig.59, ligne 2). Ces observations démontrent donc que la catalyse du complexe RrgB-RrgC par SrtC-1 est dépendante de la reconnaissance du motif IPQTG de RrgB.

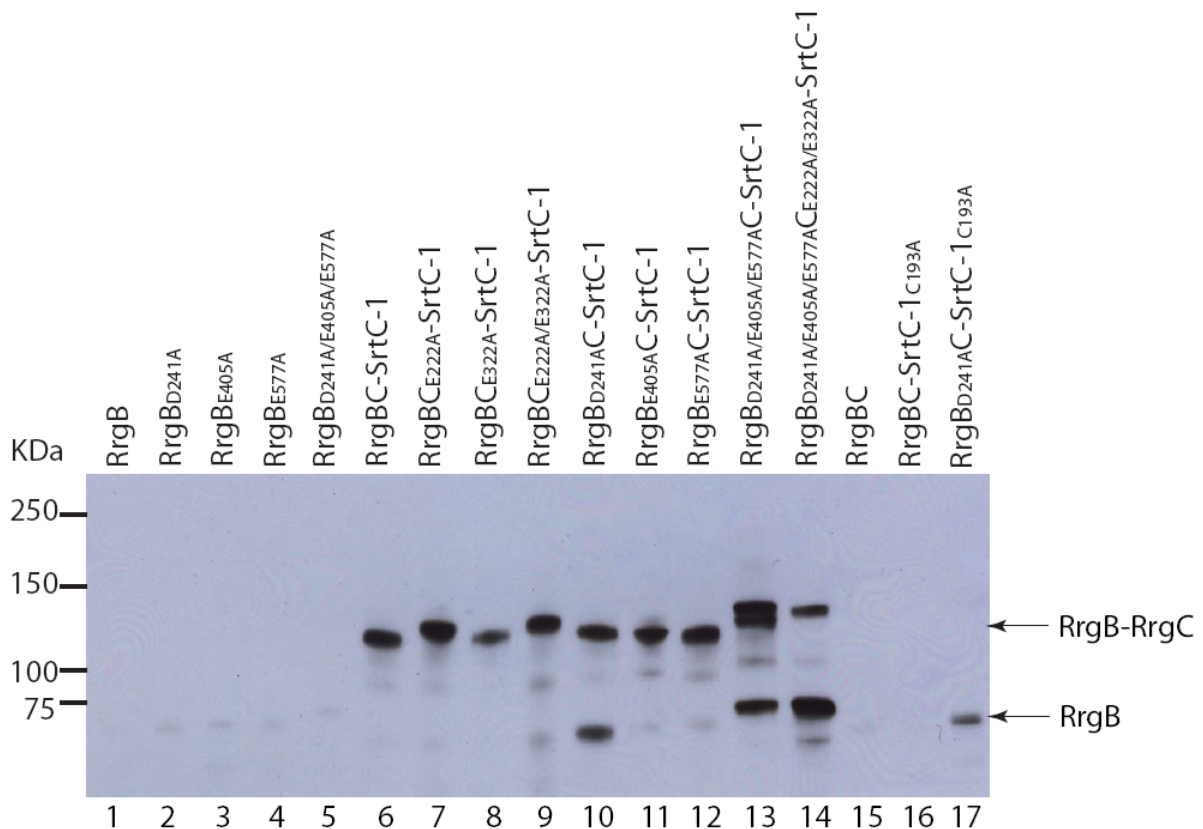


**Figure 59: Spécificité de SrtC-1 dans la reconnaissance du substrat.** Les motifs LPxTG des protéines RrgB et RrgC ont été délétés, et chaque mutant a été testé dans sa capacité à être reconnu et pris en charge par SrtC-1. Les purifications des protéines co-exprimées, RrgB, RrgC et SrtC-1 ont été réalisées sur les colonnes His-Trap<sup>TM</sup> HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par des anticorps polyclonaux anti-RrgB (El Mortaji *et al.* 2010).

### II.1.3 Les ponts intramoléculaires de RrgB jouent un rôle dans l'efficacité de reconnaissance des pilines par la SrtC-1

Les ponts intramoléculaires stabilisent les pilines RrgB et RrgC qui sont covalentement associées par SrtC-1. Par conséquent, nous avons cherché à savoir si la compaction induite par les ponts isopeptidiques pouvait influencer la formation du complexe covalent RrgB-RrgC. Les résidus glutamate et aspartate impliqués dans la formation des ponts chez RrgB et RrgC ont été mutés en alanine et testés dans le système de co-expression décrit ci-dessus (Fig.60). Les mutations individuelles Glu222 et Glu322 chez RrgC ainsi que les mutations Asp241, Glu405 et Glu577 chez RrgB n'affectent pas la formation du complexe (Fig.60, lignes 7, 8, 10, 11, et 12). Cependant, lorsque l'Asp241 est mutée en alanine chez RrgB, une

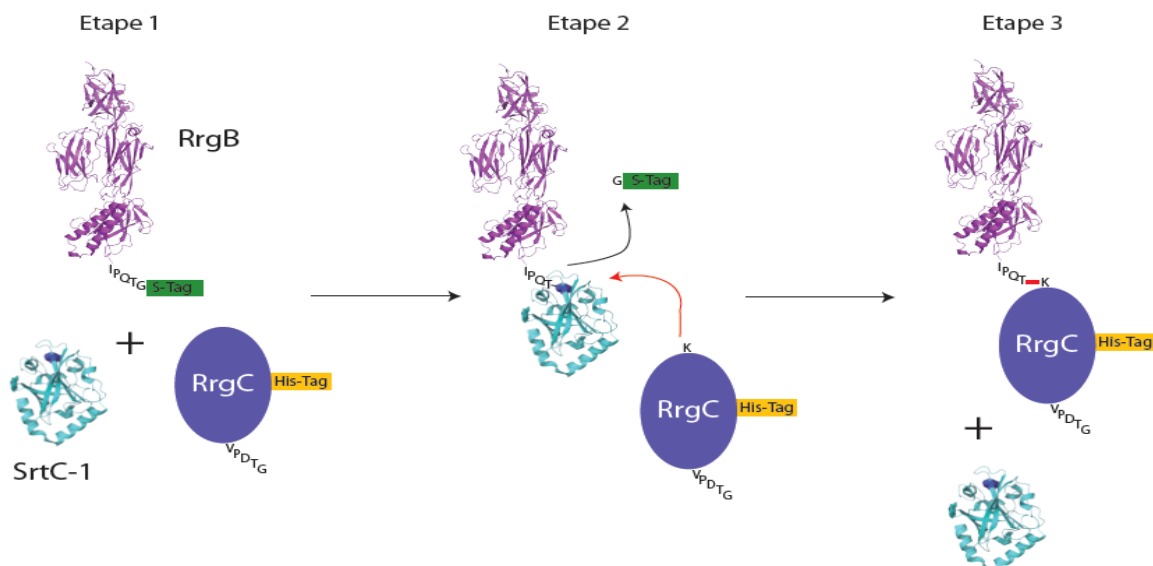
quantité significative de RrgB monomérique est présent dans les éluions des colonnes His-Trap™ HP, indiquant qu'une importante quantité de RrgB reste liée à RrgC via des interactions non covalentes (Fig.60, ligne 10). La co-élution du variant de RrgB n'est pas due à une liaison artificielle à la résine car lorsqu'il est exprimé seul, ce mutant présente le même comportement que celui de la protéine RrgB native (Fig.60, ligne 1 et 2). L'association non covalente du mutant Asp241 de RrgB à RrgC n'est pas dépendante de l'activité de la SrtC-1, car cette forme mutante demeure associée à RrgC même lorsque la SrtC-1 est inactive (Fig.60, ligne 17). Lorsque les trois ponts intramoléculeaires de RrgB sont mutés, le complexe RrgB-RrgC migre sous la forme de deux bandes de masse moléculaire apparente de 140 et 150 kDa (Fig.60, ligne 13). Enfin, l'absence de tous les ponts intramoléculeaires dans RrgB et RrgC mène à une catalyse moins efficace du complexe par la SrtC-1 comme en atteste la faible quantité de complexe RrgB-RrgC formé dans ces conditions (Fig.60, ligne 14). En conclusion, la stabilisation des structures de RrgB et RrgC par les ponts intramoléculeaires joue un rôle dans l'efficacité de reconnaissance de ces pilines par la SrtC-1.



**Figure 60: Rôle des ponts intramoléculeaires dans la formation du complexe covalent RrgB-RrgC.** Les variants des ponts intramoléculeaires de RrgB et RrgC ont été introduits dans la plateforme de co-expression. La purification de toutes les formes de RrgB (native, simple et triple mutant) et des protéines co-exprimées RrgB, RrgC et SrtC-1 a été réalisée grâce au système de colonnes His-Trap™ HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées à l'aide d'anticorps polyclonaux anti-RrgB (El Mortaji *et al.* 2010).

## II.1.4 Conclusion

Le système de co-expression que nous avons développé a permis de mettre en évidence la fonction de la SrtC-1 dans la génération du complexe covalent entre RrgB et RrgC. La nature catalytique de ce processus a été validée par des expériences de mutagenèse : l'inactivation de l'enzyme par mutation de la cystéine catalytique en alanine abolie la formation du complexe RrgB-RrgC. L'analyse de la composition de ce complexe par des méthodes complémentaires (Western-Blot avec des anticorps anti-RrgB et anti-RrgC et spectrométrie de masse MS/MS) révèle que celui-ci est formé des deux pilines RrgB et RrgC. Des mesures par spectrométrie de masse electrospray ont conclu à une stœchiométrie 1 :1 indiquant qu'un monomère HisRrgC est lié covalent à un monomère RrgB<sub>IPQT</sub>, ce dernier ayant perdu la Gly et le S-Tag C-terminaux. La spécificité de reconnaissance du motif IPQTG de RrgB par la SrtC-1 est donc ainsi définie et a été validée par les expériences de délétion des motifs de RrgB et RrgC : en effet, seule l'absence de la séquence IPQTG spécifique à RrgB empêche la formation du complexe RrgB-RrgC. En conclusion, ces résultats indiquent que la sortase C-1 présente une spécificité de reconnaissance stricte pour le motif IPQTG de RrgB et catalyse la formation de la liaison peptidique entre RrgB et RrgC (Fig.61). Enfin, nous avons montré que l'absence des ponts intramoléculaires, notamment ceux de RrgB, influence négativement l'efficacité catalytique de formation du complexe RrgB-RrgC par la SrtC-1.



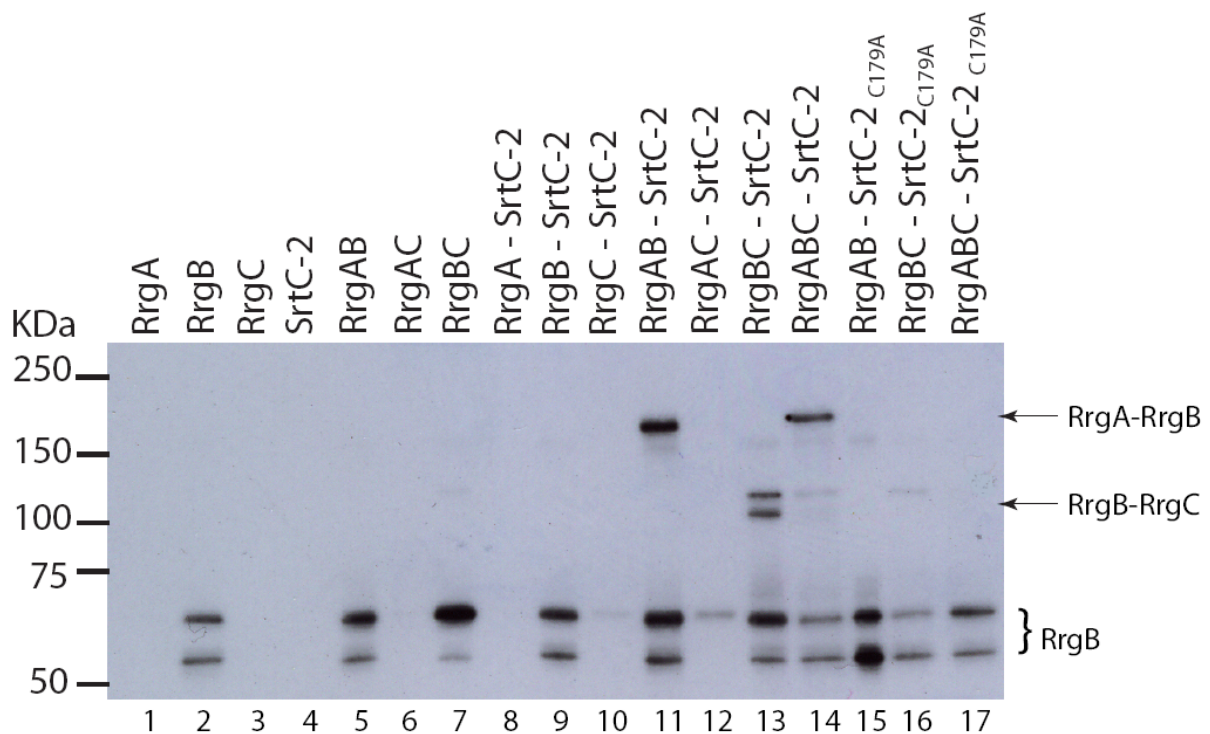
**Figure 61: Schéma récapitulatif de la catalyse du complexe RrgB-RrgC par SrtC-1.** **Etape 1.** Les protéines RrgB\_S-Tag, His\_RrgC et SrtC-1 sont co-exprimées chez *E. coli*. RrgB et SrtC-1 sont représentés par leur structure cristallographique et RrgC par une ellipse. **Etape 2.** SrtC-1 reconnaît le motif IPQTG de RrgB et clive entre les résidus Thréonine et Glycine libérant ainsi la Glycine terminale et le S-Tag (Flèche noire). Le complexe acyl-enzyme généré subit une attaque nucléophile par l'amine d'un résidu Lysine de RrgC qui reste à être identifié (Flèche rouge). **Etape 3.** Le complexe RrgB-RrgC est formé aboutissant à la libération de SrtC-1 (La liaison intermoléculaire est représentée en rouge).



## II.2 Fonction de la sortase C-2

### II.2.1. La SrtC-2 catalyse deux complexes covalents différents : RrgA-RrgB et RrgB-RrgC

L'étude de la fonction de la SrtC-2 s'est déroulée de façon similaire à celle développée pour la SrtC-1 décrite ci-dessus. Après avoir conduit les co-expressions de RrgA, RrgB, RrgC et SrtC-2 dans différentes combinaisons, les protéines sont purifiées en parallèle sur le système His-MultiTrap™ HP columns et analysées par immunodétection avec les anticorps polyclonaux anti-RrgB (Fig.62).



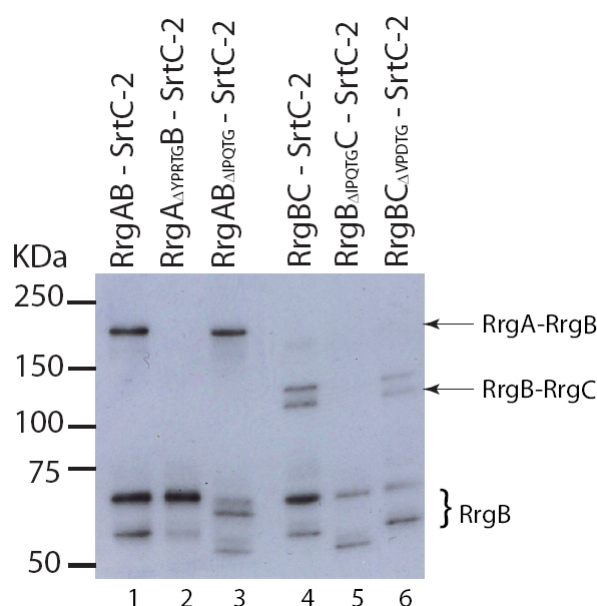
**Figure 62: Catalyse par SrtC-2 des complexes covalents RrgA-RrgB et RrgB-RrgC.** Les co-expressions de RrgA, RrgB, RrgC et SrtC-2 ont été réalisées chez *E. coli*. Les purifications ont toutes été réalisées en parallèle sur le système His-Trap™ HP columns et seules les pilines accessoires RrgA et RrgC présentent une étiquette poly-histidine à leur extrémité N-terminale. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps anti-RrgB polyclonaux (El Mortaji *et al.* soumis).

De façon similaire à ce qui a été observé précédemment, la protéine monomérique RrgB n'est pas immuno-détectée dans les éluations issues des surexpressions de RrgA, RrgC et SrtC-2 (Fig.62, lignes 1, 3 et 4) mais lorsque produite, reste en partie associée à la résine (Fig.62, lignes 2, 5, 7, 9, 11, 13, 14, 15, 16 et 17). Le résultat intéressant de cette expérience est la détection d'une bande de masse moléculaire apparente de 200 kDa, immuno-détectée par les anticorps anti-RrgB dans les conditions où la SrtC-2 et les pilines RrgA et RrgB sont co-exprimées (Fig.62, ligne 11). Cette espèce de haut poids moléculaire est également détectée par les anticorps anti-RrgA (non montré). Dans une combinaison semblable mais dans

laquelle la cystéine catalytique de SrtC-2 a été mutée en alanine, la formation du complexe est abolie (Fig.62, ligne 15). Ces résultats suggèrent donc que la SrtC-2 peut lier covalentement RrgA et RrgB. D'autre part, deux bandes immuno-réactives avec les anticorps anti-RrgB et de masse moléculaire apparente de 110 et 120 kDa sont également détectées lorsque la SrtC-2 est co-exprimée avec les pilines RrgB et RrgC, indiquant que ces deux formes peuvent correspondre à un complexe covalent RrgB-RrgC (Fig.62, ligne 13). La formation de ce complexe est également abolie lorsque la SrtC-2 est inactive (Fig.62, ligne 16). Il est intéressant de noter que lorsque les trois pilines, RrgA, RrgB et RrgC sont co-exprimées avec la SrtC-2, le complexe RrgA-RrgB est préférentiellement formé (Fig.62, ligne 14).

### II.2.2. La SrtC-2 reconnaît le motif YPRTG de RrgA et le motif IPQTG de RrgB

Afin de mieux caractériser les deux complexes RrgA-RrgB et RrgB-RrgC catalysés par la SrtC-2, nous avons délété les motifs LPxTG de chacune des pilines et analysé l'effet de ces mutations dans le système de co-expression (Fig.63).



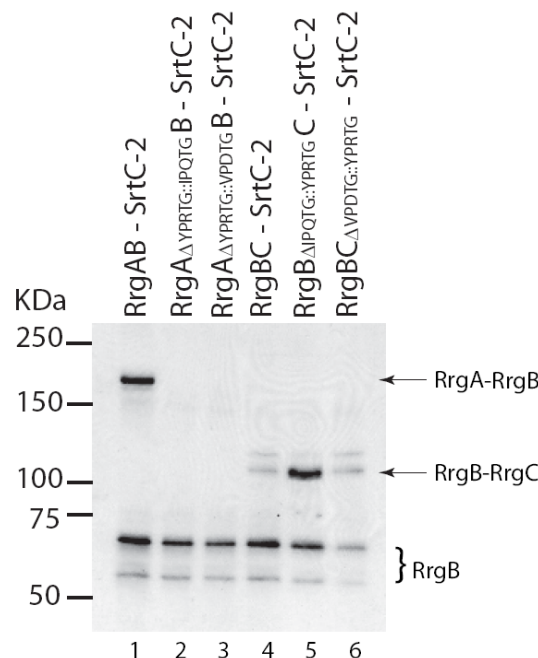
**Figure 63: Spécificité de SrtC-2 dans la reconnaissance des substrats.** Les motifs LPxTG des protéines RrgA, RrgB et RrgC ont été délétés, et chaque mutant a été testé dans sa capacité à être reconnu et pris en charge par SrtC-2. Les purifications des protéines co-exprimées, RrgA, RrgB, RrgC et SrtC-2 ont été réalisées sur les colonnes His-Trap<sup>TM</sup> HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par des anticorps polyclonaux anti-RrgB (El Mortaji *et al.* soumis).

La délétion du motif YPRTG de RrgA abolie totalement la formation du complexe RrgA-RrgB, contrairement à la délétion du motif IPQTG de RrgB qui n'a aucun effet sur ce complexe (Fig.63, lignes 2 et 3). En revanche, dans le cas de la catalyse du complexe RrgB-RrgC, la délétion du motif IPQTG inhibe totalement la formation du complexe alors que celle du motif VPDTG de RrgC n'a aucun effet. Ces résultats indiquent donc que la SrtC-2 catalyse

le complexe RrgA-RrgB via la reconnaissance du motif YPRTG de RrgA et le complexe RrgB-RrgC via la reconnaissance du motif IPQGTG de RrgB, suggérant une double spécificité de reconnaissance de substrats par la SrtC-2.

### II.2.3. Substrat privilégié de la SrtC-2 : le motif YPRTG de RrgA

La SrtC-2 catalyse la formation des complexes RrgA-RrgB et RrgB-RrgC mais lorsque les trois pilines, RrgA, RrgB et RrgC sont co-exprimées avec la SrtC-2, le complexe RrgA-RrgB est préférentiellement formé (Fig.62, ligne 14). Ce résultat suggère que la SrtC-2 interagit de façon privilégiée avec le motif YPRTG de RrgA. Afin de vérifier cette hypothèse, nous avons échangé les motifs LPxTG entre chacune des pilines (Fig.64).



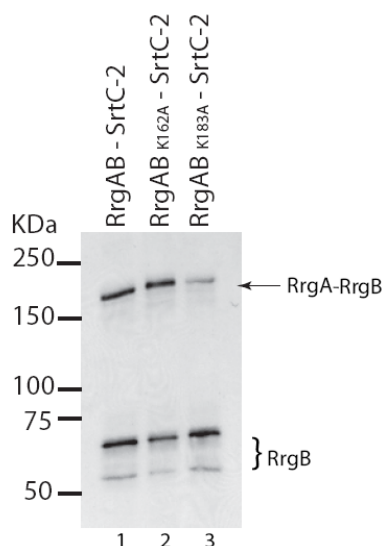
**Figure 64: SrtC-2 a un substrat privilégié : le motif YPRTG de RrgA.** Les motifs LPxTG des protéines RrgA, RrgB et RrgC ont été inversés, et chaque mutant a été testé dans sa capacité à être reconnu et pris en charge par SrtC-2. Les purifications des protéines co-exprimées, RrgA, RrgB, RrgC et SrtC-2 ont été réalisées sur les colonnes His-MultiTrap<sup>TM</sup> HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par des anticorps polyclonaux anti-RrgB (El Mortaji *et al.* soumis).

Les complexes RrgA-RrgB et RrgB-RrgC sont formés lorsque chacune des pilines présente son propre motif LPxTG (Fig.64, lignes 1 et 4). En revanche, lorsque le motif YPRTG de RrgA est remplacé par celui de RrgB ou RrgC, respectivement IPQGTG et VPDTG, la formation du complexe RrgA-RrgB est abolie (Fig.64, lignes 2 et 3). Ces résultats indiquent que le motif de reconnaissance YPRTG est essentiel à la catalyse du complexe RrgA-RrgB par la SrtC-2. Dans le cas du complexe RrgB-RrgC, les motifs IPQGTG et VPDTG de, respectivement, RrgB et RrgC ont été remplacés par le motif YPRTG de RrgA (Fig.64, lignes

5 et 6). La catalyse du complexe RrgB-RrgC est favorisée lorsque le motif de reconnaissance de RrgB est remplacé par le motif YPRTG de RrgA (Fig.64, ligne 5). Ceci suggère que le motif YPRTG, à lui seul, augmente la capacité de catalyse du complexe RrgB-RrgC par la SrtC-2, confirmant la grande spécificité de reconnaissance de cette séquence par la SrtC-2. Par contre, lorsque le motif YPRTG est associé à RrgC, aucun changement significatif n'est observé, indiquant que la SrtC-2 ne reconnaît RrgC qu'en tant qu'accepteur nucléophile (Fig.64, ligne 6). En conclusion de ces expériences, nous proposons que la SrtC-2 présente deux substrats, les pilines RrgA et RrgB, avec néanmoins une préférence marquée pour le motif YPRTG de RrgA.

#### II.2.4. Liaison intermoléculaire entre RrgA-RrgB

La structure cristallographique de RrgB suggère que deux lysines différentes (Lys 162 et Lys 183) peuvent être impliquées dans la polymérisation de la fibre de RrgB. Etant donné que la catalyse du complexe RrgA-RrgB par la sortase C-2 fait intervenir le YPRTG de RrgA, c'est l'une des Lysine de RrgB qui va jouer le rôle d'accepteur pour générer la liaison intermoléculaire. Conscient qu'à ce stade de l'étude toutes les lysines exposées de RrgB sont des candidates nucléophiles potentielles, nous avons néanmoins voulu vérifier si les Lys 162 ou 183 pouvait être impliquées dans la formation du complexe RrgA-RrgB. Ainsi, dans la co-expression impliquant RrgA, RrgB et la sortase C-2 nous avons muté indépendamment ces deux Lysine en Alanine (Fig.65).



**Figure 65: La Lysine du motif piline de RrgB est impliquée dans la liaison intermoléculaire de RrgA-RrgB.** Les Lysines 162 et 183 (motif piline) de RrgB ont été mutées en Alanine, et chaque mutant a été co-exprimé avec RrgA et SrtC-2. Les purifications des protéines co-exprimées, RrgA, SrtC-2, RrgB et ses variants ont été réalisées sur les colonnes His-Trap™ HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par des anticorps polyclonaux anti-RrgB (El Mortaji *et al.* soumis).

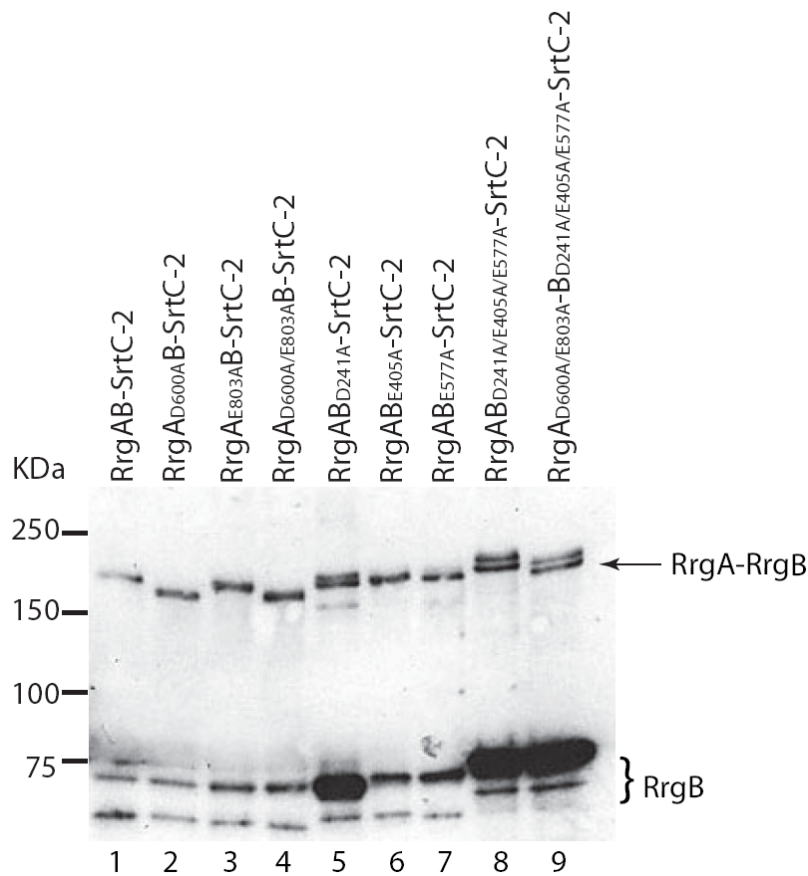
La mutation de la Lysine 162 de RrgB dans la co-expression RrgA, RrgB et SrtC-2, ne présente aucun effet sur la formation du complexe RrgA-RrgB (Fig.65, ligne 2). En revanche, lorsque la Lysine 183, qui n'est autre que la Lysine du motif piline (YPKN), est mutée en Alanine chez RrgB la catalyse du complexe est diminuée (Fig.65, ligne 3). Ces résultats suggèrent donc que la lysine du piline motif est impliquée dans la liaison intermoléculaire de RrgA-RrgB, mais également qu'une autre Lysine de RrgB est susceptible de la remplacer dans cette fonction expliquant que sa mutation n'abolie pas totalement la formation du complexe RrgA-RrgB.

### **II.2.5. Rôle des ponts isopeptidiques dans la formation du complexe RrgA-RrgB**

Les pilines RrgA et RrgB sont stabilisées par des ponts intramoléculaires Lys-Asn. RrgA contient deux ponts, présents dans les domaines D2 et D4 (Izoré 2009). Nous avons identifié trois liaisons intramoléculaires dans RrgB lors des études en solution et la structure cristallographique de RrgB a révélée la formation d'un pont supplémentaire dans le domaine D1. Dans les expériences décrites ci-après, seuls les trois ponts de RrgB présents dans la protéine en solution ont été mutés. La question posée ici est d'évaluer le rôle de ces ponts intramoléculaires dans la formation du complexe RrgA-RrgB par la SrtC-2.

Les résidus glutamate et aspartate impliqués dans la formation des ponts chez RrgA et RrgB ont été mutés en alanine et le comportement des protéines a été analysé dans le système de co-expression : les complexes RrgA-RrgB portant les mutations ponctuelles ont été purifiés et détectés à l'aide d'anticorps polyclonaux anti-RrgB (Fig.66). L'absence de l'un, ou des deux ponts de RrgA, ainsi que les mutations Glu405 et Glu577 dans RrgB n'affectent pas la formation du complexe (Fig.66, lignes 2, 3, 4, 6 et 7). Il est intéressant de noter que les différences de migration observées pour les variants du complexe RrgA-RrgB sont directement reliées à la compaction induite par les ponts intramoléculaires des deux pilines : comparer par exemple la migration du complexe RrgA<sub>D600A</sub>-RrgB par rapport à la forme native (Fig.66, lignes 1 et 2). Lorsque l'Asp241 est mutée en alanine dans RrgB, une quantité significative de RrgB monomérique est détecté, indiquant qu'une quantité supérieure de RrgB<sub>D241A</sub>, par rapport à forme native, reste liée à RrgA via des interactions non covalentes (Fig.66, ligne 5). De plus, le complexe RrgA-RrgB<sub>D241A</sub> se présente sous la forme d'une double bande (Fig.66, lignes 5, 8 et 9). Ces résultats indiquent que la mutation D241A, en déstabilisant RrgB, diminue son efficacité de reconnaissance par la SrtC2. La structure cristallographique de RrgB ne suffit pas, à elle seule, à apporter une interprétation rationnelle

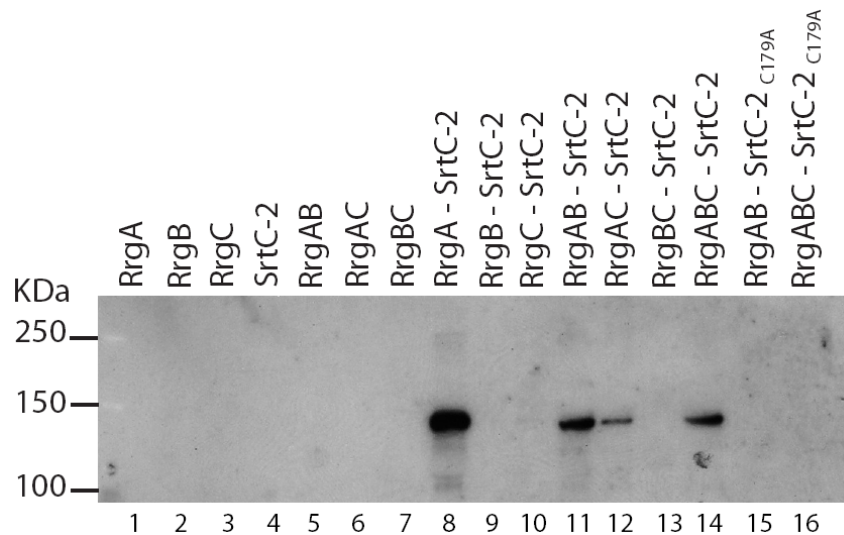
à cette observation. La détermination de la structure du complexe RrgA-RrgB pourrait par contre être utile pour répondre à cette question. Ce projet n'est, en l'état actuel des travaux, pas réalisable au vu des quantités trop faibles de complexe produit.



**Figure 66: Rôle des ponts intramoléculaires dans la formation du complexe covalent RrgA-RrgB.** Les variants des ponts intramoléculaires de RrgA et RrgB ont été introduits dans la plateforme de co-expression. La purification de toutes les formes de RrgB (native, simple et triple mutant) et des protéines co-exprimées RrgA, RrgB et SrtC-2 a été réalisée grâce au système de colonnes His-Trap<sup>TM</sup> HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées à l'aide d'anticorps polyclonaux anti-RrgB (El Mortaji *et al.* soumis).

## II.2.6. SrtC-2 forme un complexe acyl-enzyme stable avec RrgA

La SrtC-2 reconnaît le motif YPRTG de RrgA et clive la liaison peptidique entre les résidus thréonine et glycine en formant un intermédiaire covalent acyl-enzyme. Nous avons recherché la présence d'un tel complexe SrtC-2\*RrgA afin de vérifier, le mécanisme réactionnel de la SrtC-2 et également pour évaluer la stabilité du complexe acyl-enzyme en vue d'une purification à grande échelle. Les protéines éluées provenant des co-expressions impliquant RrgA, RrgB, RrgC et la SrtC-2 ont été séparées par SDS-PAGE gradient 4-12,5%, puis après transfert sur membrane de nitrocellulose, immunorévéloées par un anticorps polyclonal anti-SrtC-2 (Fig.67).

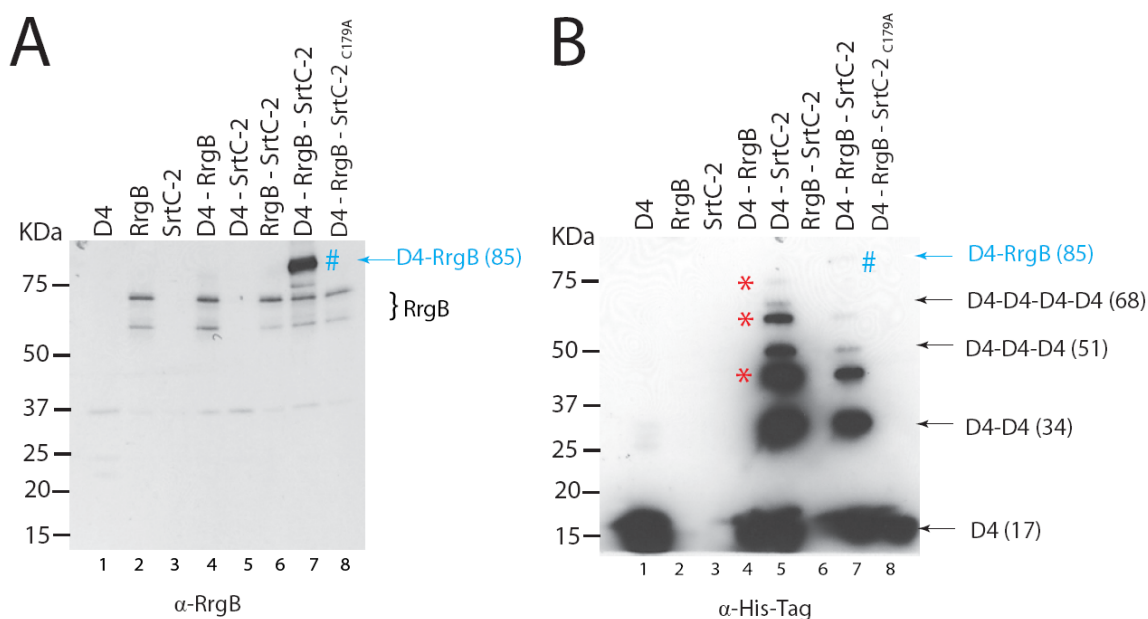


**Figure 67: Existence d'un intermédiaire acyl-enzyme SrtC-2- RrgA stable.** Les co-expressions de RrgA, RrgB, RrgC et SrtC-2 ont été réalisées chez *E. coli*. Les purifications ont toutes été réalisées en parallèle sur le système His-Trap™ HP columns et seules les pilines accessoires RrgA et RrgC présentent une étiquette poly-histidine à leur extrémité N-terminale. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps anti-SrtC-2 polyclonaux.

La SrtC-2 étant dépourvue d'étiquette poly-histidine, sa présence dans les fractions d'élution ne peut être que le fait d'une interaction avec les protéines HisRrgA ou HisRrgC. Dans toutes les co-expressions impliquant RrgA et la SrtC-2, une bande de masse moléculaire apparente de 140 kDa est immuno-détectée par l'anticorps anti-SrtC-2 (Fig.67, lignes 8, 11, 12, 14) et anti-RrgA (non montré). Ce résultat suggère que la SrtC-2 est capable de former un complexe acyl-enzyme stable et covalent avec RrgA. Afin de vérifier la nature catalytique de la formation du complexe SrtC-2\*RrgA, nous avons muté la cystéine 179 du site actif de la SrtC-2 en alanine et avons réalisé la même expérience que précédemment décrite (Fig.67, ligne 15, 16). Dans ces conditions, aucun complexe SrtC-2\*RrgA n'est identifié, indiquant que l'association entre la SrtC-2 et RrgA est dépendant du potentiel catalytique de la sortase.

### II.2.7. SrtC-2 multimérise le domaine D4 de RrgA

Au cours des travaux portant sur la caractérisation individuelle des pilines, nous avons montré que le gène codant pour le domaine D4, cloné dans le plasmide pETDuet, permettait l'expression d'une protéine soluble d'environ 17 kDa dont le pont isopeptidique était formé et exposant à son extrémité C-terminale le motif YPRTG. Afin de faciliter la caractérisation des complexes SrtC-2\*RrgA et RrgA-RrgB, nous avons décidé d'utiliser le domaine D4 seul plutôt que la protéine RrgA entière, dont la masse importante (90 kDa) rend plus difficile les analyses biochimiques. Nous avons ainsi intégré le domaine D4, fusionné à une étiquette 6xHis, dans le système de co-expression en place de la forme complète de RrgA (Fig.68).



**Figure 68: La sortase C-2 catalyse un complexe D4-RrgB et permet la multimérisation de D4.** Les purifications des protéines co-exprimées, D4, RrgB et SrtC-2 ont été réalisées sur les colonnes His-Trap<sup>TM</sup> HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par (A) des anticorps polyclonaux anti-RrgB et (B) par des anticorps anti-His-Tag. Le symbole (#) indique le complexe D4-RrgB, et le symbole (\*) les acyl-enzymes intermédiaires (immuno-réactif avec des anticorps polyclonaux anti-SrtC-2) (El Mortaji *et al.* en préparation).

Lorsque le domaine D4 est co-exprimé avec RrgB et la SrtC-2, une bande intense de masse moléculaire apparente de 85 kDa est révélée par des anticorps polyclonaux anti-RrgB correspondant au complexe D4-RrgB (Fig.68 A, ligne 7). Cette même bande est immunodétectée par des anticorps anti-His-Tag bien que l'intensité soit plus faible (Fig.68 B, ligne 7) (NB : le temps d'exposition de ce film a été délibérément réduit pour permettre de visualiser les autres formes de D4 produites en grandes quantités). Une nouvelle fois, le rôle de l'activité enzymatique de la SrtC-2 dans la formation du complexe D4-RrgB est prouvée (Fig.68 A, ligne 8).

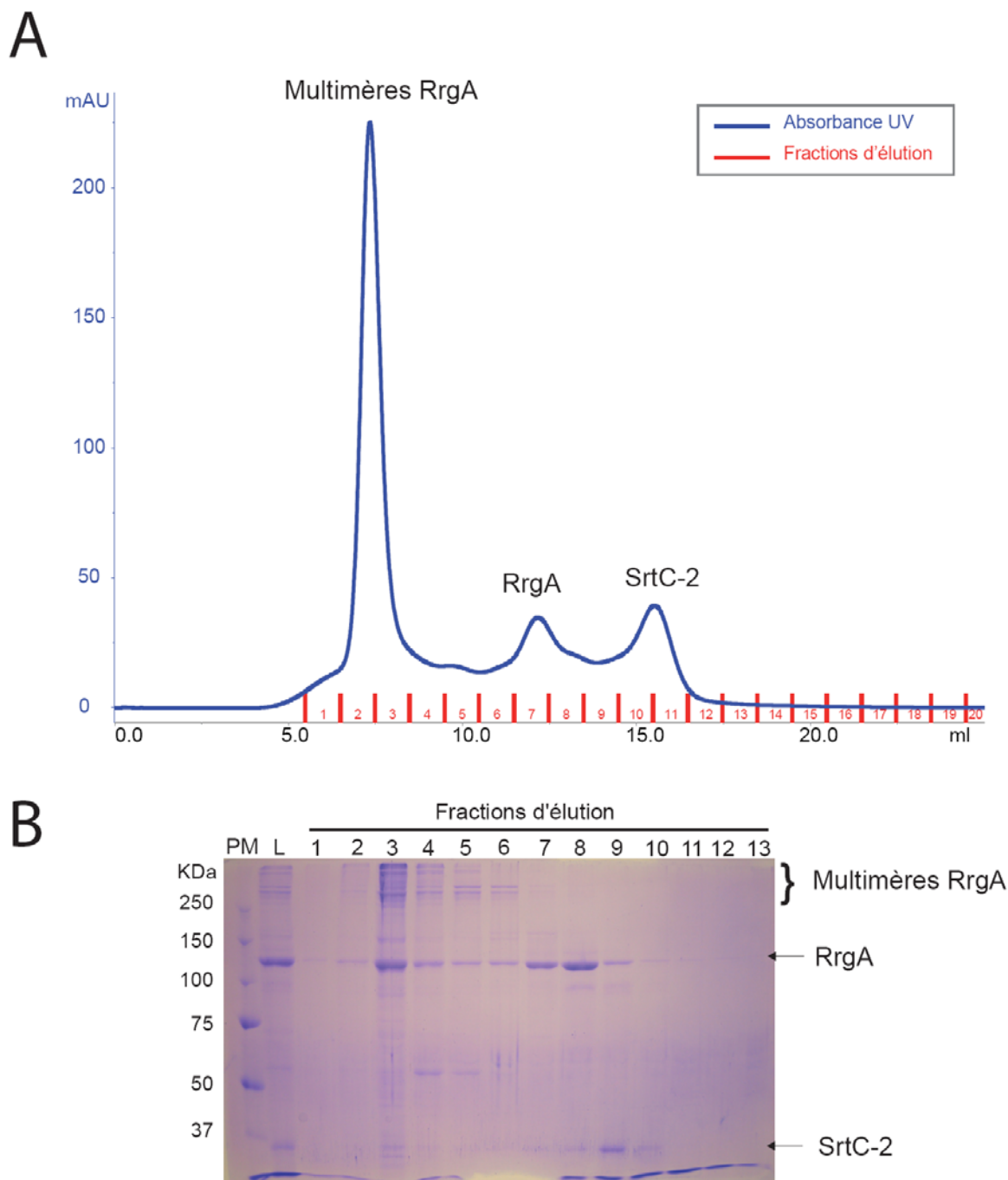
L'immunodétection par les anticorps anti-His-Tag des produits formés lors des co-expressions du domaine D4 et de la SrtC-2 a révélé des résultats surprenants (Fig.68, B). Lorsque le domaine D4 est exprimé seul, une bande de masse moléculaire apparente de 17 kDa est détectée correspondant, comme attendu, au monomère purifié (Fig.68, B ligne 1). Cette bande est également visible lorsque RrgB est co-exprimé avec le domaine D4 (Fig.68, B ligne 4). En revanche, lorsque le domaine D4 est co-exprimé avec la SrtC-2, des espèces multimériques de D4 sont détectées, présentant des masses moléculaires apparentes de 34, 51 et 68 kDa, correspondant à des formes dimériques, trimériques et tétramériques (Fig.68, B ligne 5), suggérant que la SrtC-2 est capable de multimériser le domaine D4. Des bandes intermédiaires de masses moléculaires apparentes de 42, 59 et 76 kDa sont également révélées



par les anticorps anti-His-Tag (figure X, B ligne 5) ainsi que par des anticorps polyclonaux anti-SrtC-2 (non montré) indiquant que ceux-ci correspondent aux complexes acyl-enzymes intermédiaires. L'oligomérisation du domaine D4 par la SrtC-2 a également lieu en présence de RrgB, mais avec une efficacité moindre (Fig.68, ligne 7). Ces résultats indiquent que (i) la SrtC-2 est capable de multimériser D4, (ii) la piline RrgB entre en compétition avec D4 pour la reconnaissance par la SrtC-2.

### **II.2.8. SrtC-2 multimériser RrgA**

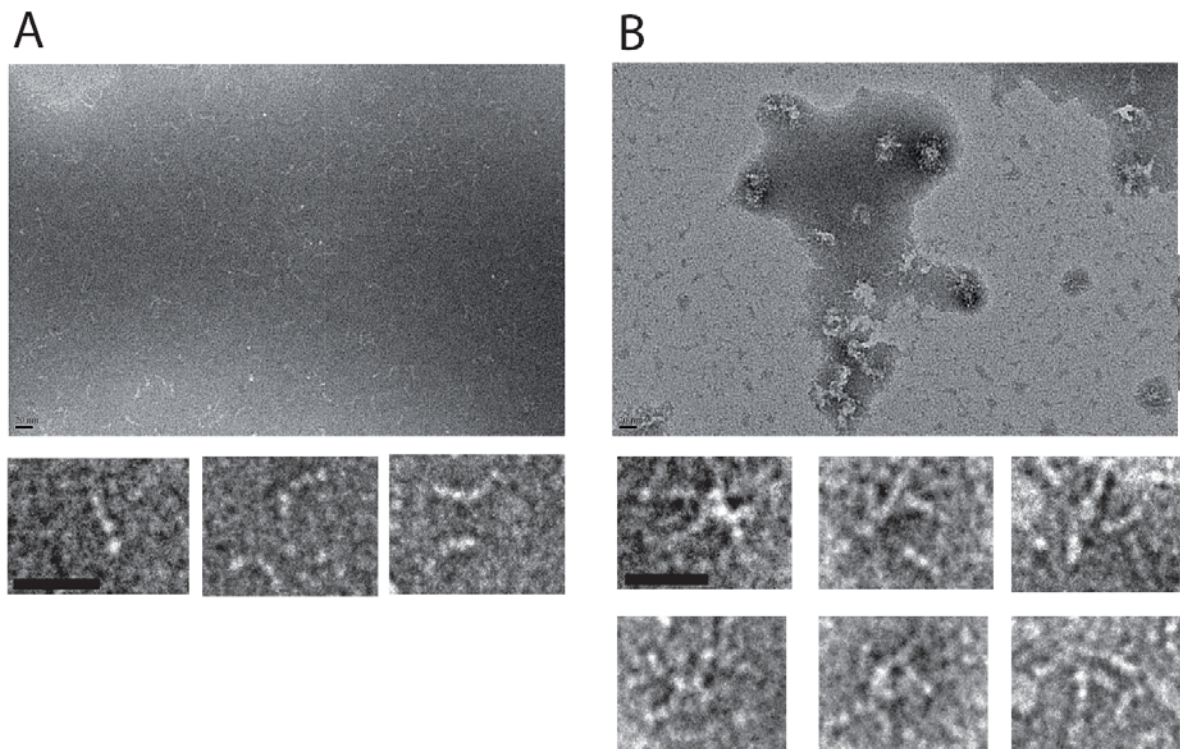
Etant donné que la SrtC-2 est capable de multimériser le domaine D4 de RrgA lorsque ces deux protéines sont co-exprimées, nous avons voulu vérifier si la SrtC-2 peut également prendre en charge la protéine RrgA complète dans cette réaction d'oligomérisation. Les produits issus de la co-expression HisRrgA et SrtC-2 ont été purifiés par une première étape de chromatographie d'affinité puis dans un second temps par filtration sur gel (Fig.69 A). Le profil d'élution de cette seconde chromatographie permet de distinguer trois pics. Le premier pic est élué dans le volume mort de la colonne et correspond à des formes multimériques de RrgA, ce qui est d'ailleurs confirmé par l'analyse en SDS-PAGE (Fig.69, B). Le second pic correspond au monomère de RrgA et le troisième pic à la SrtC-2.



**Figure 69: SrtC-2 multimérise RrgA.** **A.** Filtration sur gel des produits purifiés issus de la co-expression RrgA SrtC-2. Une colonne Superdex 200 10/300GL a été utilisée pour réaliser la chromatographie d'exclusion de taille. **B.** Analyse par SDS-PAGE 7.5% d'acrylamide des fractions d'élution post gel filtration. (PM : poids moléculaires, L : échantillon chargé sur la colonne, les numéros correspondent aux différentes fractions d'élution) (El Mortaji *et al.* en préparation).

Un échantillon des formes monomériques de RrgA (fraction 7 de la colonne de filtration sur gel) a été analysé par MET en coloration négative par Daphna Fenel à l'IBS (Fig.70 A). Les clichés révèlent que la solution protéique contient une population homogène de petits filaments de RrgA, similaires à ceux observés précédemment lors de l'étude de RrgA (Fig.43 chapitre I.1.4 des résultats).

Un échantillon de la fraction 3 issue de la colonne de filtration sur gel et correspondant aux formes multimériques de RrgA a été observé dans les mêmes conditions (Fig.70, B). L'analyse des clichés indique que la population est hétérogène et contient de nombreux agrégats. Cependant, des objets d'aspect « étoilé » sont identifiables dans chacun des clichés (Fig.70, B). Ces structures présentent toutes un « centre de nucléation » duquel émergent 4 à 5 filaments de RrgA. Au regard des résultats obtenus sur la multimérisation du domaine D4, nous pouvons supposer que ces formes étoilées multimériques de RrgA résultent de l'activité catalytique de la SrtC-2.



**Figure 70: Analyse par MET en coloration négative des formes monomériques et multimériques de RrgA.** A. Clichés de MET en coloration négative des formes monomériques de RrgA. (En haut : image correspondant au champs complet ; en bas : galerie d'images élargies) B. Clichés de MET en coloration négative des formes multimériques de RrgA (En haut : image correspondant au champs complet ; en bas : galerie d'images élargies). Pour A et B la barre d'échelle représente une distance de 20nm (El Mortaji *et al.* en préparation).

## II.2.9 Conclusions concernant l'activité de la SrtC-2

### Complexe RrgA-RrgB

Nous avons mis en évidence que la SrtC-2 est capable de générer deux complexes covalents différents faisant intervenir la piline majeure: RrgA-RrgB et RrgB-RrgC. Afin de mieux comprendre les spécificités de reconnaissance des substrats par cette sortase, nous avons procédé aux délétions des motifs LPxTG des pilines et mis en évidence que la SrtC-2 reconnaît deux motifs différents: YPRTG de RrgA pour la catalyse du complexe RrgA-RrgB

et le motif IPQTG de RrgB pour la catalyse du complexe RrgB-RrgC. Cependant, lorsque les pilines RrgA, RrgB et RrgC sont co-exprimées avec la SrtC-2, c'est le complexe RrgA-RrgB qui est préférentiellement formé, indiquant que le motif YPRTG de RrgA est reconnu avec une meilleure affinité que celui de RrgB. Cette hypothèse a été vérifiée lors des expériences d'inversion des motifs LPxTG montrant sans équivoque que le motif YPRTG de RrgA est le substrat privilégié de SrtC-2. Ce point est également renforcé par la mise en évidence de l'intermédiaire acyl-enzyme SrtC-2\*RrgA.

Afin de caractériser d'un point de vue structural le complexe RrgA-RrgB, nous avons recherché quel résidu lysine de RrgB pouvait jouer le rôle de nucléophile dans la réaction de transpeptidation. Les Lys162 et Lys183 de RrgB, qui étaient des candidates nucléophiles, ont été individuellement mutées en alanine et les formes variantes introduites dans le système de co-expression. Seule la mutation de la Lys183, qui fait partie du motif piline, réduit le taux de formation du complexe RrgA-RrgB par la SrtC-2, suggérant que ce résidu porte le groupement nucléophile privilégié dans la réaction de transpeptidation de RrgA-RrgB. Enfin, nous avons étudié l'implication des ponts intramoléculaires des pilines RrgA et RrgB dans la formation du complexe RrgA-RrgB et montré que l'absence des ponts dans les deux pilines n'empêche pas la formation du complexe, sauf en ce qui concerne la mutation D241A dans RrgB qui présente un effet légèrement délétère. Les différentes étapes conduisant la SrtC-2 à catalyser la formation du complexe RrgA-RrgB sont résumées ci-dessous (Fig.71).

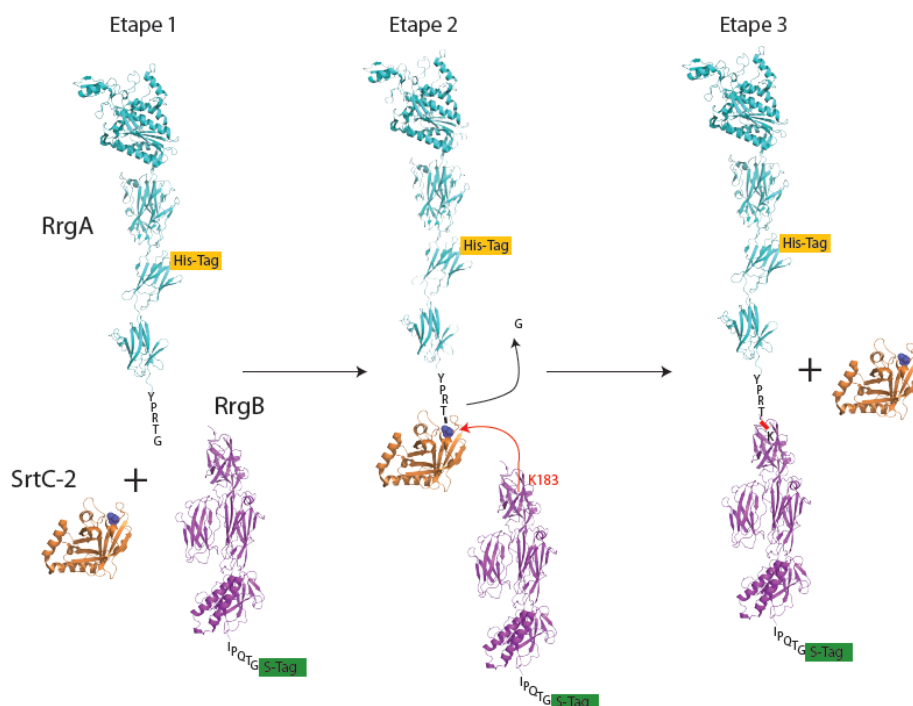


Figure 71: Récapitulatif de la catalyse du complexe RrgA-RrgB par SrtC-2.

## Multimérisation de RrgA

L'utilisation du domaine isolé D4 de RrgA dans le système de co-expression a permis de valider sa prise en charge par la SrtC-2 lors de la catalyse du complexe D4-RrgB. Ces résultats apportent deux informations majeures : (i) le domaine D4 est suffisamment bien structuré pour être pris en charge par la SrtC-2, et (ii) les autres régions de la piline RrgA (D1-D3) ne sont pas impliquées dans la réaction de transpeptidation médiée par cette sortase.

Nous avons également mis en évidence que la SrtC-2 présente une activité enzymatique de multimérisation du domaine D4. Ces résultats suggèrent que la SrtC-2 possède deux sites de reconnaissance dans le domaine D4 : le motif YPRTG et un résidu portant le groupement nucléophile accepteur. L'identification de ce résidu est en cours d'investigation. En effet, il est possible de purifier des quantités importantes de formes dimériques covalentes de D4 (de l'ordre de 1 à 5 mg par litre de culture) et l'analyse de la liaison peptidique est en cours de réalisation par MS/MS. Au vu de ces données, nous avons vérifié si la piline entière RrgA pouvait être également multimérisée. La purification de formes de hauts poids moléculaires issues de la co-expression de RrgA et SrtC-2 et leur analyse par MET en coloration négative atteste de la présence de formes étoilées correspondant aux monomères de RrgA multimérisés via leur domaine D4 (Fig.72).

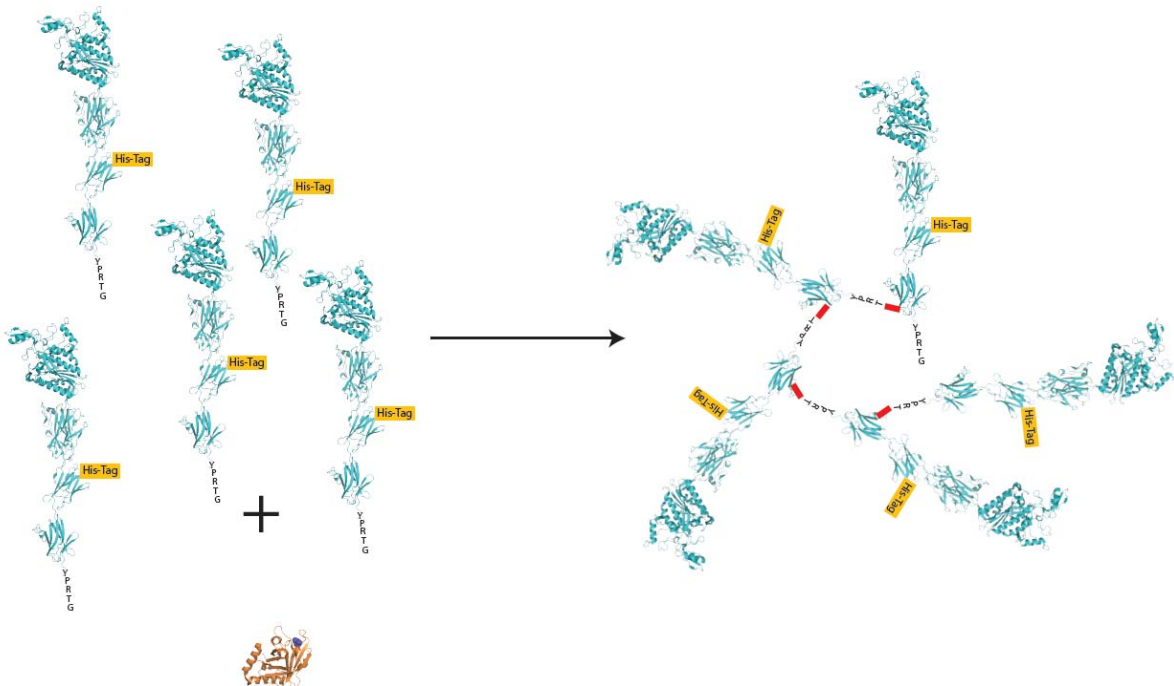
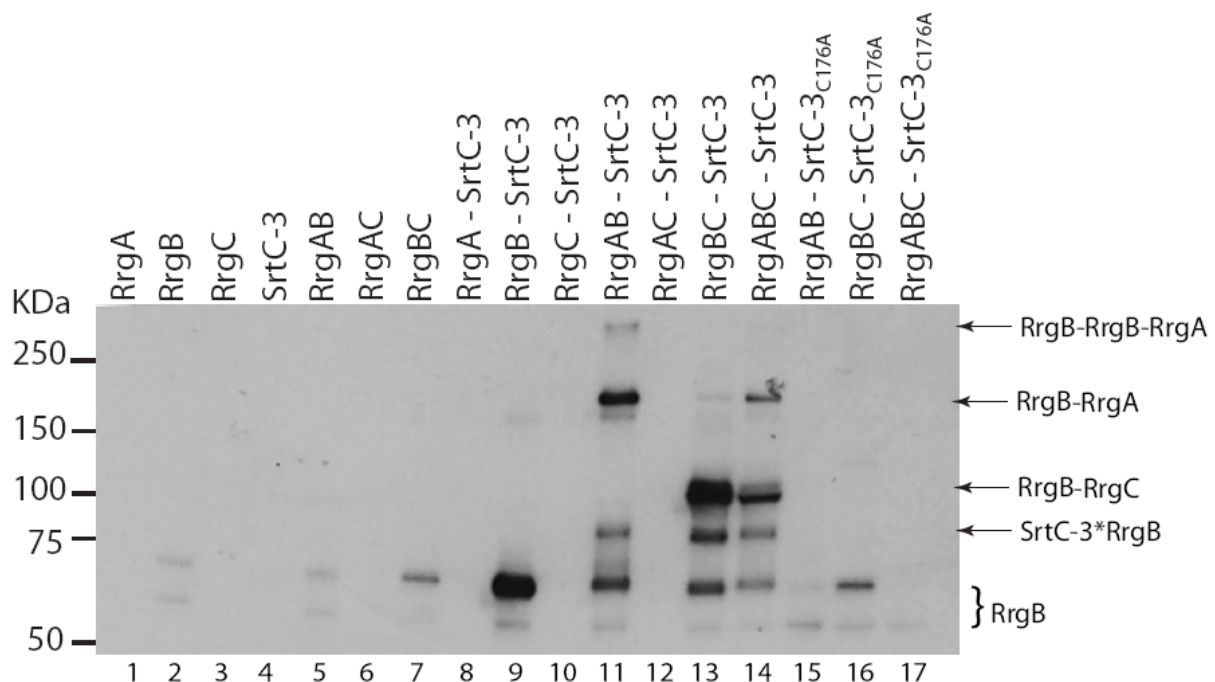


Figure 72: Récapitulatif de la multimérisation de RrgA par SrtC-2.

## II.3 Fonction de la SrtC-3

### II.3.1. La SrtC-3 catalyse deux complexes: RrgB-RrgA et RrgB-RrgC

La dernière sortase de l'opéron PI-1 de *S. pneumoniae*, SrtC-3, a été introduite dans le système de co-expression afin de déterminer sa fonction. Après avoir réalisé chacune des co-expressions impliquant RrgA, RrgB, RrgC et la SrtC-3, tous les produits issus de ces co-expressions ont été purifiés en parallèle sur le système His-Trap<sup>TM</sup> HP columns. Les protéines éluées ont été séparées par SDS-PAGE gradient 4-12,5%, puis après transfert sur membrane de nitrocellulose, ont été immunodétectées par les anticorps polyclonaux anti-RrgB (Fig.73).



**Figure 73: Catalyse par SrtC-3 des complexes covalents RrgB-RrgA et RrgB-RrgC.** Les co-expressions de RrgA, RrgB, RrgC et SrtC-3 ont été réalisées chez *E. coli*. Les purifications ont toutes été réalisées en parallèle sur le système His-Trap<sup>TM</sup> HP columns et seules les pilines accessoires RrgA et RrgC présentent une étiquette poly-histidine à leur extrémité N-terminale. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps anti-RrgB polyclonaux (El Mortaji *et al.* en préparation).

De façon similaire à ce que nous avons observé durant l'analyse des fonctions des SrtC-1 et SrtC-2, la protéine RrgB n'est pas immuno-détectée dans les éluations issues des surexpressions de RrgA, RrgC et SrtC-3 (Fig.73, lignes 1, 3 et 4) et les deux bandes caractéristiques de la liaison non spécifique du monomère de RrgB (60 et 70 kDa) sur les colonnes His-Trap sont à nouveau détectées avec une intensité plus ou moins importante selon la co-expression considérée (Fig.73, lignes 2, 5, 7, 9, 11,13, 14, 15, 16 et 17).

La co-expression de la SrtC-3 avec les pilines RrgA et RrgB montre la présence de deux espèces détectées par les anticorps anti-RrgB (Fig.73, ligne 11) et anti-RrgA (non montré) : une bande intense de masse moléculaire apparente de 200 kDa et une bande moins intense

migrant à une masse supérieure à 250 kDa. Lorsque la cystéine catalytique de SrtC-3 est mutée en alanine, la formation de ces complexes est abolie (Fig.73, ligne 15). Ceci suggère que la SrtC-3 est capable de lier covalentement RrgB et RrgA dans un complexe 1 :1 (espèce de 200kDa) et dans un complexe de stoechiométrie probablement différente (> 250kDa).

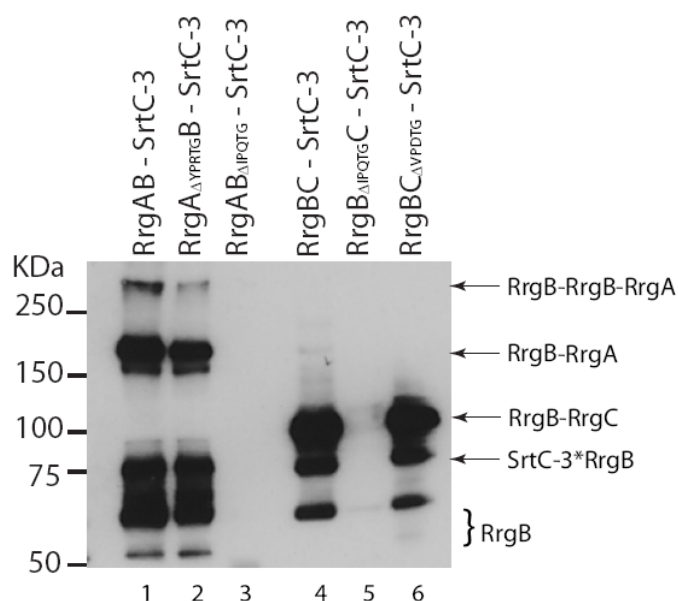
La co-expression de la SrtC-3 avec les pilines RrgB et RrgC conduit à la formation d'une bande intense, immuno-réactive avec les anticorps anti-RrgB, de masse moléculaire apparente de 100 kDa (Fig.73, lignes 13 et 14). Ce complexe covalent semble une nouvelle fois correspondre à RrgB-RrgC, produit également par les SrtC-1 et SrtC-2 (Fig.73, lignes 13 et 14). La formation de ce complexe est abolie lorsque la co-expression est réalisée avec la SrtC-3 mutée au niveau de sa cystéine catalytique (Fig.73, ligne 16). Lorsque les trois pilines, RrgA, RrgB et RrgC sont co-exprimées avec la SrtC-3, les deux complexes RrgB-RrgA et RrgB-RrgC sont formés, contrairement au complexe de masse moléculaire supérieure à 250 KDa qui n'est pas visible (Fig.73, 14).

Une dernière bande immuno-réactive avec les anticorps anti-RrgB, de masse moléculaire apparente de 75kDa est identifiée dans toutes les co-expressions menant à la formation des complexes covalents (Fig.73, lignes 11, 13, 14). Celle-ci est également révélée par l'utilisation d'anticorps polyclonaux anti-SrtC-3 suggérant qu'elle correspond à un complexe acyl-enzyme SrtC-3\*RrgB.

En conclusion, la SrtC-3 est capable de catalyser deux complexes covalents, RrgB-RrgA et RrgB-RrgC, ainsi qu'un complexe contenant RrgB et RrgA mais dont la stoechiométrie reste à déterminer. Cette même sortase forme également un complexe acyl-enzyme stable avec RrgB suggérant que les complexes covalents résultent de la prise en charge du motif IPQTG de RrgB par la SrtC-3. Cette observation laisse supposer que le complexe RrgB-RrgA de masse supérieure à 250kDa puisse correspondre à un dimère covalent de RrgB, lui-même associé à un monomère de RrgA.

### **II.3.2. La SrtC-3 reconnaît le motif IPQTG de RrgB**

Afin de confirmer la proposition selon laquelle la SrtC-3 reconnaît le motif IPQTG de RrgB, nous avons réalisé des expériences de délétion des pentapeptides LPxTG de chacune des pilines, les protéines mutées résultantes ont été introduites dans le système de co-expression (Fig.74).



**Figure 74: Spécificité de SrtC-3 dans la reconnaissance des substrats.** Les motifs LPxTG des protéines RrgA, RrgB et RrgC ont été délévés, et chaque mutant a été testé dans sa capacité à être reconnu est pris en charge par SrtC-3. Les purifications des protéines co-exprimées, RrgA, RrgB, RrgC et SrtC-3 ont été réalisées sur les colonnes His-Trap™ HP. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur membrane de nitrocellulose et immunodétectées par des anticorps polyclonaux anti-RrgB (El Mortaji *et al.* en préparation).

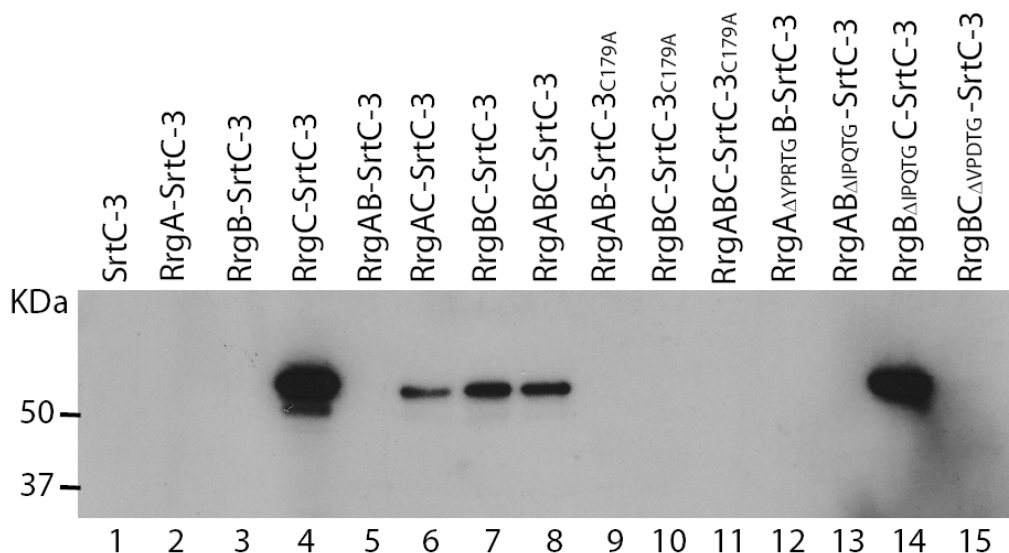
La délétion du motif YPRTG de RrgA n'a d'effet ni sur la formation du complexe RrgB-RrgA ni sur la formation du complexe de plus haut poids moléculaire, contrairement à la délétion du motif IPQTG de RrgB qui empêche toute formation des complexes (Fig.74, lignes 2 et 3). En ce qui concerne la catalyse du complexe RrgB-RrgC, la délétion du motif IPQTG inhibe aussi totalement la formation de ce complexe alors que celle du motif VPDTG de RrgC n'a aucun effet (Fig.74, lignes 5 et 6). L'ensemble de ces résultats indique que la SrtC-3 reconnaît uniquement le motif IPQTG de RrgB pour catalyser les trois complexes RrgB-RrgA, RrgB-RrgC et RrgB-RrgB-RrgA.

### II.3.3. La SrtC-3 forme un complexe acyl-enzyme stable avec RrgC

Les protéines éluées provenant des co-expressions impliquant RrgA, RrgB, RrgC et la SrtC-3 ont été séparées par SDS-PAGE gradient 4-12,5%, puis après transfert sur membrane de nitrocellulose, ont été immunodétectées par des anticorps polyclonaux anti-SrtC-3. Une exposition longue (> 5min) du film autoradiographique révèle la présence d'une bande de masse moléculaire apparente de 75 kDa correspondant à l'acyl-enzyme SrtC-3\*RrgB qui est observée dans toutes les conditions de co-expressions menant à la formation des complexes covalents (Fig.73 et 74 ci-dessus). De façon intéressante, le même blot révélé par des anticorps anti-SrtC-3, montre la présence d'une bande d'intensité plus importante (exposition courte, < 1 min) migrant à une masse moléculaire de 55 kDa (Fig.75). Cette espèce protéique



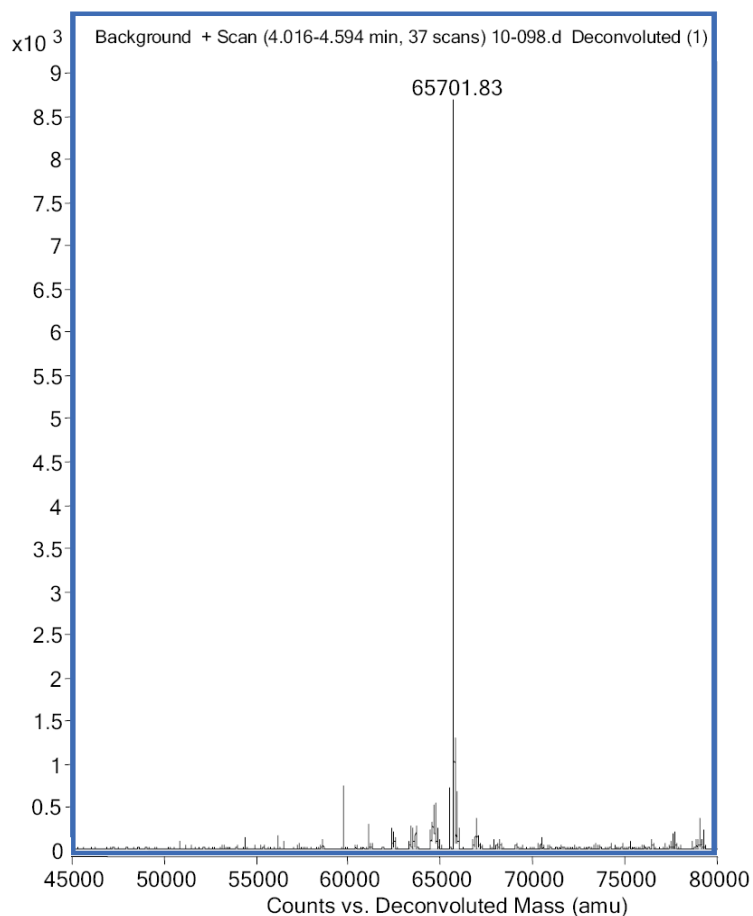
n'est visualisée que dans des co-expressions impliquant RrgC (Fig.75, lignes 4, 6, 7, 8 et 14). De plus, lorsque le motif VPDTG de RrgC est délété (Fig.75, ligne15) ou lorsque la cystéine catalytique de l'enzyme SrtC-3 est mutée en alanine (Fig.75 ligne 10 et 11) ce complexe n'est plus formé. En conclusion, ces résultats suggèrent que cette bande correspond à un complexe acyl-enzyme stable SrtC-3\*RrgC dont la formation nécessite l'intégrité du motif de reconnaissance VPDTG de RrgC et la cystéine catalytique de SrtC-3.



**Figure 75: Existence d'un intermédiaire acyl-enzyme RrgC-SrtC-3 stable.** Les co-expressions de RrgA, RrgB, RrgC et SrtC-3 ont été réalisées chez *E. coli*. Les purifications ont toutes été réalisées en parallèle sur le système His-Trap™ HP columns et seules les pilines accessoires RrgA et RrgC présentent une étiquette poly-histidine à leur extrémité N-terminale. Les protéines éluées ont été séparées par SDS-PAGE 4-12,5%, transférées sur une membrane de nitrocellulose et immunodétectées à l'aide d'anticorps anti-SrtC-3 polyclonaux (El Mortaji *et al.* en préparation).

En se basant sur le modèle de la réaction de transpeptidation médiée par les sortases, le complexe acyl-enzyme SrtC-3\*RrgC devrait présenter une masse théorique correspondant à la somme d'un monomère 6xHis-RrgC dont le motif VPDTG est clivé par la sortase après le résidu Thr (40 616,79 Da) et d'un monomère de sortase dont le groupement thiol de la cystéine a perdu son hydrogène (25 087,31 Da) : la masse théorique du complexe SrtC-3\*RrgC est donc de 65 704,10 Da.

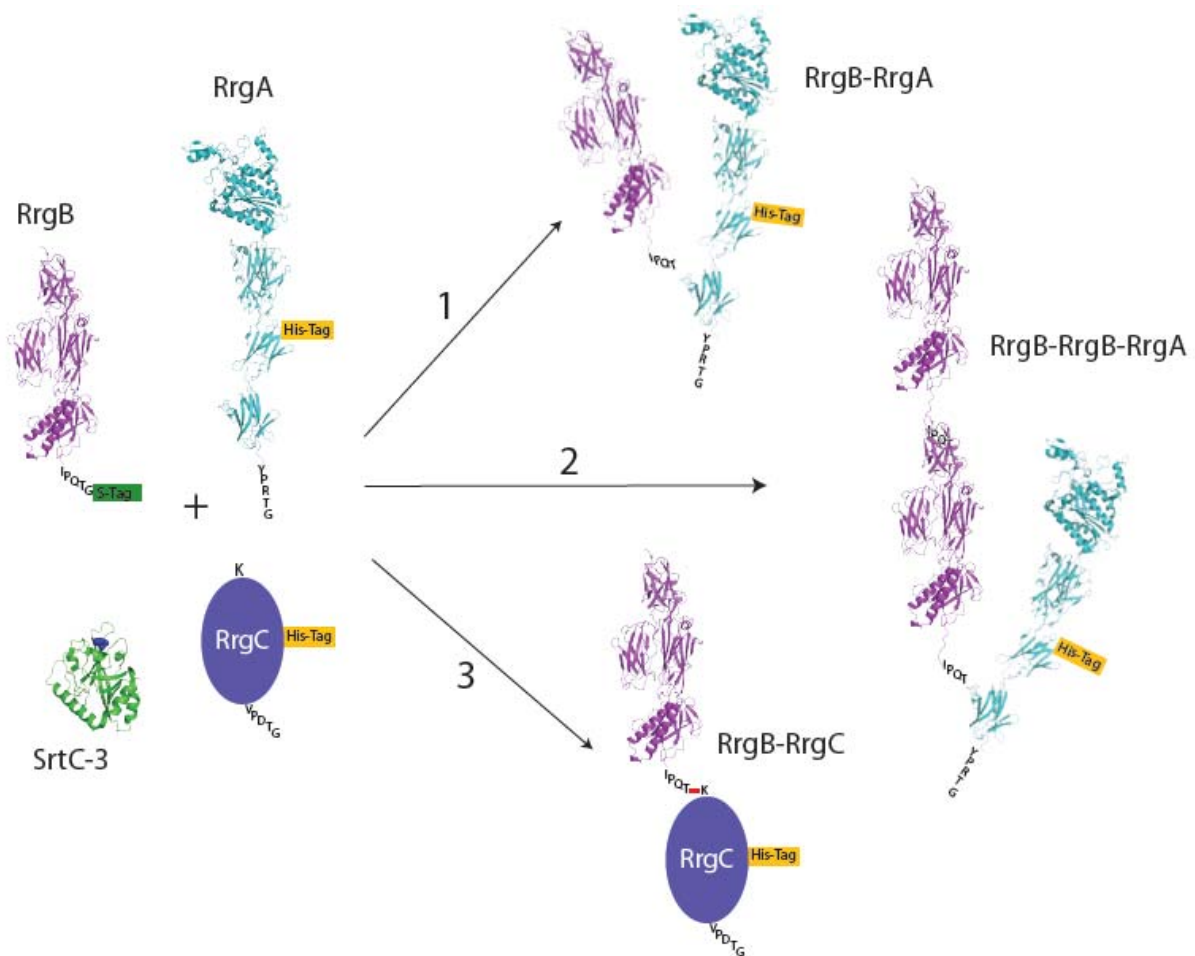
Nous avons purifié ce complexe SrtC-3\*RrgC et son analyse a été effectuée par spectrométrie de masse (Fig.76). La masse moléculaire mesurée est 65 701,83, présentant une différence de masse de 2.27 Da avec la valeur attendue. Cette différence est comprise dans la précision de mesure de la technique (1 pour 10 000 Da), validant l'identification du complexe SrtC-3\*RrgC.



**Figure 76: Analyse ESI-MS de l'acyl-enzyme SrtC-3\*RrgC.**

### II.3.4 Conclusions concernant l'activité de la SrtC-3

Nous avons montré que la SrtC-3 est capable de catalyser trois complexes différents en utilisant pour chacun d'eux le motif de reconnaissance IPQTG de RrgB (Fig.77). En effet, lorsque la SrtC-3 est co-exprimée avec RrgB et RrgA, un complexe covalent RrgB-RrgA équimolaire ainsi qu'un complexe composé d'un dimère de RrgB lié covalent à un monomère de RrgA sont générés. D'autre part, la co-expression de la SrtC-3 avec RrgB et RrgC conduit à la production du complexe RrgB-RrgC.

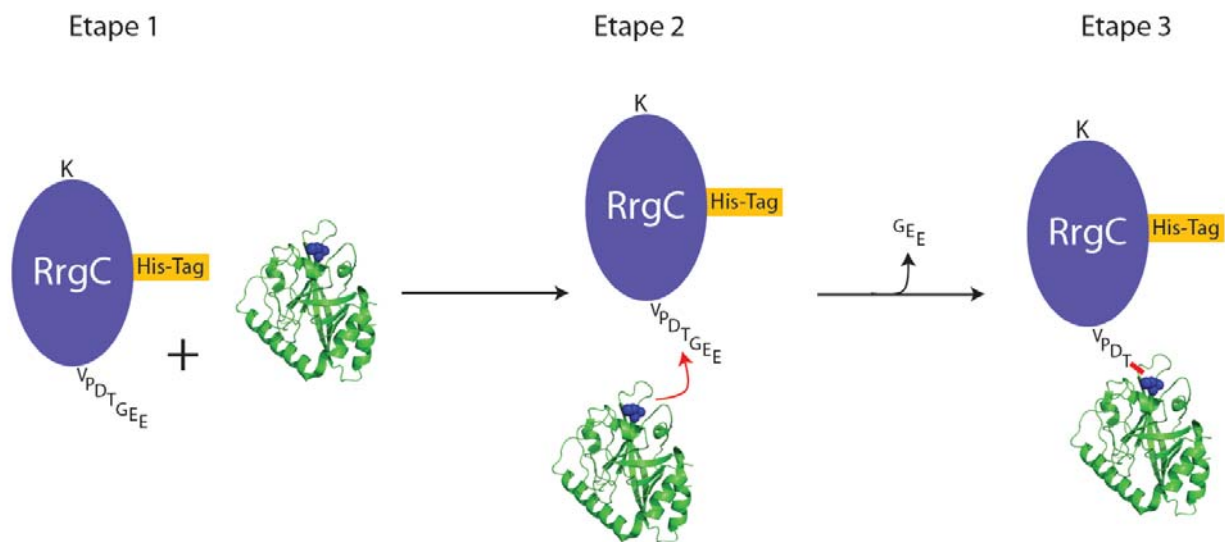


**Figure 77: Schéma récapitulatif des complexes catalysés par SrtC-3.** Les pilines RrgB\_S-tag et His\_RrgA ainsi que la sortase C-3 sont représentés par leur structure cristallographique. La piline His\_RrgC est représentée par une ellipse **1.** Représentation du complexe RrgB-RrgA catalysé par SrtC-3. La thréonine du motif IPQTG de RrgB est liée covalentement à une lysine nucléophile de RrgA qui n'a pas encore été identifiée. Néanmoins il existe une Lysine nucléophile au niveau du domaine D4 de RrgA (cf. travaux réalisés sur la sortase C-2) c'est la raison pour laquelle le modèle a été représenté ainsi. **2.** Représentation du complexe RrgB-RrgB-RrgA catalysé par SrtC-3. **3.** Représentation du complexe RrgB-RrgC catalysé par SrtC-3.

La délétion du motif IPQTG de RrgB abolit totalement la formation de ces complexes, de même que la mutation de la cystéine catalytique de la SrtC-3. De façon intéressante, l'intermédiaire acyl-enzyme SrtC-3\*RrgB est retrouvé dans toutes les co-expression menant à la formation de ces complexes indiquant qu'il est co-élué avec les pilines accessoires RrgA ou RrgC qui seules portent l'étiquette poly-histidine. Ces informations suggèrent qu'il n'existe pas de spécificité stricte de la SrtC-3 pour l'accepteur nucléophile puisque la formation de l'acyl-enzyme SrtC-3\*RrgB peut mener à l'association de RrgB avec RrgA (complexe RrgB-RrgA), ou avec RrgC (complexe RrgB-RrgC) ou encore avec RrgB (pour former RrgB-RrgB-RrgA). Ces données indiquent que la SrtC-3 reconnaît et clive le motif IPQTG de RrgB puis accepte

indifféremment chacune des trois pilines en tant qu'accepteur dans l'étape finale de la réaction de transpeptidation.

La purification et l'analyse biochimique du complexe SrtC-3\*RrgC indique que cette sortase ne présente pas de spécificité stricte pour le IPQTG de RrgB, puisqu'elle peut interagir également fortement avec le motif VPDTG de RrgC (Fig.78). La question de l'identité de l'accepteur du complexe SrtC-3\*RrgC est posé : RrgA, RrgC ou bien le lipide II si l'on suppose que cette réaction permettrait d'ancrer le pilus au peptidoglycane. Ces hypothèses sont actuellement en cours de validation.



**Figure 78: Schéma récapitulatif de la formation de l'acyl-enzyme RrgC-SrtC-3. Etape1.** His\_RrgC et la sortase C-3 sont co-exprimées chez *E. coli*. **Etape 2.** La sortase C-3 hydrolyse la liaison Thr-Gly du VPDTG de His\_RrgC qui mène au relargage des résidus GEE C-terminaux. **Etape 3.** Formation de l'acyl-enzyme covalent RrgC-SrtC-3.



# **Discussion & Perspectives**



## **DISCUSSION & PERSPECTIVES**

Les pili, exemple singulier de protéines bactériennes de surface, sont des structures exposées au milieu extracellulaire et qui donc nécessitent d'être extrêmement résistantes à toute dégradation, notamment à la protéolyse. Les pili exprimés par les bactéries Gram-positif sont particulièrement stables et résistants à la digestion protéolytique. Ces caractéristiques proviennent du mode d'assemblage unique des pilines qui sont associées entre elles de façon covalente par des enzymes spécialisées, les sortases (Ton-That 2003). Un degré supplémentaire dans la compacité et la stabilisation des pili Gram-positifs est apporté par la présence de liaisons covalentes intramoléculaires au sein des pilines (Kang 2007, Budzik 2009, Kang 2009). Depuis la première description du PI-1 chez le pneumocoque en 2006, plusieurs mécanismes de biogenèse ont été proposés conduisant à différents modèles d'architecture du pilus et à une compréhension encore très fragmentaire du rôle de chacune des sortases. Au cours de ces travaux de thèse, nous avons cherché à répondre à ces deux questions en utilisant des approches biochimiques et structurales.

### **Les ponts intramoléculaires au sein des pilines du pneumocoque**

La première partie de notre étude a porté sur l'étude structurale des pilines RrgA, RrgB et RrgC dans le but notamment d'identifier la présence de ces ponts intramoléculaires. Cette analyse a ainsi permis de mettre en évidence que chacune des pilines du pneumocoque renferme dans leur structure des ponts intramoléculaires qui ont le pouvoir d'augmenter leur stabilité thermique mais également leur stabilité face à la digestion protéolytique. De façon intéressante, la localisation des ponts dans les pilines présente un effet stabilisateur plus ou moins marqué. Prenons l'exemple du domaine D2 de RrgA, celui-ci est le résultat de l'assemblage de deux demi-domaines, N et C-terminaux de la protéine, verrouillés dans cette conformation par un pont intramoléculaire. La position clé de ce pont au sein de la structure globale de RrgA permet de comprendre l'effet déstabilisateur important observé en son absence. Dans le cas de la piline majeure, RrgB, le pont présent dans le domaine D2 semble jouer un rôle similaire au niveau de la structure globale de la piline notamment dans son rôle de stabilisation face à la digestion protéolytique. De plus, nous avons également montré que l'absence de certains ponts intramoléculaires, notamment celui du domaine D2 de RrgB, pouvait influencer l'efficacité de reconnaissance des pilines par les sortases et donc diminuer l'efficacité de la catalyse des liaisons intermoléculaires.



Un résultat surprenant est apparu lors de l'étude de la piline majeure RrgB. En effet, en solution, celle-ci ne présente que trois ponts intramoléculaires formés respectivement dans les domaines D2, D3 et D4. En revanche, la structure cristallographique de RrgB révèle la présence d'un dernier pont visible au niveau du domaine D1 de la forme cristalline de la protéine. Qu'est-ce qui permet d'expliquer cette différence majeure ? Un élément de réponse peut être proposé en se basant sur l'étude réalisée sur l'homologue de RrgB chez *B. cereus*, la piline BcpA pour laquelle la structure cristallographique montre la présence de trois ponts intramoléculaires chacun localisé dans les domaines CNA2, XNA, et CNA3 (Budzik 2009) (Fig.19, Introduction bibliographique). Le domaine CNA1 qui n'est pas inclus dans cette structure a été étudié biochimiquement et cette analyse a indiqué qu'aucun pont intramoléculaire n'était identifiable dans le domaine CNA1 de BcpA. En revanche, l'étude des pili BcpA matures, donc assemblés, a permis la mise en évidence d'un pont supplémentaire au sein de la protéine BcpA mature, localisé dans le domaine CNA1. Les auteurs proposent ainsi que les trois ponts intramoléculaires identifiés dans les domaines CNA2, XNA et CNA3 sont formés dans BcpA dès sa synthèse et avant même sa prise en charge par la sortase qui va conduire à la polymérisation de BcpA. Il semblerait donc que seule la catalyse du pont intermoléculaire Thr-Lys entre deux sous-unités successives mènerait à la formation du dernier pont intramoléculaire localisé dans le domaine CNA1. Revenons à présent sur les informations obtenues lors de l'étude de la piline RrgB. Nous avons montré qu'une fibre artificielle de RrgB est formée dans l'empilement cristallin mimant son homologue natif. De façon intéressante, le motif IPQTG d'un monomère RrgB(a) s'engouffre dans une gorge localisée au niveau du domaine D1 d'un monomère RrgB(b) dans lequel le pont intramoléculaire est formé. Dans cette conformation, la Thr du motif IPQTG (monomère a) est positionnée à proximité de deux résidus Lysines (monomère b) candidats à la transpeptidation. Nous pouvons donc supposer que la présence du motif IPQTG de RrgB dans la gorge du domaine D1 induit par sa seule présence une contrainte suffisante pour générer la formation du pont intramoléculaire dans le domaine D1. L'existence de ce pont chez les pili natifs de *S. pneumoniae* restent encore à démontrer, ce que nous nous proposons de réaliser. Ainsi, les bactéries Gram-positif semblent avoir développé des mécanismes communs permettant la formation de liaisons intramoléculaires et également intermoléculaires dans le but de générer des pili matures extrêmement résistants.

## Vers une meilleure compréhension de la fonction adhésive du pilus

La fonction d'adhésion des pili aux tissus hôtes a été démontrée chez de nombreuses bactéries, y compris chez le pneumocoque (Mandlik 2007, Abbot 2007, Manetti 2007, Nelson 2007). Pour ce dernier, et dans le cas du PI-1, c'est principalement la piline accessoire RrgA qui porte la fonction d'adhésine (Nelson 2007, Hilleringmann 2008). En effet, *in vitro*, des souches bactériennes n'exprimant plus RrgA présentent un pouvoir d'adhésion réduit d'un facteur huit sur des cellules épithéliales pulmonaires A549 par rapport à une souche sauvage (Nelson 2007). De plus, des tests d'interactions en phase solide montre qu'une forme recombinante de RrgA est capable de lier différents composants de la matrice extracellulaire tels que la fibronectine, le collagène de type I et la laminine (Hilleringmann 2008). Qu'est-ce qui permet d'expliquer qu'une telle palette d'interaction soit médiée par une seule et même protéine ? La résolution de la structure de RrgA permet de mieux comprendre ce phénomène. En effet, cette protéine modulaire présente à l'une de ces extrémités deux domaines potentiellement impliqués dans ces interactions. Le premier est le domaine D3, le plus distal, est un mélange composite de séquences d'origine eucaryote et procaryote, le tout formant un domaine structuré. Le corps de celui-ci présente une homologie structurale avec le domaine VWA humain et le domaine I de l'intégrine  $\alpha 2\beta 1$  connus pour médier l'adhérence cellulaire. Le second domaine D2, est quant à lui un homologue structural d'une région de la protéine Cna de *S. aureus* qui est impliquée dans des interaction avec le collagène 1. En outre, il n'est pas totalement exclu que les deux autres domaines D1 et D4 puissent participer à la fonction adhésive de la protéine. Il serait donc intéressant d'analyser, sur la base de la structure, l'implication de chacun des domaines dans la fonction d'adhésion de RrgA.

## Un modèle d'assemblage expliquant les résultats antérieurs contradictoires

Plusieurs modèles d'assemblage du PI-1 de *S. pneumoniae* ont été proposés ces dernières années, mais aucun d'entre eux n'a réellement fait l'unanimité (Barocchi 2006, LeMieux 2006, Hilleringmann 2008, Hilleringmann 2009). La seule information sur laquelle toutes les équipes s'accordent est que le squelette du pilus est le résultat de la polymérisation de la piline majeure RrgB. Les deux pilines accessoires RrgA et RrgC y seraient associés: à quelle localisation et dans quelle stœchiométrie? Ces interrogations restent sans réponses précises. De même, la fonction de chacune des sortases, dans la polymérisation de RrgB et dans l'incorporation des pilines accessoires, reste l'objet de discordes (LeMieux 2008,

Manzano 2008, Fälker 2008). Durant ces trois dernières années nous avons voulu apporté des réponses précises à ces questions en utilisant des approches différentes de celles qui étaient utilisées jusqu'à lors pour élucider le mécanisme de biogenèse du pilus chez *S. pneumoniae*.

### **Fonction des sortases dans la polymérisation de RrgB**

Dans un premier temps, nous avons voulu connaître le rôle de chaque sortase dans la polymérisation de RrgB. Ainsi, nous avons montré par l'utilisation de protéines recombinantes solubles, que l'incubation à 37°C pendant 16h de RrgB avec la SrtC-1 dans un ratio 1:2 conduit à la polymérisation de RrgB *in vitro* (Manzano 2008). Dans les mêmes conditions, la SrtC-3 n'est capable de générer qu'un début de fibre contrairement à la SrtC-2 qui ne présente aucune activité de polymérisation de RrgB *in vitro*. Ces résultats ont été en partie validés, *in vivo*, par la délétion de la SrtC-1 chez la souche TIGR4. Celle-ci présente alors une capacité réduite de polymérisation de RrgB. Ainsi, ces résultats suggèrent que la polymérisation de RrgB est principalement médiée par la SrtC-1 et que la SrtC-3 puisse en partie suppléer cette fonction. Cette notion de redondance dans l'activité de polymérisation de RrgB a également été soulevée dans les travaux réalisés par LeMieux en 2008 où la mutation individuelle de chacune des sortases a été étudiée dans le contexte de la souche TIGR4 (LeMieux 2008). Chacune des souches mutantes, dont l'une des sortases a été inactivée, demeure capable de produire un polymère de RrgB. Il semble que la nécessité de produire une fibre de RrgB soit suffisamment importante pour que la bactérie produise des enzymes présentant une redondance fonctionnelle permettant ainsi de suppléer la déficience de l'une d'entre elles.

### **Fonction des sortases dans l'incorporation des pilines accessoires**

En ce qui concerne l'incorporation des pilines accessoires et afin de tenter d'élucider le mécanisme de biogenèse du PI-1, nous avons développé une plateforme de co-expression chez *E. coli* permettant de produire chacune des trois pilines et des trois sortases dans toutes les combinaisons possibles. Ce système a ainsi permis de mieux comprendre les mécanismes moléculaires d'action de chaque sortase.

Dans le cas de la piline accessoire RrgA, sa liaison à RrgB peut être médiée par deux sortases différentes. Nous avons montré que la SrtC-3 catalyse un complexe RrgB-RrgA, via le motif IPQTG de RrgB, mais aussi que la SrtC-2 catalyse un complexe RrgA-RrgB via cette fois-ci le motif YPRTG de RrgA. Ces données indiquent donc que RrgA peut être localisée à l'extrémité distale d'une fibre de RrgB, résultat de l'activité de la SrtC-2 et que l'incorporation

de RrgA dans une fibre de RrgB est aussi possible, résultat des activités synergiques des SrtC-2 et SrtC-3. En se référant à la littérature, les équipes de Fälker et de LeMieux étaient en désaccord quant au rôle des sortases dans l'incorporation de RrgA (Fälker 2008, LeMieux 2008). La première a montré que la SrtC-2 permettait l'association de RrgA au pilus, alors que la seconde proposait que les formes RrgA<sub>HMW</sub> nécessitent l'activité de la SrtC-3. Les bases moléculaires de reconnaissance de chacune des SrtC-2 et SrtC-3 obtenues dans notre étude permettent d'expliquer ces deux résultats qui semblaient *a priori* contradictoires mais sont en réalité complémentaires. Par ailleurs, au vu de ces nouvelles données, le modèle d'assemblage proposé par Hilleringmann en 2009 devient incomplet puisque l'incorporation de RrgA dans la fibre de RrgB est moléculairement possible (Hilleringmann 2009). De plus, nous avons montré dans cette étude que la SrtC-2 catalyse la multimérisation de RrgA. La nature catalytique de la formation de ces multimères a été prouvée par l'inactivation de la cystéine de SrtC-2. Néanmoins, la question de la pertinence physiologique de ce multimère peut être posée. En se référant à la littérature une fois de plus, deux modèles différents d'assemblage du PI-1 ont proposé l'existence de groupements de RrgA localisés tout au long de la fibre (LeMieux 2006, Hilleringmann 2008) pouvant correspondre à des formes multimériques associées entre elles de RrgA telles que celles formés dans notre système *in vitro*. D'autre part, l'existence de tels multimères de RrgA situés à extrémité distale du pilus permettrait d'une part de faciliter l'adhérence aux cellules hôtes et d'autre part d'augmenter l'avidité des interactions médiées par les pili.

Concernant l'incorporation de la piline accessoire RrgC, les expériences de délétions génétiques obtenues par l'équipe de Fälker montraient que seule la SrtC-1 est capable d'incorporer RrgC aux pili (Fälker 2008). D'après les travaux de LeMieux (LeMieux 2008), les SrtC-1 et SrtC-3 sont capable d'accomplir cette fonction. Nous sommes en mesure une nouvelle fois d'expliquer ces résultats. En effet, nos données indiquent que les trois sortases sont capable d'incorporer RrgC aux pili, car capables de lier RrgB à RrgC via la reconnaissance du motif IPQTG de RrgB. Néanmoins, de part la spécificité de reconnaissance inhérente à chacune des sortases, SrtC-2 semble la moins favorable à la catalyse de ce complexe. En effet, nous avons montré que SrtC-2 privilégie la piline RrgA comme substrat plutôt que RrgB. Nous avons également montré que contrairement à la sortase C-1 qui ne présente qu'une spécificité stricte pour RrgB, la sortase C-3 peut également avoir pour substrat RrgC. Ainsi, la sortase C-1 semble être l'enzyme principalement responsable de la catalyse du complexe RrgB-RrgC, et la sortase C-3 une nouvelle fois pourrait suppléer SrtC-1 dans cette fonction. A partir de ces nouvelles données moléculaires, nous pouvons proposer

un modèle d'assemblage du pilus en adéquation avec des données antérieures obtenues *in vivo* et donc attribuer à chacune des sortases un rôle spécifique dans la biogenèse du PI-1 (Fig.79).

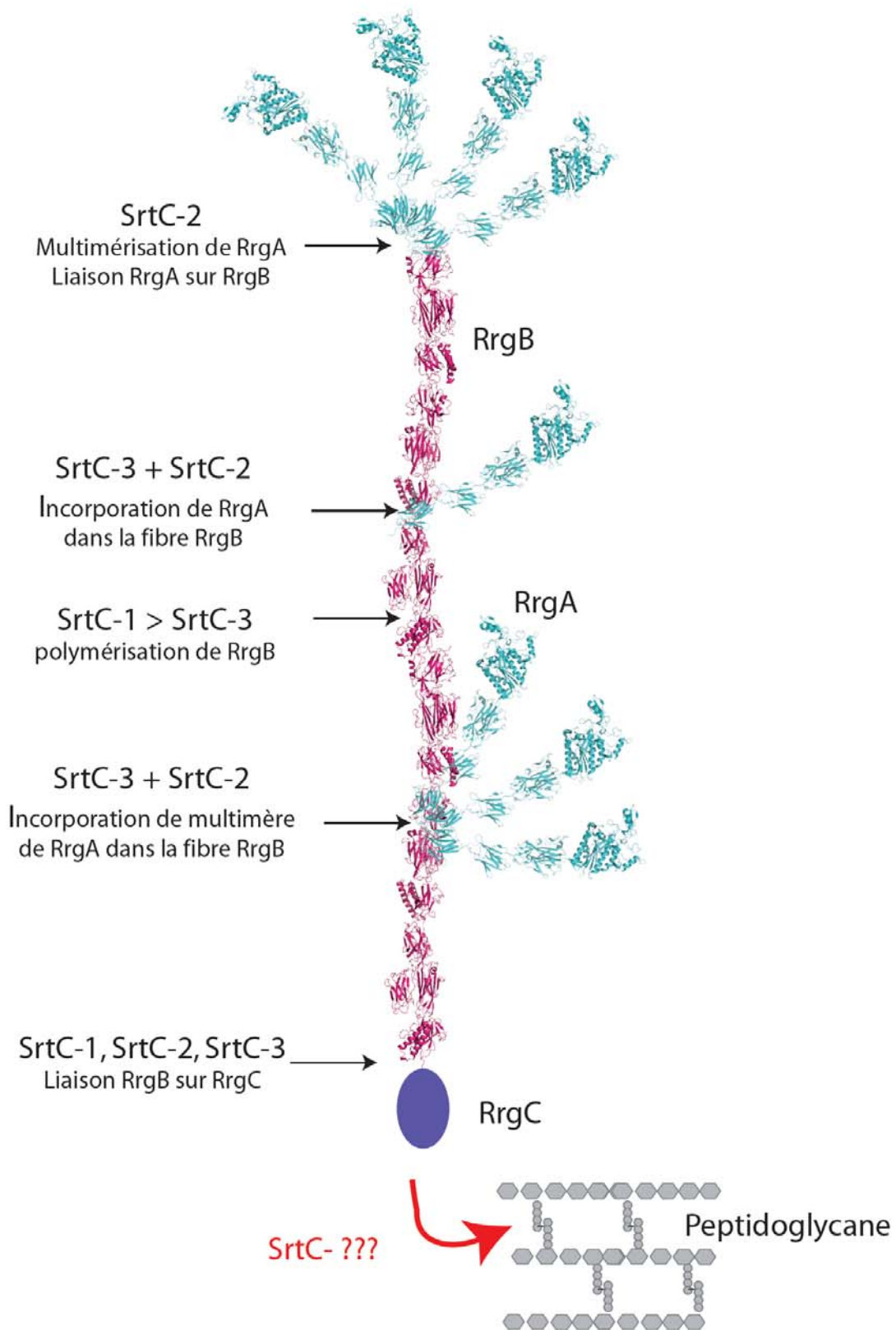


Figure 79: Modèle d'assemblage du pilus PI-1 de *S. pneumoniae*

## Les perspectives

Néanmoins, une information essentielle reste à déterminer: comment les pili sont-ils liés covalentement au peptidoglycane? Contrairement aux autres bactéries Gram-positif, chez le pneumocoque, la sortase ménagère SrtA, n'est nécessaire ni à l'assemblage du pilus, ni à son ancrage à la paroi bactérienne (LeMieux 2008). Cette information indique donc qu'au moins l'une des sortases de classe C est en charge de l'ancrage du PI-1 au peptidoglycane. De plus, les travaux de Hilleringmann menés en 2009 indique que chez une souche TIGR4*ArrgC*, les pili, qui ne présente plus cette piline accessoire, se détachent plus facilement de la surface bactérienne et sont retrouvés en quantité plus importante dans le milieu de culture (Hilleringmann 2009). Ces résultats, associés à la localisation basale de cette piline au sein de la fibre, laissent supposer que RrgC a une fonction d'ancrage des pili. Par ailleurs dans notre système de co-expression, seule la sortase C-3 est capable de prendre en charge le motif VPDTG de RrgC, générant ainsi un acyl-enzyme SrtC-3\*RrgC intermédiaire. Mais aucun complexe impliquant le motif VPDTG de RrgC n'a encore été identifié. En réalisant un bilan de ces données, l'hypothèse la plus probable est que l'accepteur de cette réaction de transpeptidation soit le lipide II, précurseur du peptidoglycane, et que son absence dans les co-expressions, bloque la réaction de transpeptidation à l'étape d'acylation. C'est précisément cette hypothèse que nous nous proposons de vérifier dans les prochaines semaines.

Il est important de noter que le système de co-expression que nous avons développé nous a permis d'obtenir un nombre important d'informations qualitatives concernant les spécificités de reconnaissance des sortases pour leurs substrats ainsi que pour leurs accepteurs. Néanmoins, les quantités de complexes obtenues par cette méthode restent très faibles (souvent inférieur à 500 µg par litre de culture). Ce phénomène est probablement directement relié à l'utilisation de constructions solubles. En effet, il est important de rappeler que chez la bactérie *S. pneumoniae* TIGR4, toutes ces protéines sont membranaires. Ceci implique que les concentrations locales ainsi que l'orientation des protéines sont optimales pour favoriser la catalyse par les sortases. Nous avons déjà observé ce phénomène dans les tests de reconstitution de la fibre RrgB *in vitro* qui nécessitait d'utiliser des protéines solubles dans de très faibles volumes (40 µl). Nous pensons que c'est également pour la même raison que nous n'arrivons pas à générer des objets de stœchiométrie plus importante que les hétérotrimères obtenus en co-expression avec SrtC-3.

Par ailleurs, ce système s'est avéré suffisamment efficace pour envisager une analyse plus poussée de certains complexes. C'est le cas du complexe RrgB-RrgC catalysé par SrtC-1 qui a pu être produit en quantité et en pureté suffisante pour réaliser des premiers tests de cristallogénèse encourageants. De même, nous espérons prochainement pouvoir purifier le complexe acyl-enzyme RrgC\*SrtC-3 en quantité suffisante pour envisager une analyse structurale. Celle-ci permettrait notamment de mieux comprendre les spécificités de reconnaissance de chacune des sortases et également de déterminer l'implication des couvercles (spécifiques des sortases de classe C) dans ce mécanisme. Maintenant que nous connaissons les spécificités de substrats de chacune des sortases, nous pouvons envisager d'analyser les paramètres cinétiques de ces enzymes.

Les données recueillies dans la bibliographie semblent valider nos résultats. Néanmoins, nous envisageons de les confirmer, *in vivo*, au laboratoire durant les prochains mois.

# **Références bibliographiques**





## REFERENCES BIBLIOGRAPHIQUES

- Abbot, E.L., Smith, W.D., Siou, G.P., Chiriboga, C., Smith, R.J., Wilson, J.A., Hirst, B.H., and Kehoe, M.A. (2007). Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. *Cellular microbiology* 9, 1822-1833.
- Aguiar, S.I., Serrano, I., Pinto, F.R., Melo-Cristino, J., and Ramirez, M. (2008). The presence of the pilus locus is a clonal property among pneumococcal invasive isolates. *BMC microbiology* 8, 41.
- Attali, C., Durmort, C., Vernet, T., and Di Guilmi, A.M. (2008). The interaction of *Streptococcus pneumoniae* with plasmin mediates transmigration across endothelial and epithelial monolayers by intercellular junction cleavage. *Infection and immunity* 76, 5350-5356.
- Attali, C., Frolet, C., Durmort, C., Offant, J., Vernet, T., and Di Guilmi, A.M. (2008). *Streptococcus pneumoniae* choline-binding protein E interaction with plasminogen/plasmin stimulates migration across the extracellular matrix. *Infection and immunity* 76, 466-476.
- Avery, O.T., and Morgan, H.J. (1925). Immunological Reactions of the Isolated Carbohydrate and Protein of *Pneumococcus*. *The Journal of experimental medicine* 42, 347-353.
- Bae, T., and Schneewind, O. (2003). The YSIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing. *Journal of bacteriology* 185, 2910-2919.
- Bagnoli, F., Moschioni, M., Donati, C., Dimitrovska, V., Ferlenghi, I., Facciotti, C., Muzzi, A., Giusti, F., Emolo, C., Sinisi, A., et al. (2008). A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. *Journal of bacteriology* 190, 5480-5492.
- Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Massignani, V., et al. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. *Proceedings of the National Academy of Sciences of the United States of America* 103, 2857-2862.
- Bender, M.H., Cartee, R.T., and Yother, J. (2003). Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae*. *Journal of bacteriology* 185, 6057-6066.
- Bergmann, S., Rohde, M., and Hammerschmidt, S. (2004). Glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pneumoniae* is a surface-displayed plasminogen-binding protein. *Infection and immunity* 72, 2416-2419.
- Berry, A.M., Yother, J., Briles, D.E., Hansman, D., and Paton, J.C. (1989). Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infection and immunity* 57, 2037-2042.
- Blumberg, P.M., and Strominger, J.L. (1972). Five penicillin-binding components occur in *Bacillus subtilis* membranes. *The Journal of biological chemistry* 247, 8107-8113.
- Boel, G., Jin, H., and Pancholi, V. (2005). Inhibition of cell surface export of group A streptococcal anchorless surface dehydrogenase affects bacterial adherence and antiphagocytic properties. *Infection and immunity* 73, 6237-6248.
- Bootsma, H.J., Egmont-Petersen, M., and Hermans, P.W. (2007). Analysis of the in vitro transcriptional response of human pharyngeal epithelial cells to adherent *Streptococcus pneumoniae*: evidence for a distinct response to encapsulated strains. *Infection and immunity* 75, 5489-5499.

- Bourgogne, A., Singh, K.V., Fox, K.A., Pflughoeft, K.J., Murray, B.E., and Garsin, D.A. (2007). EbpR is important for biofilm formation by activating expression of the endocarditis and biofilm-associated pilus operon (ebpABC) of *Enterococcus faecalis* OG1RF. *Journal of bacteriology* *189*, 6490-6493.
- Brandileone, M.C., Casagrande, S.T., Guerra, M.L., Zanella, R.C., Andrade, A.L., and Di Fabio, J.L. (2006). Increase in numbers of beta-lactam-resistant invasive *Streptococcus pneumoniae* in Brazil and the impact of conjugate vaccine coverage. *Journal of medical microbiology* *55*, 567-574.
- Braun, J.S., Sublett, J.E., Freyer, D., Mitchell, T.J., Cleveland, J.L., Tuomanen, E.I., and Weber, J.R. (2002). Pneumococcal pneumolysin and H<sub>2</sub>O<sub>2</sub> mediate brain cell apoptosis during meningitis. *The Journal of clinical investigation* *109*, 19-27.
- Briles, E.B., and Tomasz, A. (1973). Pneumococcal Forssman antigen. A choline-containing lipoteichoic acid. *The Journal of biological chemistry* *248*, 6394-6397.
- Brisou, P., T. Gaillard, Y. Muzellec (2004). Infections à pneumocoque. In *Encyclopédie Médico-chirurgicale*. (France).
- Budzik, J.M., Marraffini, L.A., and Schneewind, O. (2007). Assembly of pili on the surface of *Bacillus cereus* vegetative cells. *Molecular microbiology* *66*, 495-510.
- Budzik, J.M., Marraffini, L.A., Souda, P., Whitelegge, J.P., Faull, K.F., and Schneewind, O. (2008). Amide bonds assemble pili on the surface of bacilli. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 10215-10220.
- Budzik, J.M., Oh, S.Y., and Schneewind, O. (2008). Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. *The Journal of biological chemistry* *283*, 36676-36686.
- Budzik, J.M., Oh, S.Y., and Schneewind, O. (2009). Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili. *The Journal of biological chemistry* *284*, 12989-12997.
- Budzik, J.M., Poor, C.B., Faull, K.F., Whitelegge, J.P., He, C., and Schneewind, O. (2009). Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 19992-19997.
- Canvin, J.R., Marvin, A.P., Sivakumaran, M., Paton, J.C., Boulnois, G.J., Andrew, P.W., and Mitchell, T.J. (1995). The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. *The Journal of infectious diseases* *172*, 119-123.
- Cao, J., Chen, D., Xu, W., Chen, T., Xu, S., Luo, J., Zhao, Q., Liu, B., Wang, D., Zhang, X., et al. (2007). Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. *Vaccine* *25*, 4996-5005.
- Carvalho Mda, G., Pimenta, F.C., Gertz, R.E., Jr., Joshi, H.H., Trujillo, A.A., Keys, L.E., Findley, J., Moura, I.S., Park, I.H., Hollingshead, S.K., et al. (2009). PCR-based quantitation and clonal diversity of the current prevalent invasive serogroup 6 pneumococcal serotype, 6C, in the United States in 1999 and 2006 to 2007. *Journal of clinical microbiology* *47*, 554-559.
- Casadevall, A., and Pirofski, L.A. (1999). Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infection and immunity* *67*, 3703-3713.
- Chain, E., Florey, H.W., Adelaide, M.B., Gardner, A.D., Heatley, N.G., Jennings, M.A., Orr-Ewing, J., and Sanders, A.G. (1993). Penicillin as a chemotherapeutic agent. 1940. *Clinical orthopaedics and related research*, 3-7.
- Chen, C.C., Kurokawa, T., Shaw, S.Y., Tillotson, L.G., Kalled, S., and Isselbacher, K.J. (1986). Human erythrocyte glucose transporter: normal asymmetric orientation and function in liposomes. *Proceedings of the National Academy of Sciences of the United States of America* *83*, 2652-2656.

- Chen, P., Cisar, J.O., Hess, S., Ho, J.T., and Leung, K.P. (2007). Amended description of the genes for synthesis of *Actinomyces naeslundii* T14V type 1 fimbriae and associated adhesin. *Infection and immunity* 75, 4181-4185.
- De las Rivas, B., Garcia, J.L., Lopez, R., and Garcia, P. (2001). Molecular characterization of the pneumococcal teichoic acid phosphorylcholine esterase. *Microbial drug resistance* (Larchmont, N.Y. 7, 213-222.
- Decousser, J.W., Pina, P., Viguier, F., Picot, F., Courvalin, P., and Allouch, P. (2004). Invasive *Streptococcus pneumoniae* in France: antimicrobial resistance, serotype, and molecular epidemiology findings from a monthly national study in 2000 to 2002. *Antimicrobial agents and chemotherapy* 48, 3636-3639.
- Deivanayagam, C.C., Rich, R.L., Carson, M., Owens, R.T., Danthuluri, S., Bice, T., Hook, M., and Narayana, S.V. (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8, 67-78.
- Den Blaauwen, T., de Pedro, M.A., Nguyen-Disteche, M., and Ayala, J.A. (2008). Morphogenesis of rod-shaped sacculi. *FEMS microbiology reviews* 32, 321-344.
- Dintilhac, A., and Claverys, J.P. (1997). The *adc* locus, which affects competence for genetic transformation in *Streptococcus pneumoniae*, encodes an ABC transporter with a putative lipoprotein homologous to a family of streptococcal adhesins. *Research in microbiology* 148, 119-131.
- Dorléans, F., Varon, E., Lepoutre, A., Georges, S., Gutmann, L., Lévy-Bruhl, D. et les microbiologistes du réseau Epibac et du réseau des Observatoires régionaux du pneumocoque (2010). Impact de la vaccination par le vaccin anti-pneumococcique conjugué heptavalent sur l'incidence des infections invasives à pneumocoques en France.
- Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M.C., Kojadinovic, M., Lalioui, L., Poyart, C., and Trieu-Cuot, P. (2006). Assembly and role of pili in group B streptococci. *Molecular microbiology* 60, 1401-1413.
- Dramsi, S., Trieu-Cuot, P., and Bierne, H. (2005). Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Research in microbiology* 156, 289-297.
- Duguid, J.P., Smith, I.W., Dempster, G., and Edmunds, P.N. (1955). Non-flagellar filamentous appendages (fimbriae) and haemagglutinating activity in *Bacterium coli*. *The Journal of pathology and bacteriology* 70, 335-348.
- El Mortaji, L., Terrasse, R., Dessen, A., Vernet, T., and Di Guilmi, A.M. (2010). Stability and assembly of pilus subunits of *Streptococcus pneumoniae*. *The Journal of biological chemistry* 285, 12405-12415.
- Elsner, A., Kreikemeyer, B., Braun-Kiewnick, A., Spellerberg, B., Buttaro, B.A., and Podbielski, A. (2002). Involvement of Lsp, a member of the LraI-lipoprotein family in *Streptococcus pyogenes*, in eukaryotic cell adhesion and internalization. *Infection and immunity* 70, 4859-4869.
- Falker, S., Nelson, A.L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008). Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Molecular microbiology* 70, 595-607.
- Fernandez-Tornero, C., Lopez, R., Garcia, E., Gimenez-Gallego, G., and Romero, A. (2001). A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nature structural biology* 8, 1020-1024.
- Fischer, W. (2000). Pneumococcal Lipoteichoic and Teichoic Acid. In *Streptococcus pneumoniae: Molecular Biology & Mechanisms of Disease*, A. Tomasz, ed. (New York: Mary Ann Liebert), pp. 155-177.

- Frolet, C., Beniazza, M., Roux, L., Gallet, B., Noirclerc-Savoie, M., Vernet, T., and Di Guilmi, A.M. (2010). New adhesin functions of surface-exposed pneumococcal proteins. *BMC microbiology* *10*, 190.
- Garau, G., Lemaire, D., Vernet, T., Dideberg, O., and Di Guilmi, A.M. (2005). Crystal structure of phosphorylcholine esterase domain of the virulence factor choline-binding protein e from streptococcus pneumoniae: new structural features among the metallo-beta-lactamase superfamily. *The Journal of biological chemistry* *280*, 28591-28600.
- Gaspar, A.H., and Ton-That, H. (2006). Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. *Journal of bacteriology* *188*, 1526-1533.
- Geerlof, A., Brown, J., Coutard, B., Egloff, M.P., Enguita, F.J., Fogg, M.J., Gilbert, R.J., Groves, M.R., Haouz, A., Nettleship, J.E., et al. (2006). The impact of protein characterization in structural proteomics. *Acta crystallographica* *62*, 1125-1136.
- Glaser, P.E., and Gross, R.W. (1995). Rapid plasmenylethanolamine-selective fusion of membrane bilayers catalyzed by an isoform of glyceraldehyde-3-phosphate dehydrogenase: discrimination between glycolytic and fusogenic roles of individual isoforms. *Biochemistry* *34*, 12193-12203.
- Goebel, W.F., and Adams, M.H. (1943). The Immunological Properties of the Heterophile Antigen and Somatic Polysaccharide of Pneumococcus. *The Journal of experimental medicine* *77*, 435-449.
- Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I., and Masure, H.R. (2000). Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. *Infection and immunity* *68*, 5690-5695.
- Gotwals, P.J., Chi-Rosso, G., Ryan, S.T., Sizing, I., Zafari, M., Benjamin, C., Singh, J., Venyaminov, S.Y., Pepinsky, R.B., and Koteliansky, V. (1999). Divalent cations stabilize the alpha 1 beta 1 integrin I domain. *Biochemistry* *38*, 8280-8288.
- Greenberg, D. (2009). The shifting dynamics of pneumococcal invasive disease after the introduction of the pneumococcal 7-valent conjugated vaccine: toward the new pneumococcal conjugated vaccines. *Clin Infect Dis* *49*, 213-215.
- Gros, P., Croop, J., and Housman, D. (1986). Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. *Cell* *47*, 371-380.
- Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Muller, E., and Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infection and immunity* *73*, 4653-4667.
- Hava, D.L., and Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Molecular microbiology* *45*, 1389-1406.
- Hava, D.L., Hemsley, C.J., and Camilli, A. (2003). Transcriptional regulation in the *Streptococcus pneumoniae* rlrA pathogenicity islet by RlrA. *Journal of bacteriology* *185*, 413-421.
- Hemsley, C., Joyce, E., Hava, D.L., Kawale, A., and Camilli, A. (2003). MgrA, an orthologue of Mga, Acts as a transcriptional repressor of the genes within the rlrA pathogenicity islet in *Streptococcus pneumoniae*. *Journal of bacteriology* *185*, 6640-6647.
- Hendriksen, W.T., Silva, N., Bootsma, H.J., Blue, C.E., Paterson, G.K., Kerr, A.R., de Jong, A., Kuipers, O.P., Hermans, P.W., and Mitchell, T.J. (2007). Regulation of gene expression in *Streptococcus pneumoniae* by response regulator 09 is strain dependent. *Journal of bacteriology* *189*, 1382-1389.
- Hermoso, J.A., Lagartera, L., Gonzalez, A., Stelter, M., Garcia, P., Martinez-Ripoll, M., Garcia, J.L., and Menendez, M. (2005). Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce. *Nature structural & molecular biology* *12*, 533-538.

- Hilleringmann, M., Giusti, F., Baudner, B.C., Massignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ferlenghi, I. (2008). Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. *PLoS pathogens* 4, e1000026.
- Hilleringmann, M., Ringler, P., Muller, S.A., De Angelis, G., Rappuoli, R., Ferlenghi, I., and Engel, A. (2009). Molecular architecture of *Streptococcus pneumoniae* TIGR4 pili. *The EMBO journal* 28, 3921-3930.
- Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszcak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., et al. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *Journal of bacteriology* 183, 5709-5717.
- Houwink, A.L., and van, I.W. (1950). Electron microscopical observations on bacterial cytology; a study on flagellation. *Biochimica et biophysica acta* 5, 10-44.
- Hsieh, Y.C., Lin, P.Y., Chiu, C.H., Huang, Y.C., Chang, K.Y., Liao, C.H., Chiu, N.C., Chuang, Y.C., Chen, P.Y., Chang, S.C., et al. (2009). National survey of invasive pneumococcal diseases in Taiwan under partial PCV7 vaccination in 2007: emergence of serotype 19A with high invasive potential. *Vaccine* 27, 5513-5518.
- Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003). Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry* 42, 11307-11315.
- Hutchings, M.I., Palmer, T., Harrington, D.J., and Sutcliffe, I.C. (2009). Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends in microbiology* 17, 13-21.
- Hyams, C., Tam, J.C., Brown, J.S., and Gordon, S.B. (2010). C3b/iC3b deposition on *Streptococcus pneumoniae* is not affected by HIV infection. *PloS one* 5, e8902.
- Hyams, C., Yuste, J., Bax, K., Camberlein, E., Weiser, J.N., and Brown, J.S. (2010). *Streptococcus pneumoniae* resistance to complement-mediated immunity is dependent on the capsular serotype. *Infection and immunity* 78, 716-725.
- Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R.T. (2001). Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences of the United States of America* 98, 6056-6061.
- Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96, 23-28.
- Izore, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A.M., and Dessen, A. (2010). Structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*. *Structure* 18, 106-115.
- Jedrzejewski, M.J. (2007). Unveiling molecular mechanisms of bacterial surface proteins: *Streptococcus pneumoniae* as a model organism for structural studies. *Cell Mol Life Sci* 64, 2799-2822.
- Jin, H., Song, Y.P., Boel, G., Kochar, J., and Pancholi, V. (2005). Group A streptococcal surface GAPDH, SDH, recognizes uPAR/CD87 as its receptor on the human pharyngeal cell and mediates bacterial adherence to host cells. *Journal of molecular biology* 350, 27-41.
- Johnston, J.W., Briles, D.E., Myers, L.E., and Hollingshead, S.K. (2006). Mn<sup>2+</sup>-dependent regulation of multiple genes in *Streptococcus pneumoniae* through PsaR and the resultant impact on virulence. *Infection and immunity* 74, 1171-1180.
- Kadioglu, A., Gingles, N.A., Grattan, K., Kerr, A., Mitchell, T.J., and Andrew, P.W. (2000). Host cellular immune response to pneumococcal lung infection in mice. *Infection and immunity* 68, 492-501.

- Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nature reviews* 6, 288-301.
- Kamerling, J.P. (2000). Pneumococcal polysaccharides: A chemical view. In *Streptococcus pneumoniae: Molecular Biology & Mechanisms of disease*, A. Tomasz, ed. (New York: Mary Ann Liebert), pp. 81-114.
- Kang, H.J., and Baker, E.N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *The Journal of biological chemistry* 284, 20729-20737.
- Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science (New York, N.Y)* 318, 1625-1628.
- Kang, H.J., Paterson, N.G., Gaspar, A.H., Ton-That, H., and Baker, E.N. (2009). The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proceedings of the National Academy of Sciences of the United States of America* 106, 16967-16971.
- Khandavilli, S., Homer, K.A., Yuste, J., Basavanna, S., Mitchell, T., and Brown, J.S. (2008). Maturation of *Streptococcus pneumoniae* lipoproteins by a type II signal peptidase is required for ABC transporter function and full virulence. *Molecular microbiology* 67, 541-557.
- Kharat, A.S., and Tomasz, A. (2003). Inactivation of the *srtA* gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells in vitro. *Infection and immunity* 71, 2758-2765.
- Kim, J.O., Romero-Steiner, S., Sorensen, U.B., Blom, J., Carvalho, M., Barnard, S., Carlone, G., and Weiser, J.N. (1999). Relationship between cell surface carbohydrates and intrastrain variation on opsonophagocytosis of *Streptococcus pneumoniae*. *Infection and immunity* 67, 2327-2333.
- Konto-Ghiorghi, Y., Mairey, E., Mallet, A., Dumenil, G., Caliot, E., Trieu-Cuot, P., and Dramsi, S. (2009). Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. *PLoS pathogens* 5, e1000422.
- Krishnan, V., Gaspar, A.H., Ye, N., Mandlik, A., Ton-That, H., and Narayana, S.V. (2007). An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion. *Structure* 15, 893-903.
- Lacks, S.A. (2004). Transformation. In *The Pneumococcus*, E.I. Tuomanen, ed. (Washington, D.C.: ASM Press), pp. 89-115
- Laible, G., and Hakenbeck, R. (1987). Penicillin-binding proteins in beta-lactam-resistant laboratory mutants of *Streptococcus pneumoniae*. *Molecular microbiology* 1, 355-363.
- Laible, G., Spratt, B.G., and Hakenbeck, R. (1991). Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Molecular microbiology* 5, 1993-2002.
- Lancefield, R.C., and Dole, V.P. (1946). The Properties of T Antigens Extracted from Group a Hemolytic Streptococci. *The Journal of experimental medicine* 84, 449-471.
- Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck, W., and Amrein, K.E. (1999). Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene* 237, 223-234.
- Lauer, P., Rinaudo, C.D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A.R., Mora, M., Rappuoli, R., Grandi, G., et al. (2005). Genome analysis reveals pili in Group B *Streptococcus*. *Science (New York, N.Y)* 309, 105.

- Lee, C.J., Banks, S.D., and Li, J.P. (1991). Virulence, immunity, and vaccine related to *Streptococcus pneumoniae*. *Critical reviews in microbiology* 18, 89-114.
- LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006). RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet. *Infection and immunity* 74, 2453-2456.
- LeMieux, J., Woody, S., and Camilli, A. (2008). Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. *Journal of bacteriology* 190, 6002-6013.
- Ligon, B.L. (2004). Penicillin: its discovery and early development. *Seminars in pediatric infectious diseases* 15, 52-57.
- Ling, E., Feldman, G., Portnoi, M., Dagan, R., Overweg, K., Mulholland, F., Chalifa-Caspi, V., Wells, J., and Mizrahi-Nebenzahl, Y. (2004). Glycolytic enzymes associated with the cell surface of *Streptococcus pneumoniae* are antigenic in humans and elicit protective immune responses in the mouse. *Clinical and experimental immunology* 138, 290-298.
- Linke, C., Young, P.G., Kang, H.J., Bunker, R.D., Middleditch, M.J., Caradoc-Davies, T.T., Proft, T., and Baker, E.N. (2010). Crystal structure of the minor pilin FctB reveals determinants of Group A streptococcal pilus anchoring. *The Journal of biological chemistry* 285, 20381-20389.
- Loisel, E., Jacquamet, L., Serre, L., Bauvois, C., Ferrer, J.L., Vernet, T., Di Guilmi, A.M., and Durmort, C. (2008). AdcAII, a new pneumococcal Zn-binding protein homologous with ABC transporters: biochemical and structural analysis. *Journal of molecular biology* 381, 594-606.
- Lynch, J.P., 3rd, and Zhanel, G.G. (2009). *Streptococcus pneumoniae*: epidemiology, risk factors, and strategies for prevention. *Seminars in respiratory and critical care medicine* 30, 189-209.
- Lynch, J.P., 3rd, and Zhanel, G.G. (2010). *Streptococcus pneumoniae*: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. *Current opinion in pulmonary medicine* 16, 217-225.
- Macleod, C.M., and Krauss, M.R. (1953). Control by factors distinct from the S transforming principle of the amount of capsular polysaccharide produced by type III pneumococci. *The Journal of experimental medicine* 97, 767-771.
- Madureira, P., Baptista, M., Vieira, M., Magalhaes, V., Camelo, A., Oliveira, L., Ribeiro, A., Tavares, D., Trieu-Cuot, P., Vilanova, M., et al. (2007). *Streptococcus agalactiae* GAPDH is a virulence-associated immunomodulatory protein. *J Immunol* 178, 1379-1387.
- Magee, A.D., and Yother, J. (2001). Requirement for capsule in colonization by *Streptococcus pneumoniae*. *Infection and immunity* 69, 3755-3761.
- Mandlik, A., Das, A., and Ton-That, H. (2008). The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14147-14152.
- Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2007). *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Molecular microbiology* 64, 111-124.
- Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2008). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends in microbiology* 16, 33-40.
- Manetti, A.G., Zingaretti, C., Falugi, F., Capo, S., Bombaci, M., Bagnoli, F., Gambellini, G., Bensi, G., Mora, M., Edwards, A.M., et al. (2007). *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Molecular microbiology* 64, 968-983.



- Manzano, C., Contreras-Martel, C., El Mortaji, L., Izore, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A.M., and Dessen, A. (2008). Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* *16*, 1838-1848.
- Manzano, C., Izore, T., Job, V., Di Guilmi, A.M., and Dessen, A. (2009). Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. *Biochemistry* *48*, 10549-10557.
- Marraffini, L.A., and Schneewind, O. (2006). Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Molecular microbiology* *62*, 1402-1417.
- Marraffini, L.A., Ton-That, H., Zong, Y., Narayana, S.V., and Schneewind, O. (2004). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. A conserved arginine residue is required for efficient catalysis of sortase A. *The Journal of biological chemistry* *279*, 37763-37770.
- Mazmanian, S.K., Liu, G., Ton-That, H., and Schneewind, O. (1999). *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science (New York, N.Y)* *285*, 760-763.
- Mazmanian, S.K., Ton-That, H., Su, K., and Schneewind, O. (2002). An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* *99*, 2293-2298.
- McAllister, L.J., Tseng, H.J., Ogunniyi, A.D., Jennings, M.P., McEwan, A.G., and Paton, J.C. (2004). Molecular analysis of the *psa* permease complex of *Streptococcus pneumoniae*. *Molecular microbiology* *53*, 889-901.
- Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature* *449*, 819-826.
- Mishra, A., Das, A., Cisar, J.O., and Ton-That, H. (2007). Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*. *Journal of bacteriology* *189*, 3156-3165.
- Mitchell, T.J., Andrew, P.W., Saunders, F.K., Smith, A.N., and Boulnois, G.J. (1991). Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Molecular microbiology* *5*, 1883-1888.
- Moore, M.R., Gertz, R.E., Jr., Woodbury, R.L., Barkocy-Gallagher, G.A., Schaffner, W., Lexau, C., Gershman, K., Reingold, A., Farley, M., Harrison, L.H., et al. (2008). Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. *The Journal of infectious diseases* *197*, 1016-1027.
- 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. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 15641-15646.
- Morgan, P.J., Hyman, S.C., Rowe, A.J., Mitchell, T.J., Andrew, P.W., and Saibil, H.R. (1995). Subunit organisation and symmetry of pore-forming, oligomeric pneumolysin. *FEBS letters* *371*, 77-80.
- Morona, J.K., Morona, R., Miller, D.C., and Paton, J.C. (2002). *Streptococcus pneumoniae* capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. *Journal of bacteriology* *184*, 577-583.
- Morona, J.K., Morona, R., Miller, D.C., and Paton, J.C. (2003). Mutational analysis of the carboxy-terminal (YGX)<sub>4</sub> repeat domain of CpsD, an autophosphorylating tyrosine kinase required for capsule biosynthesis in *Streptococcus pneumoniae*. *Journal of bacteriology* *185*, 3009-3019.

- Morona, J.K., Paton, J.C., Miller, D.C., and Morona, R. (2000). Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in streptococcus pneumoniae. *Molecular microbiology* 35, 1431-1442.
- Moschioni, M., Donati, C., Muzzi, A., Massignani, V., Censini, S., Hanage, W.P., Bishop, C.J., Reis, J.N., Normark, S., Henriques-Normark, B., et al. (2008). Streptococcus pneumoniae contains 3 rlrA pilus variants that are clonally related. *The Journal of infectious diseases* 197, 888-896.
- Munoz-Elias, E.J., Marcano, J., and Camilli, A. (2008). Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization. *Infection and immunity* 76, 5049-5061.
- Nallapareddy, S.R., Singh, K.V., Sillanpaa, J., Garsin, D.A., Hook, M., Erlandsen, S.L., and Murray, B.E. (2006). Endocarditis and biofilm-associated pili of Enterococcus faecalis. *The Journal of clinical investigation* 116, 2799-2807.
- Neiers, F., Madhurantakam, C., Falker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009). Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *Journal of molecular biology* 393, 704-716.
- Nelson, A.L., Ries, J., Bagnoli, F., Dahlberg, S., Falker, S., Rounioja, S., Tschop, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., et al. (2007). RrgA is a pilus-associated adhesin in Streptococcus pneumoniae. *Molecular microbiology* 66, 329-340.
- Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., and Weiser, J.N. (2007). Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infection and immunity* 75, 83-90.
- Nobbs, A.H., Lamont, R.J., and Jenkinson, H.F. (2009). Streptococcus adherence and colonization. *Microbiol Mol Biol Rev* 73, 407-450, Table of Contents.
- Nobbs, A.H., Rosini, R., Rinaudo, C.D., Maione, D., Grandi, G., and Telford, J.L. (2008). Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in Streptococcus agalactiae. *Infection and immunity* 76, 3550-3560.
- ORP (2010). Notes ORP 2010 In Observatoires régionaux du pneumocoque
- Pallen, M.J., Lam, A.C., Antonio, M., and Dunbar, K. (2001). An embarrassment of sortases - a richness of substrates? *Trends in microbiology* 9, 97-102.
- Pancholi, V., and Fischetti, V.A. (1988). Isolation and characterization of the cell-associated region of group A streptococcal M6 protein. *Journal of bacteriology* 170, 2618-2624.
- Pancholi, V., and Fischetti, V.A. (1989). Identification of an endogenous membrane anchor-cleaving enzyme for group A streptococcal M protein. Its implication for the attachment of surface proteins in gram-positive bacteria. *The Journal of experimental medicine* 170, 2119-2133.
- Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., et al. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* 6, 429-440.
- Paterson, G.K., and Mitchell, T.J. (2004). The biology of Gram-positive sortase enzymes. *Trends in microbiology* 12, 89-95.
- Paterson, G.K., and Mitchell, T.J. (2006). The role of Streptococcus pneumoniae sortase A in colonisation and pathogenesis. *Microbes and infection / Institut Pasteur* 8, 145-153.
- Paton, J.C., Rowan-Kelly, B., and Ferrante, A. (1984). Activation of human complement by the pneumococcal toxin pneumolysin. *Infection and immunity* 43, 1085-1087.

- Perry, A.M., Ton-That, H., Mazmanian, S.K., and Schneewind, O. (2002). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *The Journal of biological chemistry* 277, 16241-16248.
- Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998). Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infection and immunity* 66, 5620-5629.
- Pozzi, G., Masala, L., Iannelli, F., Manganelli, R., Havarstein, L.S., Piccoli, L., Simon, D., and Morrison, D.A. (1996). Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *Journal of bacteriology* 178, 6087-6090.
- Proft, T., and Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66, 613-635.
- Rigden, D.J., Galperin, M.Y., and Jedrzejewski, M.J. (2003). Analysis of structure and function of putative surface-exposed proteins encoded in the *Streptococcus pneumoniae* genome: a bioinformatics-based approach to vaccine and drug design. *Critical reviews in biochemistry and molecular biology* 38, 143-168.
- Rosch, J.W., Mann, B., Thornton, J., Sublett, J., and Tuomanen, E. (2008). Convergence of regulatory networks on the pilus locus of *Streptococcus pneumoniae*. *Infection and immunity* 76, 3187-3196.
- Rosenstein, R., and Gotz, F. (2000). Staphylococcal lipases: biochemical and molecular characterization. *Biochimie* 82, 1005-1014.
- Ruzin, A., Severin, A., Ritacco, F., Tabei, K., Singh, G., Bradford, P.A., Siegel, M.M., Projan, S.J., and Shlaes, D.M. (2002). Further evidence that a cell wall precursor [C(55)-MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. *Journal of bacteriology* 184, 2141-2147.
- Schneewind, O., Fowler, A., and Faull, K.F. (1995). Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science (New York, N.Y)* 268, 103-106.
- Schneewind, O., Mihaylova-Petkov, D., and Model, P. (1993). Cell wall sorting signals in surface proteins of gram-positive bacteria. *The EMBO journal* 12, 4803-4811.
- Schneewind, O., Model, P., and Fischetti, V.A. (1992). Sorting of protein A to the staphylococcal cell wall. *Cell* 70, 267-281.
- Seib, K.L., Dougan, G., and Rappuoli, R. (2009). The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS genetics* 5, e1000612.
- Sjoquist, J., Movitz, J., Johansson, I.B., and Hjelm, H. (1972). Localization of protein A in the bacteria. *European journal of biochemistry / FEBS* 30, 190-194.
- Song, X.M., Connor, W., Hokamp, K., Babiuk, L.A., and Potter, A.A. (2009). The growth phase-dependent regulation of the pilus locus genes by two-component system TCS08 in *Streptococcus pneumoniae*. *Microbial pathogenesis* 46, 28-35.
- Sorensen, U.B., Henriksen, J., Chen, H.C., and Szu, S.C. (1990). Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. *Microbial pathogenesis* 8, 325-334.
- Spellerberg, B., Rozdzinski, E., Martin, S., Weber-Heynemann, J., Schnitzler, N., Luttkien, R., and Podbielski, A. (1999). Lmb, a protein with similarities to the Lral adhesin family, mediates attachment of *Streptococcus agalactiae* to human laminin. *Infection and immunity* 67, 871-878.

- Spraggon, G., Koesema, E., Scarselli, M., Malito, E., Biagini, M., Norais, N., Emolo, C., Barocchi, M.A., Giusti, F., Hilleringmann, M., et al. (2010). Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. *PloS one* 5, e10919.
- Steinmoen, H., Knutsen, E., and Havarstein, L.S. (2002). Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7681-7686.
- Stringaris, A.K., Geisenhainer, J., Bergmann, F., Balshusemann, C., Lee, U., Zysk, G., Mitchell, T.J., Keller, B.U., Kuhnt, U., Gerber, J., et al. (2002). Neurotoxicity of pneumolysin, a major pneumococcal virulence factor, involves calcium influx and depends on activation of p38 mitogen-activated protein kinase. *Neurobiology of disease* 11, 355-368.
- Sulek, K. (1968). [Nobel prize in 1945 for Alexander Fleming, Ernest Boris Chain and Howard Walter Florey for the discovery of penicillin and its therapeutic effect in infections]. *Wiad Lek* 21, 1388-1390.
- Suree, N., Liew, C.K., Villareal, V.A., Thieu, W., Fadeev, E.A., Clemens, J.J., Jung, M.E., and Clubb, R.T. (2009). The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *The Journal of biological chemistry* 284, 24465-24477.
- Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A., and Ton-That, H. (2007). Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Molecular microbiology* 66, 961-974.
- Tam, R., and Saier, M.H., Jr. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological reviews* 57, 320-346.
- Tanner, M.J., and Gray, W.R. (1971). The isolation and functional identification of a protein from the human erythrocyte 'ghost'. *The Biochemical journal* 125, 1109-1117.
- Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., and Grandi, G. (2006). Pili in gram-positive pathogens. *Nature reviews* 4, 509-519.
- Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science (New York, N.Y)* 293, 498-506.
- Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., et al. (2000). A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Molecular microbiology* 35, 566-576.
- Tillet, W.S., Goebel, W.F., Avery, O.T. (1930). Chemical and immunological properties of species-specific carbohydrate of pneumococci. *J. Exp. Med* 52, 895-900.
- Tilley, S.J., Orlova, E.V., Gilbert, R.J., Andrew, P.W., and Saibil, H.R. (2005). Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* 121, 247-256.
- Tipper, D.J., and Strominger, J.L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proceedings of the National Academy of Sciences of the United States of America* 54, 1133-1141.
- Tomasz, A. (1967). Choline in the cell wall of a bacterium: novel type of polymer-linked choline in *Pneumococcus*. *Science (New York, N.Y)* 157, 694-697.
- Tomasz, A., and Fischer, W. (2000). The cell wall of *Streptococcus pneumoniae*. In *Gram-Positive Pathogens*, V.A.F.e. al., ed., pp. 191-200.

- Tomasz, A., and Hotchkiss, R.D. (1964). Regulation of the Transformability of Pneumococcal Cultures by Macromolecular Cell Products. *Proceedings of the National Academy of Sciences of the United States of America* 51, 480-487.
- Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F., and Schneewind, O. (1999). Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proceedings of the National Academy of Sciences of the United States of America* 96, 12424-12429.
- Ton-That, H., Marraffini, L.A., and Schneewind, O. (2004). Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Molecular microbiology* 53, 251-261.
- Ton-That, H., Mazmanian, S.K., Alksne, L., and Schneewind, O. (2002). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *The Journal of biological chemistry* 277, 7447-7452.
- Ton-That, H., Mazmanian, S.K., Faull, K.F., and Schneewind, O. (2000). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *The Journal of biological chemistry* 275, 9876-9881.
- Ton-That, H., and Schneewind, O. (1999). Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *The Journal of biological chemistry* 274, 24316-24320.
- Ton-That, H., and Schneewind, O. (2003). Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Molecular microbiology* 50, 1429-1438.
- Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. *Trends in microbiology* 12, 228-234.
- Tyrrell, G.J., Lovgren, M., Chui, N., Minion, J., Garg, S., Kellner, J.D., and Marrie, T.J. (2009). Serotypes and antimicrobial susceptibilities of invasive *Streptococcus pneumoniae* pre- and post-seven valent pneumococcal conjugate vaccine introduction in Alberta, Canada, 2000-2006. *Vaccine* 27, 3553-3560.
- Varon, E., Janoir, C., Gutmann, L. (2009). Rapport CNPR d'activité 2009: Epidémiologie 2008.
- Vollmer, W., and Tomasz, A. (2001). Identification of the teichoic acid phosphorylcholine esterase in *Streptococcus pneumoniae*. *Molecular microbiology* 39, 1610-1622.
- Waltz, D.A., Natkin, L.R., Fujita, R.M., Wei, Y., and Chapman, H.A. (1997). Plasmin and plasminogen activator inhibitor type 1 promote cellular motility by regulating the interaction between the urokinase receptor and vitronectin. *The Journal of clinical investigation* 100, 58-67.
- Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., and Johnson, J.E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science (New York, N.Y.)* 289, 2129-2133.
- Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., et al. (2001). Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infection and immunity* 69, 1593-1598.
- Wu, H., and Fives-Taylor, P.M. (2001). Molecular strategies for fimbrial expression and assembly. *Crit Rev Oral Biol Med* 12, 101-115.
- Yanagawa, R., and Honda, E. (1976). Presence of pili in species of human and animal parasites and pathogens of the genus *Corynebacterium*. *Infection and immunity* 13, 1293-1295.

- Yother, J. (2004). Capsules. In *The Pneumococcus*, E.I.T.e. al., ed. (Washington, D.C.: ASM Press), pp. 30-48.
- Zong, Y., Xu, Y., Liang, X., Keene, D.R., Hook, A., Gurusiddappa, S., Hook, M., and Narayana, S.V. (2005). A 'Collagen Hug' model for *Staphylococcus aureus* CNA binding to collagen. *The EMBO journal* 24, 4224-4236.
- Zysk, G., Bongaerts, R.J., ten Thoren, E., Bethe, G., Hakenbeck, R., and Heinz, H.P. (2000). Detection of 23 immunogenic pneumococcal proteins using convalescent-phase serum. *Infection and immunity* 68, 3740-3743.



# **Annexes**





## ANNEXES

## Annexe 1 : Récapitulatif des vecteurs construits pour cette étude

N°	Vecteurs de base	Réalisé par *
1	pLIM01 RrgA	AMDG
2	pLIM01 RrgB	AMDG
3	pLIM01 RrgC	AMDG
4	pETDuet RrgA	LEM
5	pETDuet RrgB	LEM
6	pACYCDuet RrgC	LEM
7	pACYCDuet SrtC-1	LEM
8	pCDFDuet SrtC-2	LEM
9	pCDFDuet SrtC-3	LEM
10	pETDuet RrgA-RrgB (FigureX)	LEM
11	pACYCDuet RrgC-SrtC-1 (FigureX)	LEM
12	pCDFDuet SrtC-2-SrtC-3 (FigureX)	LEM
N°	Vecteurs sortases mutées	Réalisé par *
13	pACYCDuet SrtC-1 C193A	LEM
14	pACYCDuet RrgC SrtC-1 C193A	LEM
15	pCDFDuet SrtC-2 C179A	LEM
16	pCDFDuet SrtC-3 C176A	LEM
N°	Vecteurs pilines mutées (ponts intra et inter-moléculaires)	Réalisé par *
17	pLIM01 RrgB K41A	RT
18	pLIM01 RrgB E143A	RT
19	pLIM01 RrgB N184A	RT
20	pLIM01 RrgB D241A	LEM
21	pLIM01 RrgB E405A	LEM
22	pLIM01 RrgB E577A	RT
23	pLIM01 RrgB N623A	RT
24	pLIM01 RrgB D241A/E405A/E577A	LEM
25	pETDuet RrgA K191A	LEM
26	pETDuet RrgA D600A	LEM
27	pETDuet RrgA N695A	LEM
28	pETDuet RrgA E803A	LEM
29	pETDuet RrgA D600A/E803A	LEM
30	pETDuet RrgB D241A	LEM
31	pETDuet RrgB E405A	LEM
32	pETDuet RrgB E577A	LEM
33	pETDuet RrgB E405A/D241A/E577A	LEM
34	pETDuet RrgA RrgB K162A	LEM
35	pETDuet RrgA RrgB K183A	LEM
36	pACYCDuet RrgC K155A	LEM
37	pACYCDuet RrgC E222A	LEM
38	pACYCDuet RrgC N252A	LEM
39	pACYCDuet RrgC K264A	LEM
40	pACYCDuet RrgC E322A	LEM
41	pACYCDuet RrgC N354A	LEM
42	pACYCDuet RrgC E222A/E322A	LEM
43	pACYCDuet RrgC E222A SrtC-1	LEM
44	pACYCDuet RrgC E322A SrtC-1	LEM
45	pACYCDuet RrgC E222A/E322A SrtC-1	LEM
46	pETDuet RrgA D600A RrgB	RT
47	pETDuet RrgA E803A RrgB	RT
48	pETDuet RrgA D600A/E803A RrgB	LEM
49	pETDuet RrgA RrgB D241A	TF
50	pETDuet RrgA RrgB E405A	TF
51	pETDuet RrgA RrgB E577A	LEM
52	pETDuet RrgA RrgB D241A/E405A/E577A	TF
53	pETDuet RrgA D600A/E803A RrgB D241A/E405A/E577A	TF

N°	Vecteurs pilines mutées (délétions LPXTG)	Réalisé par *
54	pETDuet RrgBΔIPQTG	LEM
55	pETDuet RrgA RrgBΔIPQTG	LEM
56	pETDuet RrgAΔYPRTG	LEM
57	pETDuet RrgAΔYPRTG RrgB	LEM
58	pACYCDuet RrgCΔVPDTG	LEM
59	pACYCDuet RrgCΔVPDTG SrtC-1	LEM
N°	Vecteurs pilines mutées (inversions LPxTG)	Réalisé par *
60	RrgAΔYPRTG::IPQTG RrgB	LEM
61	RrgAΔVPDTG::VPDTG RrgB	LEM
62	RrgBΔIPQTG::YPRTG	LEM
63	RrgCΔVPDTG::YPRTG	LEM

\*AMDG (Anne Marie Di Guilmi) LEM (Lamy El Mortaji) RT (Rémi Terrasse) TF (Tony Fourny)

NB : Les oligonucléotides utilisés pour réaliser les mutants sont listés dans l'annexe 2

## Annexe 2 : Liste des oligonucléotides utilisés pour réaliser les mutants

Oligonucléotides de mutagenèse (cystéine catalytique)	
66 C193A_D	Clothilde Manzano (LPM)
66 C193A_R	Clothilde Manzano (LPM)
67 C179A_D	5'-aagattatgacgaccttggactgctacaccgtatgatgattaacagtc-3'
67 C179A_R	5'-gactgttaatcatatacgggtgtagcagtcaacaaggtcgcataatcctt-3'
68 C176A_D	5'-taaaaaatcatgaccttgataaccgcccgatccgattcctacctttaataaa-3'
68 C176A_R	5'-tttattaaaggttaggaatcggatcggcggttatcaaggtcatgatatttta-3'
Oligonucléotides de mutagenèse (ponts intramoléculeaires RrgA)	
62 K191A_D	5'-tgtgattccagaaggtacaccttccagcgagaatttatcaagtgaataatttgg-3'
62 K191A_R	5'-ccaaattattcacttgataaattctcgtgaaagtgtaccttctggaatcaca-3'
62 D600A_D	5'-gagaatgggtacgattacagctccgatgggtgagt-3'
62 D600A_R	5'-actcaccatcggagctgtaatcgtaccattctc-3'
62 N695A_D	5'-agtttgaagcaataaattttatgataccgctgggtcgaacaaccttacatcctaa-3'
62 N695A_R	5'-ttaggatgtaaggttggctcgaccagcggatcataaaatttattgcttacaact-3'
62 E803A_D	5'-agatgggaaatcagattatttgcgaattctgaaccagctggttat-3'
62 E803A_R	5'-ataaccagctggttcagaatttgcgaataatcagatatttcccactc-3'
Oligonucléotides de mutagenèse (ponts intramoléculeaires RrgB)	
63 K41A_D	5'-acaacaacatctgttaccggttcatgcactattggcaacagatggg-3'
63 K41A_R	5'-cccattctgttgcgaatagtgcatgaacggtaaacagatgttgggt-3'
63 E143A_D	5'-accagctgctaagtataaaatttatgcaattccagctttatcaacttatgtc-3'
63 E143A_R	5'-gacataagttgataaactgtgaattgcataaattttatacttagcagctgg-3'
63 N184A_D	5'-gtggatggcctgtgtatccaaaagctacagaagcaaaagc-3'
63 N184A_R	5'-ttttggctttgcttctgtagcttttggatacacatgcccac-3'
63 D241A_D	5'-caacagcaaaactggagcgttagaatgactgaaggtt-3'
63 D241A_R	5'-aaacctcagtcattctagcgtccagtttgcgtgtg-3'
63 E405A_D	5'-taaaaatacagaataaattcgttgcagctagtataaaagggattcagcag-3'
63 E405A_R	5'-ctgctgaataaccttttataactacgtgcaacgaatttatattctgtatttta-3'
63 N623A_D	5'-taaagatgacgctacaaaagtagtcgcaaaaaaatcactatcccacaacg-3'
63 N623A_R	5'-cgtttggggatagtgatttttggcgactactttttagcgtcatcttta-3'
63 E577A_D	5'-caggtacataattacttagaagcaaaaaacagcctgctggtta-3'
63 E577A_R	5'-taaccagcaggctggtttggttcttaagtaaatatgtacctg-3'
Oligonucléotides de mutagenèse (lysines exposées de RrgB)	
63 K162A_D	5'-gaagatggagcaaccttaacaggttctgcagcagttccaattg-3'
63 K162A_R	5'-caattggaactgctgcagaacctgttaagggtgctccactctc-3'
63 K183A_D	5'-gtggatggcctgtgtatccagcaaaatcagaagcaaaagc-3'
63 K183A_R	5'-tttggcttggcttctgtatttggctggatacacatgcccac-3'
Oligonucléotides de mutagenèse (ponts intramoléculeaires RrgB)	
64 K155A_D	5'-acaatgacaacaagggtgaagctgatagcgggtgatcaagacc-3'
64 K155A_R	5'-ggctcttgatccaccgctatcagcttcaccttggctgctcattgt-3'
64 E222A_D	5'-actatcgtttcaaggcgggtggagccactggc-3'
64 E222A_R	5'-gccagtggtccaccgcttgaacagatagt-3'
64 N252A_D	5'-gggtgacgattacggttgtcgtcagaaattaccacgtggc-3'
64 N245A_R	5'-gccacgtggtaatttctgagcgaacaaccgtaatcgtcacc-3'
64 K264A_D	5'-gtggcaatggtgactttatggcgggtggatggcggaccaa-3'
64 K264A_R	5'-ttggctccgaccatccaccgccaataaagtaaacattgccac-3'
64 E322A_D	5'-gacatactatttatggcgctccaagctccaactg-3'
64 E322A_R	5'-cagttggagcttggagcggccataaatagtatgtc-3'
64 N354A_D	5'-aggaactggtaacagtggttaaaagctacaagcggaccagg-3'
64 N354A_R	5'-ccgtggtcgccttggtagctttaaccactggttaccagttcct-3'
Oligonucléotides de mutagenèse (délétions motifs LPxTG)	
R62ΔYPRTG	5'-gcggtcgactcattctctcttggaggaaatagg-3'
R63ΔIPQTG	5'-gcgctcgagctaaagtgatttttgggtgactactttttagcgtc-3'
R64ΔVPDTG	5'-gccaagctttcaatcaatccggtgctgctgttatttttaaccac-3'
Oligonucléotides de mutagenèse (inversions motifs LPxTG)	
D62IPQ	5'-caatgaacctattcctccaaagagagaattcctcaaaactgggtggtgagt-3'
R62IPQ	5'-actcaaccaccagtttgaggaatttctctcttggaggaaataggttcattg-3'
D62VPD	5'-aatgaacctattcctccaaagagagaagttcctgatactgggtggtgagtcgacaagctt-3'
R62VPD	5'-caagcttggcactcaaccaccagatcaggaacttctctcttggaggaaataggttcatt-3'
D63YPR	5'-cgctacaaaagtagtcaacaaaaaatcacttatccaagaacgggtggtctcagctctggt-3'
R63YPR	5'-accagactcgagaccaccgcttcttggataagtgatttttggtagactactttttagc-3'
D64YPR	5'-agtggttaaaaaataacaagcggaccaggtgattatccacgtacaggggaagaatgaaagc-3'
R64YPR	5'-gctttcattcttcccctgtacgtggataatcaatccggtgctgctgttatttttaaccact-3'

## Annexe 3 : Récapitulatif des souches d'expression réalisées pour cette étude

### A. Liste des transformations des vecteurs pLIM01 réalisées chez *E. coli* BL21 (DE3) RIL

Vecteurs	Protéines exprimées	Nomenclature utilisée
1	His <sup>+</sup> -RrgA	H6-RrgA
2	His <sup>+</sup> -RrgA	H6-RrgB
3	His <sup>+</sup> -RrgA	H6-RrgC

### B. Liste des co-transformations des vecteurs de base réalisées chez *E. coli* STAR

Vecteurs	Protéines exprimées	Nomenclature utilisée
4	H6-RrgA	RrgA
5	RrgB-S-tag	RrgB
6	H6-RrgC	RrgC
7	SrtC-1	SrtC-1
8	SrtC-2	SrtC-2
9	SrtC-3	SrtC-3
10	H6-RrgA ; RrgB-S-tag	RrgAB
11	H6-RrgC ; SrtC-1	RrgC-SrtC-1
12	SrtC-2 ; SrtC-3	SrtC-2 SrtC-3
4 + 6	H6-RrgA ; H6-RrgC	RrgAC
5 + 6	RrgB-S-tag ; H6-RrgC	RrgBC
10 + 6	H6-RrgA ; RrgB-S-tag ; H6-RrgC	RrgABC
4 + 7	H6-RrgA ; SrtC-1	RrgA - SrtC-1
5 + 7	RrgB-S-tag ; SrtC-1	RrgB - SrtC-1
10 + 7	H6-RrgA ; RrgB-S-tag ; SrtC-1	RrgAB - SrtC-1
4 + 11	H6-RrgA ; H6-RrgC ; SrtC-1	RrgAC - SrtC-1
5 + 11	RrgB-S-tag ; H6-RrgC ; SrtC-1	RrgBC - SrtC-1
10 + 11	H6-RrgA ; RrgB-S-tag ; H6-RrgC ; SrtC-1	RrgABC- SrtC-1
4 + 8	H6-RrgA ; SrtC-2	RrgA - SrtC-2
5 + 8	RrgB-S-tag ; SrtC-2	RrgB - SrtC-2
6 + 8	H6-RrgC ; SrtC-2	RrgC - SrtC-2
10 + 8	H6-RrgA ; RrgB-S-tag ; SrtC-2	RrgAB - SrtC-2
4 + 6 + 8	H6-RrgA ; H6-RrgC ; SrtC-2	RrgAC - SrtC-2
5 + 6 + 8	RrgB-S-tag ; H6-RrgC ; SrtC-2	RrgBC - SrtC-2
10 + 6 + 8	H6-RrgA ; RrgB-S-tag ; H6-RrgC ; SrtC-2	RrgABC- SrtC-2
4 + 9	H6-RrgA ; SrtC-3	RrgA - SrtC-3
5 + 9	RrgB-S-tag ; SrtC-3	RrgB - SrtC-3
6 + 9	H6-RrgC ; SrtC-3	RrgC - SrtC-3
10 + 9	H6-RrgA ; RrgB-S-tag ; SrtC-3	RrgAB - SrtC-3
4 + 6 + 9	H6-RrgA ; H6-RrgC ; SrtC-3	RrgAC - SrtC-3
5 + 6 + 9	RrgB-S-tag ; H6-RrgC ; SrtC-3	RrgBC - SrtC-3
10 + 6 + 9	H6-RrgA ; RrgB-S-tag ; H6-RrgC ; SrtC-3	RrgABC- SrtC-3

C. Liste des co-transformations des vecteurs mutés réalisées chez *E. coli* STAR

Vecteurs	Nomenclature utilisée
2 + 14	RrgBC-SrtC-1 C193A
10 + 14	RrgABC-SrtC-1 C193A
30 + 13	RrgB D241A -SrtC-1 C193A
10 + 15	RrgAB - SrtC-2 C179A
5 + 6 + 15	RrgBC - SrtC-2 C179A
10 + 6 + 15	RrgABC -SrtC-2 C179A
10 + 16	RrgAB - SrtC-3 C176A
5 + 6 + 16	RrgBC - SrtC-3 C176A
10 + 6 + 16	RrgABC - SrtC-3 C176A
25	RrgA K191A
26	RrgA D600A
27	RrgA N695A
28	RrgA E803A
29	RrgA D600A/E803A
30	RrgB D241A
31	RrgB E405A
32	RrgB E577A
33	RrgB D241A/E405A/E577A
36	RrgC K155A
37	RrgC E222A
38	RrgC N252A
39	RrgC K264A
40	RrgC E322A
41	RrgC N354A
42	RrgC E222A/E322A
5 + 43	RrgBC E222A-SrtC-1
5 + 44	RrgBC E322A-SrtC-1
5 + 45	RrgBC E222A/E322A-SrtC-1
30 + 11	RrgB D241A C - SrtC-1
31 + 11	RrgB E405A C-SrtC-1
32 + 11	RrgB E577A C-SrtC-1
33 + 11	RrgB D241A/E405/E577A C-SrtC-1
33 + 45	RrgB D241A/E40A/E5771 C E222A/E322A-SrtC-1
46 + 8	RrgA D600A B - SrtC-2
47 + 8	RrgA E803A B - SrtC-2
48 + 8	RrgA D600A/E803A B - SrtC-2
49 + 8	RrgAB D241A - SrtC-2
50 + 8	RrgAB E405A - SrtC-2
51 + 8	RrgAB E577A - SrtC-2
52 + 8	RrgAB D241A/E405A/E577A - SrtC-2
53 + 8	RrgA D600A/E803A B D241A/E405A/E577A - SrtC-2
54 + 11	RrgBΔIPQTG C -SrtC-1
5 + 59	RrgBCΔVPDTG - SrtC-1
57 + 8	RrgAΔYPRTG B -SrtC-2

Vecteurs	Nomenclature utilisée
55+ 8	RrgABΔIPQTG - SrtC-2
54 + 6 + 8	RrgBΔIPQTG C -SrtC-2
5 + 58 + 8	RrgBCΔVPDTG - SrtC-2
57 + 9	RrgAΔYPRTG B -SrtC-3
55 + 9	RrgABΔIPQTG - SrtC-3
54 + 6 + 9	RrgBΔIPQTG C -SrtC-3
5 + 58 + 9	RrgBCΔVPDTG - SrtC-3
60 + 8	RrgAΔYPRTG::IPQTG B - SrtC-2
61 + 8	RrgAΔYPRTG::VPDTG B - SrtC-2
62 + 6 + 8	RrgBΔIPQTG::YPRTG C - SrtC-2
5 + 63 + 8	RrgBCΔIPQTG::YPRTG - SrtC-2
34 + 8	RrgAB K162A - SrtC-2
35 + 8	RrgAB K183A - SrtC-2

# **Publications**





## PUBLICATIONS

### I. Publication 1

#### **Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*.**

Clothilde Manzano, Carlos Contreras-Martel, Lamya El Mortaji, Thierry Izoré , Daphna Fenel, Thierry Vernet, Guy Schoehn, Anne Marie Di Guilmi and Andréa Dessen  
*Structure* **16**, 1838-48 (2008).

#### *Contribution aux travaux*

*Les résultats présentés dans la publication 1 ne sont pas décrits dans la partie résultat du manuscrit de thèse, néanmoins ceux-ci ont été en partie obtenus au laboratoire. En effet, les clonages des sortases C-1 et C-3 ont été réalisés au LIM et les protocoles de purification de ces deux protéines ont été optimisés par nos soins. De même, les premiers tests de cristallogénèse ainsi que les premiers cristaux ont été obtenus avec le robot Cartesian de l'EMBL. La poursuite de l'étude cristallographique a été réalisée par Clothilde Manzano dans l'équipe d'Andréa Dessen. Par ailleurs, la validation in vivo de la fonction de la sortase C-1 dans la polymérisation de RrgB a été réalisée au laboratoire par construction d'une souche TIGR4ΔsrtC-1 dont le phénotype a été comparé à celui de la souche sauvage.*

### II. Publication 2

#### **Structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*.**

Thierry Izoré, Carlos Contreras-Martel, Lamya El Mortaji, Clothilde Manzano, Remi Terrasse, Thierry Vernet, Anne Marie Di Guilmi and Andréa Dessen  
*Structure* **18**, 106-15 (2010).

#### *Contribution aux travaux*

*Le clonage de la piline RrgA a été réalisé au LIM et le protocole de purification de la protéine a été optimisé au laboratoire. La mutagenèse des résidus clés impliqués dans la formation des ponts intramoléculaires a été réalisée par nos soins de même que l'analyse TSA et ESI-MS des protéines natives et de leurs variants.*

### III. Publication 3

#### **Stability and Assembly of Pilus Subunits of *Streptococcus pneumoniae***

Lamya El Mortaji, Remi Terrasse, Andréa Dessen, Thierry Vernet and Anne Marie Di Guilmi  
*J Biol Chem* **285**, 12405-15 (2010).

*Un article additionnel présentant la structure de RrgB a été soumis et un second article traitant des fonctions des sortases C-2 et C-3 est en préparation.*



# Sortase-Mediated Pilus Fiber Biogenesis in *Streptococcus pneumoniae*

Clothilde Manzano,<sup>1</sup> Carlos Contreras-Martel,<sup>1</sup> Lamy El Mortaji,<sup>2</sup> Thierry Izoré,<sup>1</sup> Daphna Fenel,<sup>3</sup> Thierry Vernet,<sup>2</sup> Guy Schoehn,<sup>4</sup> Anne Marie Di Guilmi,<sup>2,\*</sup> and Andréa Dessen<sup>1,\*</sup>

<sup>1</sup>Laboratoire des Protéines Membranaires

<sup>2</sup>Laboratoire d'Ingénierie des Macromolécules

<sup>3</sup>Laboratoire de Microscopie Electronique et Structurale

Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 (CEA, CNRS, UJF, PSB), 41 rue Jules Horowitz, F-38027 Grenoble, France

<sup>4</sup>Unit for Virus Host Cell Interactions, UMR5233 UJF-EMBL-CNRS, CIBB, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France

\*Correspondence: diguilmi@ibs.fr (A.M.D.G.), dessen@ibs.fr (A.D.)

DOI 10.1016/j.str.2008.10.007

## SUMMARY

*Streptococcus pneumoniae* is a piliated pathogen whose ability to circumvent vaccination and antibiotic treatment strategies is a cause of mortality worldwide. Pili play important roles in pneumococcal infection, but little is known about their biogenesis mechanism or the relationship between components of the pilus-forming machinery, which includes the fiber pilin (RrgB), two minor pilins (RrgA, RrgC), and three sortases (SrtC-1, SrtC-2, SrtC-3). Here we show that SrtC-1 is the main pilus-polymerizing transpeptidase, and electron microscopy analyses of RrgB fibers reconstituted *in vitro* reveal that they structurally mimic the pneumococcal pilus backbone. Crystal structures of both SrtC-1 and SrtC-3 reveal active sites whose access is controlled by flexible lids, unlike in non-pilus sortases, and suggest that substrate specificity is dictated by surface recognition coupled to lid opening. The distinct structural features of pilus-forming sortases suggest a common pilus biogenesis mechanism that could be exploited for the development of broad-spectrum antibacterials.

## INTRODUCTION

*Streptococcus pneumoniae*, the causative agent of pneumonia, bacteremia, otitis, and meningitis, is a major cause of community-acquired illnesses, especially in the very young and the elderly, and causes over 1.6 million deaths worldwide each year (Levine et al., 2006). The spread of antibiotic-resistant strains (Bronzwaer et al., 2002), coupled to the limited efficiency and/or the restricted strain coverage of currently commercialized vaccines (Bogaert et al., 2004; Spratt and Greenwood, 2000), underline the need for the study of novel therapeutic targets for pneumococcal infection.

Secreted virulence factors and extracellular appendages are not only key factors in the initiation of bacterial infection, but also noteworthy vaccine antigens (Gianfaldoni et al., 2007; Maione et al., 2005; Mora et al., 2005). Pili, hair-like, elongated fibers associated to the bacterial cell wall, play important roles

in adhesion, biofilm formation, competence for DNA transformation, and motility, and are thus key elements of a bacterium's pathogenicity arsenal (Burrows, 2005; Manetti et al., 2007; Mattick, 2002). Pili expressed by Gram-negative pathogens have been the targets of multiple studies that have revealed that these filaments are composed of non-covalently associated subunits; coordination of subunit association involves, in many cases, strand exchange mechanisms orchestrated by periplasmic chaperones (Sauer et al., 2002; Vetsch et al., 2004). The pilus formation process in Gram-positive bacteria, however, is still poorly understood. It is widely accepted that in pathogens such as *Corynebacterium diphtheriae*, the causative agent of diphtheria, *Streptococcus pyogenes*, which causes toxic shock syndrome, and *S. pneumoniae*, pili are formed through the covalent association of subunits by sortases that are encoded within pilus-specific pathogenicity islands (Budzik et al., 2008; LeMieux et al., 2006; Mandlik et al., 2008; Mora et al., 2005; Ton-That et al., 2004; Ton-That and Schneewind, 2003). Notably, sortases are membrane-associated transpeptidases that play roles in functions ranging from iron acquisition to sporulation (Marraffini and Schneewind, 2006; Mazmanian et al., 2003). The quintessential, best-studied sortase, SrtA, catalyzes the covalent attachment of surface-exposed proteins, such as virulence factors, to the *Staphylococcus aureus* peptidoglycan (Schneewind et al., 1992). Its catalytic mechanism involves recognition of an LPXTG-like motif on the target protein followed by nucleophilic attack of the Thr-Gly covalent bond by a catalytic Cys residue. The ensuing acyl-enzyme intermediate can only be resolved by attack from a free  $\epsilon$ -amino group from the secondary substrate that, in the case of SrtA, is the bridging residue of the stem peptide of the peptidoglycan (Marraffini et al., 2006; Ton-That et al., 2002). Pilus formation, however, requires dedicated sortases, and the precise roles these molecules play in the pilus biogenesis process are still poorly understood. It is of interest that the recent structure of the pilus-forming subunit of *S. pyogenes* revealed that intramolecular isopeptide bonds are used to stabilize the pilus building block, with one bond located in the vicinity of the sortase recognition site (Kang et al., 2007), although the precise role played by this specific arrangement in the pilus-forming process is still unclear.

In *S. pneumoniae* the entire macromolecular machinery responsible for pilus formation is encoded on a mobile pathogenicity islet (*rfrA*). Strains that lack the *rfrA* islet display less adherence to lung epithelial cells and are less virulent in murine models of

pneumonia and bacteremia, revealing that the pilus plays an important role in infection (Barocchi et al., 2006; Hava and Camilli, 2002). The *rtrA* islet harbors seven genes, three of which encode structural proteins (RrgA, RrgB, and RrgC) and three of which encode sortases (SrtC-1, SrtC-2, SrtC-3). Transmission electron microscopy images of infectious *S. pneumoniae* strains have revealed that the pilus shaft is composed of a RrgB polymer, with the minor pilins RrgA and RrgC associated to it in patches (Hilleringmann et al., 2008; LeMieux et al., 2006); in the case of RrgA, association has also been shown directly onto the cell wall (Barocchi et al., 2006). All six proteins are potentially linked to the cell membrane through a single, predicted transmembrane (TM) region. Notably, the three structural proteins carry an LPXTG-like motif immediately upstream of the TM region, which potentially enables them to be acted upon by the dedicated sortases. In addition, RrgB also displays a conserved Lys residue within a “pilin” motif that is present within most pilus-forming subunits from Gram-positive organisms. Thus, RrgB polymerization could proceed via a mechanism in which the LPXTG-like motif serves as a donor substrate, whereas a conserved Lys residue within a “pilin” motif potentially acts as the acceptor during a sortase-catalyzed transpeptidation reaction, in a process that would ensure stepwise, covalent incorporation of RrgB subunits into the growing fiber (Marraffini et al., 2006; Scott and Zahner, 2006; Telford et al., 2006). Participation of the  $\epsilon$ -amino group of a conserved lysine residue in a nucleophilic pilus-forming reaction has now been shown for *Bacillus cereus* (Budzik et al., 2008) and *Corynebacterium diphtheriae* (Mandlik et al., 2008), suggesting that this might be a conserved mechanism in Gram-positive bacteria.

The equal number of structural and sortase-encoding genes in pneumococci has prompted the “one protein, one sortase” proposition (Hilleringmann et al., 2008; LeMieux et al., 2006), in which one sortase protein would be responsible for the polymerization of the RrgB fiber, whereas the two others could play roles in associating RrgA and RrgC (to RrgB and potentially to each other). In this work, we show that SrtC-1 is the main fiber-forming transpeptidase, and SrtC-3 has a modest capability of forming RrgB oligomers. Purified RrgB fibers produced in vitro by SrtC-1 activity are structural mimics of the pneumococcal pilus backbone. In addition, the high-resolution crystal structures of the two sortases that display in vitro reactivity toward RrgB, SrtC-1 and SrtC-3, reveal substrate-recognition regions that are covered by a lid that maintains the active sites in closed conformation, suggesting that pilus formation is regulated by a mechanism that involves activation of sortase-mediated catalysis. The identification of sequences corresponding to lid regions in sortases from a variety of pilated Gram-positive bacteria, as well as the important role played by pili in infection, suggest that pilus-forming sortases could represent tractable targets for antibiotic development for a wide spectrum of bacterial pathogens.

## RESULTS

### RrgB Fiber Can Be Assembled In Vitro by a Pilus-Polymerizing Transpeptidase

The pneumococcal *rtrA* pathogenicity islet encodes all of the genes required for formation of the pilus-forming machinery, namely backbone fiber (*rrgB*) and ancillary proteins (*rrgA*,

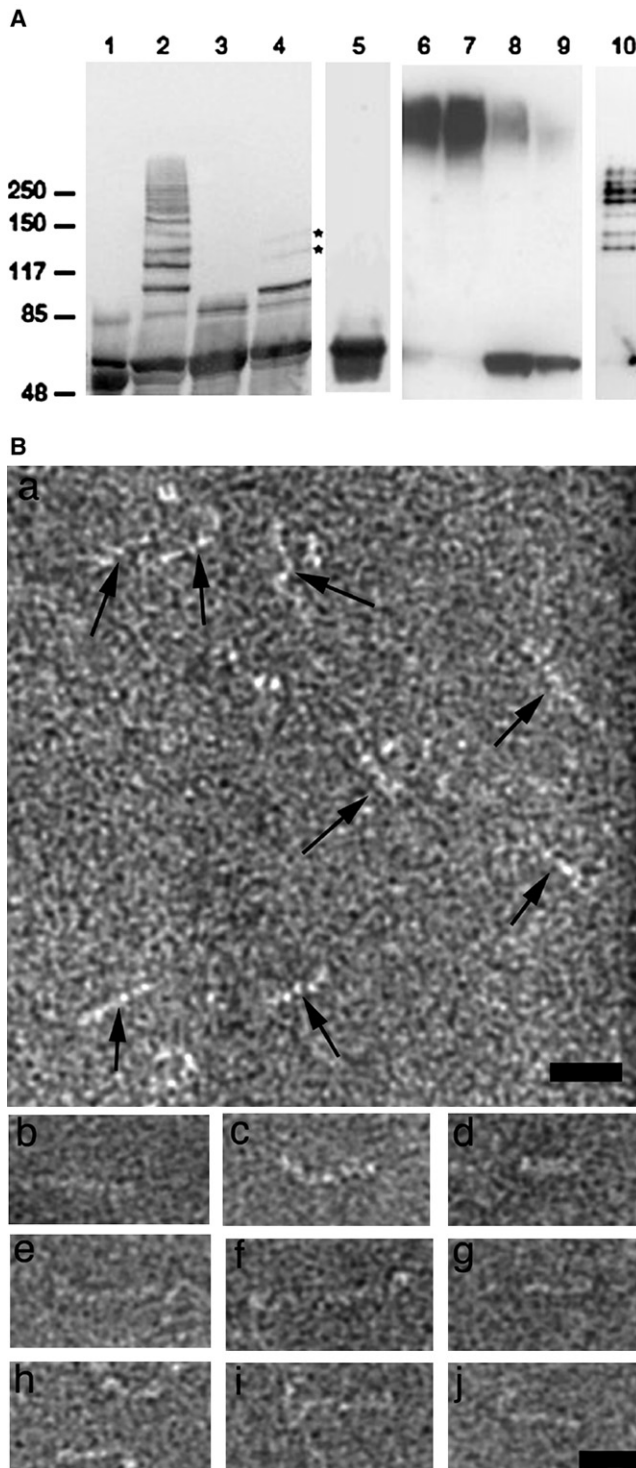
*rrgC*), as well as three sortases (SP0466, SP0467, SP0468, whose protein products are SrtC-1, SrtC-2, SrtC-3) (Barocchi et al., 2006; Hava and Camilli, 2002; LeMieux et al., 2008). To initiate our study of the pilus fiber formation process, we expressed the soluble forms of SrtC-1, SrtC-2, SrtC-3, and RrgB from the highly infectious *S. pneumoniae* serotype strain T4 (TIGR4) and tested the ability of each sortase to polymerize RrgB monomers. The presence of pili on the pneumococcal surface is characterized by a pattern of high-molecular-weight bands on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) upon analysis of surface preparations, which provide evidence of the covalent nature of the RrgB pilus backbone (Barocchi et al., 2006; Hilleringmann et al., 2008). Hence, the ability to form RrgB fibers in vitro was tested by searching for a pattern of high-molecular-weight bands on gradient SDS-PAGE upon incubation of RrgB monomers at 37°C with SrtC-1, SrtC-2, or SrtC-3. A ladder of bands with molecular weights above 100 kDa, which was recognized by antibodies raised against RrgB, could be generated when monomeric RrgB was incubated with SrtC-1 (Figure 1A, lane 2). In light of the fact that sortases catalyze transpeptidation reactions by employing a Cys nucleophile, we mutated the unique Cys residue within SrtC-1, Cys 193, into Ala, and tested this mutant protein for its ability to generate the RrgB high-molecular-weight ladder. Upon incubation with SrtC-1-Cys193Ala, RrgB remained monomeric, and no higher-molecular-weight bands could be detected (Figure 1A, lane 5), pointing to the fact that in vitro RrgB polymerization is a product of SrtC-1-related activity.

No evidence of RrgB polymerization could be identified upon incubation of RrgB with SrtC-2 (Figure 1A, lane 3); however, in the presence of SrtC-3, two faint bands could be reproducibly detected in the vicinity of the 117 kDa marker (Figure 1A, lane 4, indicated with stars), suggesting that SrtC-3 can catalyze the formation of higher-molecular-weight forms of RrgB, albeit much less efficiently than SrtC-1. These results suggest that although SrtC-1 is the most efficient transpeptidase in the RrgB polymerization process, SrtC-3 also displays a modest capability to form polymerized forms of RrgB.

To confirm these results, we deleted the gene encoding SrtC-1 (SP0466) from the *rtrA* islet in strain TIGR4. Total extracts of mutant TIGR4 as well as of the wild-type strain were treated with mutanolysin, loaded onto gradient polyacrylamide gels, and immunoblotted with antibodies specific for RrgB (Figure 1A, lanes 6–9). Wild-type TIGR4 presented high-molecular-weight bands characteristic of the presence of a polymerized pilus (Figure 1A, lanes 6 and 7), whereas the SP0466 deletion mutant presented a smaller quantity of fully polymerized pilus that was concomitant with an increase in monomeric RrgB. These results confirm that SrtC-1 plays a key role in the pilus polymerization process, and that its absence can be partly compensated by a “background” polymerization activity of other sortases.

### In Vitro Reconstitution of the Pneumococcal Pilus Backbone

To obtain structural insight into the RrgB fibers prepared in vitro, we performed a large-scale incubation of SrtC-1 and RrgB and gel filtered the mix. The RrgB fiber eluted early in the gel-filtration experiment, whereas SrtC-1 and unpolymerized RrgB both eluted in volumes indicative of monomers (see Figure S1



**Figure 1. SrtC-1 Is a Pilus-Polymerizing Transpeptidase**

(A) Western blots using polyacrylamide gradient gels developed with RrgB antiserum. Lanes: (1), monomeric RrgB; (2), RrgB + SrtC-1; (3), RrgB + SrtC-2; (4), RrgB + SrtC-3; (5) RrgB + SrtC-1Cys193Ala; (6) total extract from mutanolysed TIGR4; (7) supernatant from mutanolysed TIGR4; (8) total extract from mutanolysed TIGR4ΔSP0466; (9) supernatant from mutanolysed TIGR4ΔSP0466. Mutanolysed samples display pili that harbor not only polymerized RrgB, but also associated RrgA and RrgC, as well as peptidoglycan

available online). The high-molecular-weight sample (Figure 1, lane 10) was submitted to mass spectrometry (LCMS/MS) experimentation, which confirmed that the bands consisted of RrgB (Table S1); subsequently, it was analyzed by negative staining electron microscopy with a Philips CM12 microscope employing 2% sodium silicotungstate at pH 7.5. Electron micrographs show populations of fibers (Figure 1B, panel a) with roughly two distinct diameters: the most prevalent types resemble “beads on a string” and have diameters in the range of 3.5 nm, whereas the less common types of fibers have diameters of approximately 7.0 nm, thus being potentially representative of lateral associations of the thinner filaments (compare panels d and g in Figure 1B). Assuming that 1 image pixel = 1.17 Å, we estimate that the longest fibers have an approximate molecular weight of 480 kDa, whereas the shortest, of 140 kDa (see Experimental Procedures for precise calculations). In addition, several fibers display a “kink” (Figure 1B panels c, g, h, and i), giving evidence of flexibility. Notably, electron microscopy analyses of pili on the surface of *S. pneumoniae* have shown them to be flexible structures formed by 3.5-nm-thick protofilaments that intersect laterally to form pili with diameters of 6.8 nm or, in areas of more relaxed intersection, 9.5 nm (Hilleringmann et al., 2008). Our observations suggest that the fibers formed in vitro through the covalent association of RrgB monomers by SrtC-1 are structural mimics of the 3.5 nm and 6.8 nm filaments formed on the pneumococcal surface.

#### A Closed Substrate-Recognition Channel within Pilus-Forming Sortases

To gain insight into the catalytic aspects of the pilus-forming machinery at an atomic level, we crystallized and solved the high-resolution structures of SrtC-1 and SrtC-3. Both enzymes elute in volumes indicative of monomers in gel filtration, but crystallize as dimers in the asymmetric unit of a P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (SrtC-1) or P2<sub>3</sub> (SrtC-3) cell. The structure of SrtC-1 was solved by performing a 4 wavelength MAD experiment on a selenomethionylated SrtC-1 crystal at the ESRF synchrotron in Grenoble; this model was subsequently used to phase native data to 1.24 Å. SrtC-3 data were collected to 2.2 Å on an in-house X-ray generator and the structure was solved by employing the refined structure of SrtC-1 (with modifications; see Experimental Procedure) as a search model in a molecular replacement experiment. Data collection and refinement statistics can be found in Tables 1 and 2.

SrtC-1 (residues 21–214; Figure 2A) is composed of eight central, mostly anti-parallel β-strands that associate to form a β barrel that is surrounded by three major α helices and three 3<sub>10</sub> helices. Helices α1, α2, and α3 form the roof of the SrtC-1 β-barrel, with the region that includes α3 forming a “lid” that covers the active site (pink in Figures 2A and 2C). SrtC-3 (residues 39–241; Figures 2B and 2C), despite sharing 26.4% sequence identity with SrtC-1, displays a very similar fold to the latter enzyme (rms deviation of 1.3 Å over 138 Cα atoms), with

fragments. (10) Gel-filtered sample analyzed by electron microscopy and LCMS/MS.

(B) Top, negative staining electron microscopy images of the RrgB fibers generated by SrtC-1 activity. Black arrows point to fibers whose diameters are in the range of 3.5 nm or 7.0 nm. Bottom, a gallery of typical fibers seen on grids. Scale bar, 20 × 5 nm.

**Table 1. Data Collection and Phasing Statistics for SrtC-1 and SrtC-3**

Data Collection						
Data set	Peak	Inflection Point	High Remote	Low Remote	SrtC-1 Native	SrtC-3 Native
Wavelength (Å)	0.98055	0.980750	0.982450	0.978550	0.93100	1.5418
Space Group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P23
a (Å)	68.286	68.619	68.757	68.662	68.091	122.638
b (Å)	70.070	70.366	70.469	70.337	70.189	122.638
c (Å)	86.187	86.647	86.807	86.673	86.542	122.638
Resolution (Å)	1.73 (1.83–1.73)	1.86 (1.98–1.86)	1.86 (1.98–1.86)	1.86 (1.98–1.86)	1.24 (1.32–1.24)	2.14 (2.27–2.14)
No. obs/unique ref.	297168/80448	240375/65317	240388/65393	236792/64343	1493780/114617	205227/30999
Completeness (%)	96.3 (88.0)	96.7 (89.3)	95.6 (84.8)	95.1 (83.5)	98.3 (93.9)	90.7 (75.8)
R <sub>sym</sub> (last shell)	5.9 (37.4)	5.7 (44.6)	6.1 (63.6)	7.5 (83.4)	4.7 (50.7)	10.0 (51.0)
//σ(I) (last shell)	13.4 (3.6)	13.7 (3.2)	12.1 (2.3)	10.1 (1.8)	33.5 (5.1)	15.3 (4.4)
SrtC-1 Phasing						
SHARP FOM (acentric/centric)	0.43849/0.14854	0.43849/0.14854	0.43849/0.14854	0.43849/0.14854		
Phasing power ISO (acentric/centric)		0.253/0.208	0.844/0.879	0.170/0.160		
Phasing power ANO	1.807	1.110	0.102	0.643		

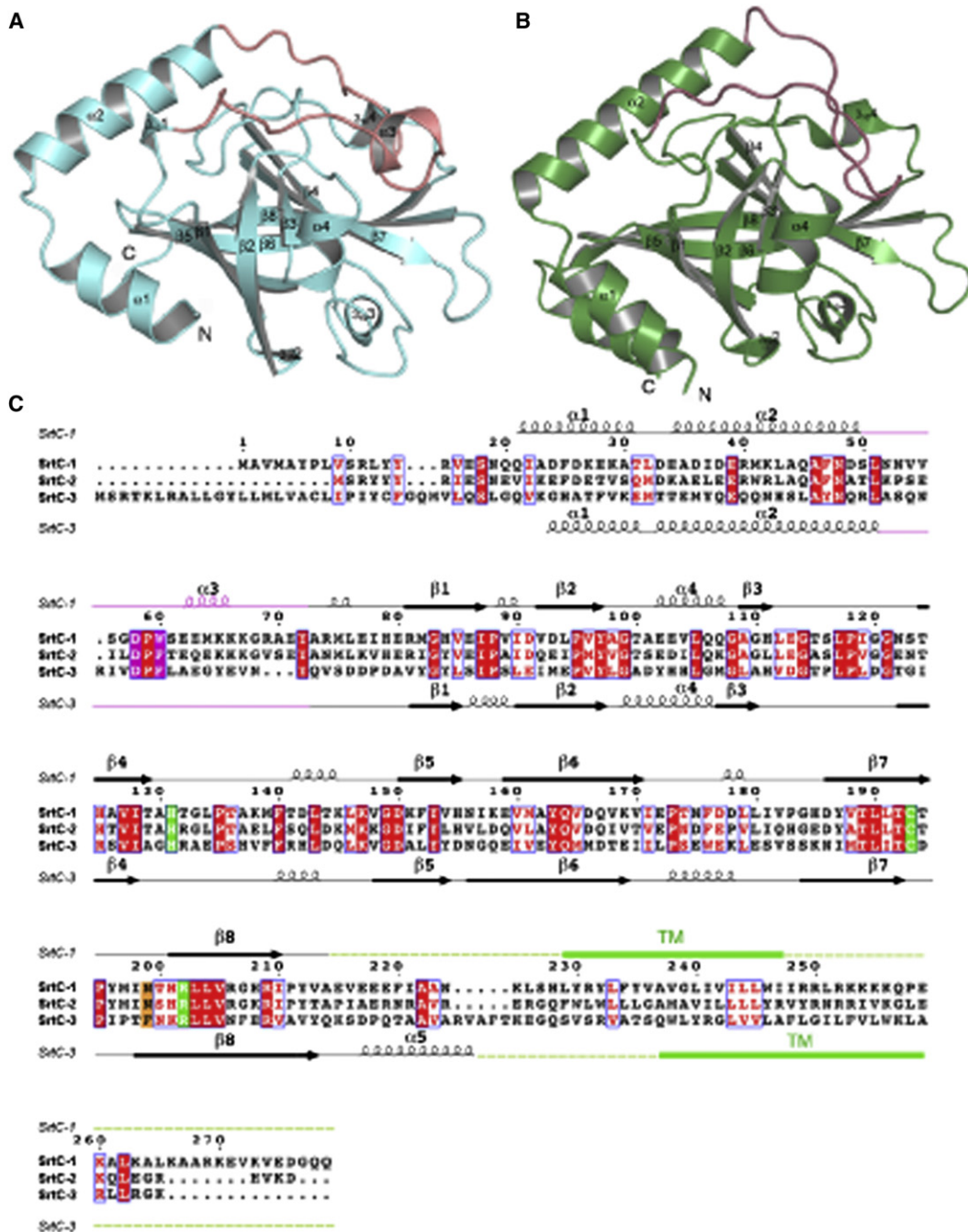
the major difference being the presence of an additional  $\alpha$ -helix at the C terminus of SrtC-3. SrtC-3 also displays a “lid” over the active site region; however, in both SrtC-1 and SrtC-3, the lid regions display the highest temperature factors in the structures, suggesting that they could be flexible or mobile elements. Analysis of the SrtC-3 structure reveals that residues 68–71 and 78–82, located on the SrtC-3 lid, display B-factors of 47.4 Å<sup>2</sup> and 44.8 Å<sup>2</sup>, respectively; these values are higher than for those of the intervening region (residues 73–76, B-factor = 20.2 Å<sup>2</sup>) as well as for the overall structure B-factor (23.8 Å<sup>2</sup>; Figure S2). These observations suggest these residues could play the roles of hinges in a potential lid opening mechanism (see below). The central section of SrtC-1 and SrtC-3, which encompasses the  $\beta$ -barrel, is reminiscent of that seen in the structure of sortases involved in association of proteins to the peptidoglycan (SrtA) or iron-heme transport (Ilangovan et al., 2001; Zhang et al., 2004). However, the arrangement of  $\alpha$  helices in the surrounding regions is distinct, and the lid seen in the structures of the pilus-

forming sortases related here is absent from all other sortases whose structures have been solved to date (Figure S3). In addition, biochemical studies performed on SrtA from *S. aureus* have revealed that its  $\beta$ 6/ $\beta$ 7 loop is a potential substrate specificity determinant (Bentley et al., 2007), and it is affected by the presence of Ca<sup>2+</sup> (Naik et al., 2006). Although the specificity of the  $\beta$ 6/ $\beta$ 7 region of SrtC-1/SrtC-3 was not studied here, we did not observe any effect of Ca<sup>2+</sup> on RrgB fiber-forming activity (not shown).

All sortases recognize the LPXTG motif of a target protein, employ their catalytic cysteine in the nucleophilic attack of the Thr-Gly peptidic bond, and subsequently catalyze a transepeptidation reaction by employing the free amino group from an acceptor substrate (Ton-That et al., 2002; Zong et al., 2004a). The SrtC-1 active site includes a Cys-His-Arg triad that has been implicated in catalysis in other surface-attachment sortases: Cys 193, at the C terminus of  $\beta$ 7; Arg 202, at the N-terminus of  $\beta$ 8; and His 131, within the loop that follows  $\beta$ 4 (Figure 3A). The side chain of Cys 193 is present in two conformations (in both monomers of the asymmetric unit) and was modeled with 0.5 occupancy in each case. In one of the conformations, it points away from the center of the active site and makes a unique, 3.1 Å hydrogen bond with the O $\delta$ 1 atom of Asn 199, itself located on the loop between  $\beta$ 7 and  $\beta$ 8. In the second conformation, the S $\gamma$  moiety of Cys 193 points directly into the active site, and makes a weak 3.4 Å hydrogen bond with the NH1 atom of Arg 202. Notably, as is the case for other sortases whose structures have been solved to date (Ilangovan et al., 2001; Zhang et al., 2004; Zong et al., 2004b), the nucleophilic cysteine does not interact directly with His 131 (the closest distance is 5.4 Å). However, it is of note that the lid contributes two residues that are anchored within the active site: Asp 58 and Trp 60. Asp 58 guarantees that the Arg 202 faces the Cys nucleophile by interacting with both NH2 and N $\epsilon$  atoms, whereas Trp 60, in addition to considerably increasing the hydrophobic nature of the cleft, interacts with the N $\delta$ 1 moiety of His 131 through its N $\epsilon$ 1 atom.

**Table 2. Refinement Statistics for SrtC-1 and SrtC-3**

Refinement	SrtC-1	SrtC-3
Resolution (Å)	1.24	2.14
R <sub>work</sub> (%)	17.31	18.05
R <sub>free</sub> (%)	19.31	24.14
No. of protein atoms	3076	3340
No. of solvent atoms	471	505
No. of glycerol molecule	1	0
RMS deviation, bond lengths (Å)	0.007	0.028
RMS deviation, bond angles (°)	1.250	2.296
Mean B-factor (Å <sup>2</sup> )	16.66	37.70
Res. most favored/allowed regions of Ramachandran (%)	100	100



**Figure 2. Sortases of the Pilus-Formation Machinery Display Similar Folds**

(A and B) SrtC-1 (A) and SrtC-3 (B) fold into  $\beta$ -barrels surrounded by helices; their active sites, centered on the region surrounding the C terminus of  $\beta 7$  and the N terminus of  $\beta 8$ , are covered by a lid (shown in pink).

(C) Structure-based sequence alignment of the three pilus-forming sortases of *S. pneumoniae*. Identical residues are shown with a red background, whereas similar residues are shown in red and highlighted with blue boxes. Residues located within the active site cleft of SrtC-1/SrtC-3 are highlighted in green, whereas the Asn residue on  $\beta 7/\beta 8$  that contacts the nucleophilic Cys in SrtC-1 is shown with an orange background. The lid region is shown in magenta. The TM (transmembrane) region was predicted with the SMART server ([www.smart.embl-heidelberg.de](http://www.smart.embl-heidelberg.de)).

The active site of SrtC-3 shows clear similarities with that of SrtC-1; Cys 206, Arg 215, and His 144 occupy analogous positions to their counterparts in SrtC-1. In this case, however, a single conformation of the Cys 206 side chain was detected, and it

points within the active site, toward Arg 215 (Figure 3B). This difference could be caused by the fact that Asn 199 of SrtC-1 is has been substituted by a Phe side chain in SrtC-3, and thus the Cys-Asn polar interaction observed in SrtC-1 is not possible



here. In addition, the SrtC-3 lid gives evidence of being more flexible than that of SrtC-1, with only the region encompassing the anchor residues being clearly traceable in the electron density map. These residues include Asp 73, which, as is the case in SrtC-1, interacts directly with Arg 215, and Phe 75, which points snugly into the pocket and guarantees its hydrophobic nature (much like Trp 60 in SrtC-1). It is of note that in both structures, the Asp residue located at the edge of the lid locks the Arg side chain in place, positioning it in close proximity to the nucleophilic Cys. Interestingly, the structures of surface-attachment sortases have revealed active sites that are not covered by lids and are much more solvent-exposed than those observed here (Figure 3C). In addition, these exposed active clefts display catalytic residues whose side-chain positions are not constrained; in SrtA from *S. aureus*, even in the presence of substrate, the side chain of the catalytic Arg residue points away from the nucleophilic Cys (Zong et al., 2004a). This observation underlines the possibility that the role of the interaction between the catalytic Arg and the Asp side chain located on the SrtC-1/SrtC-3 lid is not only to “lock” the lid in place in the absence of substrate, but also to guarantee that the Arg side chain will be appropriately positioned, potentially for oxyanion transition state stabilization (Frankel et al., 2007), once the lid is removed and substrate is bound. Thus, pilus-forming sortases present tightly knit active sites whose access is restricted by the presence of a potentially flexible lid, and whose residues are pre-positioned for catalysis.

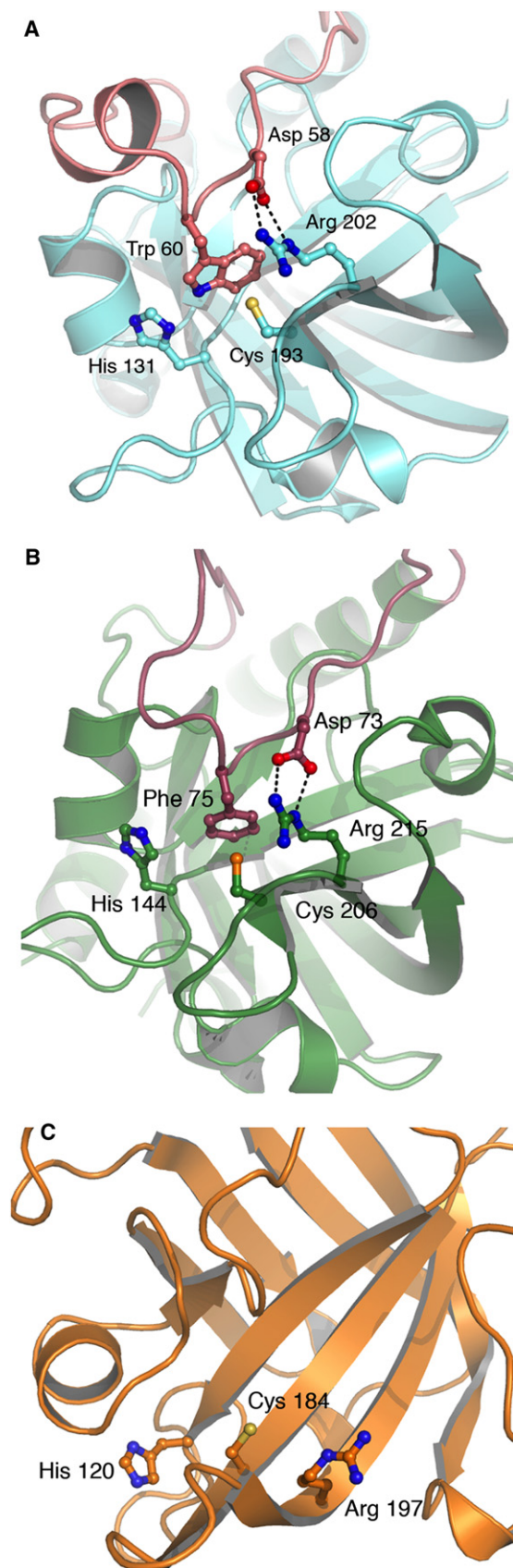
## DISCUSSION

Although pili have been identified on the surface of approximately 30% of the strains of the major human pathogen *S. pneumoniae* and shown to be involved in pathogenicity, little is known about the complex macromolecular machinery that regulates their biogenesis. In this work, we have identified that formation of the pilus backbone in the pneumococcus, known to be composed of the covalently linked subunits of RrgB, is catalyzed by a pilus-polymerizing enzyme, SrtC-1, the first sortase within the *rlrA* pathogenicity islet. We show that SrtC-1 is able to polymerize monomeric RrgB in vitro, whereas the third sortase encoded by the islet, SrtC-3, displays a modest capacity to associate RrgB monomers. A pneumococcal strain in which the SP0466 gene was knocked out displays a clearly diminished capacity of producing polymerized pili on its surface as well as an increased amount of RrgB monomers associated to the bacterial cell wall. Interestingly, a small amount of pili can still be detected on the surface of the TIGR4ΔSP0466 strain, suggesting that the absence of this pilus-polymerizing transpeptidase can be partly compensated by another sortase. These observations are corroborated by those of LeMieux and coworkers (LeMieux et al., 2008), who show that pneumococcal strains lacking individual pilus-forming sortases can still express pili, although the amount of surface-associated pili formed was not explored in their study.

In addition, we show that the RrgB fibers can be reconstituted in vitro from purified RrgB monomers and SrtC-1. In this sense, SrtC-1 acts as a catalyst to promote the covalent association of RrgB monomers independent of cellular energy. Pilus fibers produced in vitro and visualized in this work by electron microscopy

have approximately the same diameters as the single-coiled (3.5 nm) or double-coiled coil (6.8 nm) protofilaments identified and measured on the surface of pneumococci (Hilleringmann et al., 2008). Nevertheless, the assemblies produced in vitro (average length 35 nm) are shorter than the longest pili yet visualized in vivo, which have been reported as having lengths of up to 1 μm (Hilleringmann et al., 2008). It is of interest, however, that surface-associated pili are often heterogeneous in length, with short forms also being visible in several instances (Barocchi et al., 2006; Hilleringmann et al., 2008; Nallapareddy et al., 2006; Rosini et al., 2006; Weerkamp et al., 1986). It is thus possible that the structures isolated and visualized in this work could represent the minimal building blocks of the pilus fiber. However, it is also possible that the RrgB fibers formed in vitro are shorter than their wild-type counterparts because SrtC-1 polymerizes RrgB inefficiently (as shown by the finite number of high-molecular-weight bands on Figure 1A, lane 2). This could indicate that efficient fiber formation requires the concomitant participation of other members of the pilus-forming machinery, such as the other sortases. However, when this possibility was tested in vitro by incubation of RrgB with SrtC-1, SrtC-2, and SrtC-3 in different combinations, the pattern of formed fibers was identical to that observed with SrtC-1 only (not shown). It is thus conceivable that, in vivo, fiber formation is a well-orchestrated event that requires that the entire pilus-forming macromolecular machinery be associated to the bacterial outer membrane in order to optimize partner recognition events and catalysis, an organization that cannot easily be emulated in vitro.

Our results also reflect the potential requirement that SrtC-1 be maintained in an “activated” state through the entire fiber-formation process. Evidence for this possibility comes from the mapping of conserved/identical residues between sortases on the surface of SrtC-1, and comparison with the surface diagram of a well-studied sortase, SrtA from *S. aureus* (Figures 2C and 4). All sortase isoforms studied to date possess a highly conserved Thr-Leu-X-Thr motif that precedes the Cys nucleophile, and is located along β7 (shown in orange and red in Figure 4). Modifications within this sequence in SrtA suggest that it plays a role in substrate recognition (Frankel et al., 2007), a conclusion that is supported by the crystal structure of SrtA in complex with the LPETG peptide, which reveals that the substrate is located in proximity to this sequence (Zong et al., 2004a) (Figure 4B). Both SrtC-1 and SrtC-3 also possess this highly conserved region within an elongated cleft; however, in the pilus-forming sortases, this zone is covered by the lid, whose function also involves providing an anchor for the active site (shown in dark blue). These observations suggest that, in SrtC-1/SrtC-3, a closed lid (as seen in the structure) potentially prevents recognition of LPXTG-like sequences, necessitating a mechanism of lid opening in order for catalysis to occur. This process could be initiated by an initial recognition between the Rrg-substrate and the closed sortase; only binding of the appropriate substrate would lead to lid opening and enzyme activation. This model would be in agreement with the one recently put forward by Mandlik and coworkers (Mandlik et al., 2008), who suggest that, in order for proper pilus assembly to occur, two distinct sortases must act sequentially: the pilus-forming and the housekeeping enzymes, the latter of which will recognize lipid II as a substrate and anchor the formed polymers to the cell wall. Because, as



**Figure 3. The Active Site of Pilus-Forming Sortases Is Anchored by a Lid**

(A) The nucleophilic Cys 193 in SrtC-1 displays two conformations, but only one is shown for clarity.

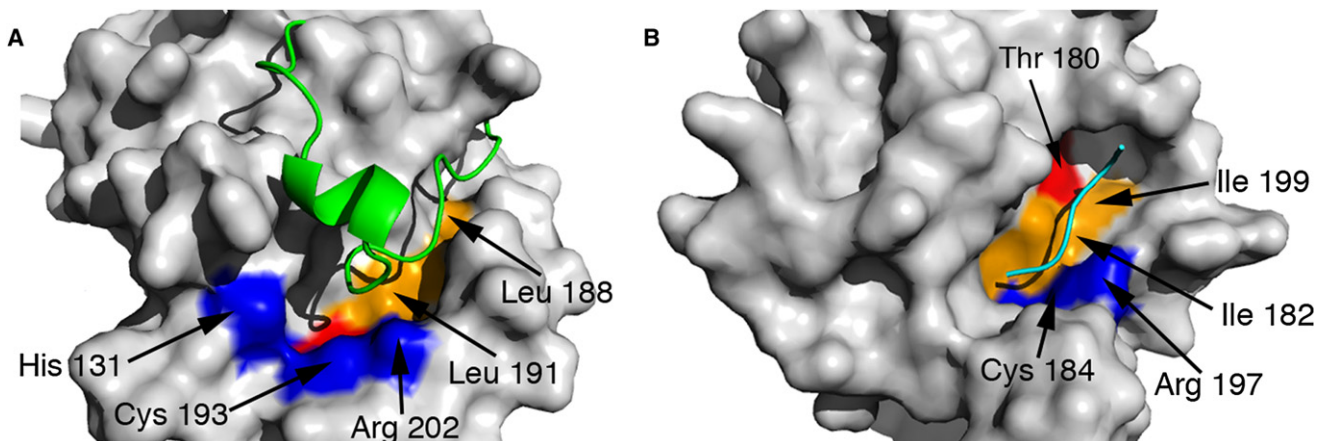
(B) The active site of SrtC-3, like SrtC-1, displays a catalytic Arg residue whose side chain is stably anchored by an Asp residue on the lid “anchor” region. In both active sites, the conserved His residue is approximately 5.4 Å away from the Cys nucleophile.

(C) The active site of SrtA from *S. aureus* (Protein Data Bank code 1T2P) is mostly solvent-exposed and does not harbor a lid.

shown in this work, the active sites of both sortase types are very similar, encapsulation of the catalytic cleft of the pilus-forming enzyme by a lid that can only be opened by the appropriate pilin subunit would ensure that lipid II is not recognized, thus guaranteeing proper orchestration of polymerization and cell wall attachment steps.

All three structural proteins that compose the pneumococcal pilus carry LPXTG-like motifs on their C termini, albeit with slightly distinct sequences (RrgA: YPRTG; RrgB: IPQTG; RrgC: VPDTG), which in principle are the substrates of a comparable transpeptidation reaction. Their exquisite specificity was illustrated by the elegant work performed by LeMieux and coworkers (LeMieux et al., 2006), who showed that exchange of the YPRTG motif of RrgA into the analogous motifs of RrgB or RrgC abrogated association of RrgA onto the RrgB fiber. Hence, slight modifications within the LPXTG-like sequences might have profound consequences for catalysis. To gain insight into the question of specificity of pilus-forming sortases, we compared surface potentials of all three enzymes after generating a model of SrtC-2 based on the structure of SrtC-1 (with which it shares 55.7% sequence identity; Figure 2C). This analysis (Figure 5) reveals clear differences within the active sites of SrtC-1, SrtC-2, and SrtC-3, as well as the regions that surround them. Most striking is the vast acidic region on the surface of SrtC-3, which contrasts sharply with the much smaller acidic region on the surfaces of SrtC-1 and SrtC-2, centered on Glu 103 and Glu 94, respectively; these differences are located on the closed lid regions. In addition, the catalytic clefts themselves also show differences, with that of SrtC-1 being more basic than that of SrtC-2, whereas that of SrtC-3 displays a clear negative charge. Thus, sortase specificity and differential recognition of RrgA, RrgB, and/or RrgC potentially involve interactions not only within the active site, but also with the surrounding region, which includes the flexible lid, as suggested above. It is of interest, thus, that our experiments show that both SrtC-1 and SrtC-3 can associate RrgB monomers, confirming that a certain level of “cross-specificity” exists among pilus-forming sortases. In light of our results, we suggest that this might arise as a consequence of the considerable flexibility observed for the lid region, coupled to the conservation of active site residues and the nature of the transpeptidation reaction catalyzed. These observations give the sortase lid region a key role in the processes of pilin substrate recognition and initiation of catalysis.

To verify if these observations could be extended to the pilus biogenesis process in other Gram-positive organisms, we performed a BLAST analysis using the pneumococcal SrtC-1 sequence and searched for the telltale lid anchor residues



**Figure 4. The Substrate-Binding Cleft of Sortases Is Conserved**

(A) Surface of SrtC-1, onto which residues that are conserved (orange) and identical (red) among the three pilus-forming sortases and have been mapped. The analogous residues were mapped onto the surface of *S. aureus* SrtA (B). This analysis identifies that the lid of pilus-forming sortases (shown in green) is positioned analogously to the LPETG substrate peptide (shown in cyan) in the structure of SrtA, suggesting that the substrate-recognition cleft in pilus-forming sortases is sterically blocked by the lid. Active site residues are shown in dark blue.

(Asp-Pro-Hyd) located approximately 125 amino acids N terminal to the nucleophilic Cys (magenta in Figure 2C). Interestingly, we were able to identify this sequence in pilus-forming sortases from a variety of other pilated species, including *Streptococcus agalactiae*, *Streptococcus pyogenes*, *Bacillus cereus*, *Enterococcus faecium*, *Actinomyces naeslundii*, and the well-studied *Corynebacterium diphtheriae* (Figure S4). These observations suggest that Gram-positive organisms could employ a universal mechanism for pilus biogenesis, where sortases encoded by the pilus-related pathogenicity islets harbor encapsulated active sites, thus necessitating substrate-induced activation for pilus formation.

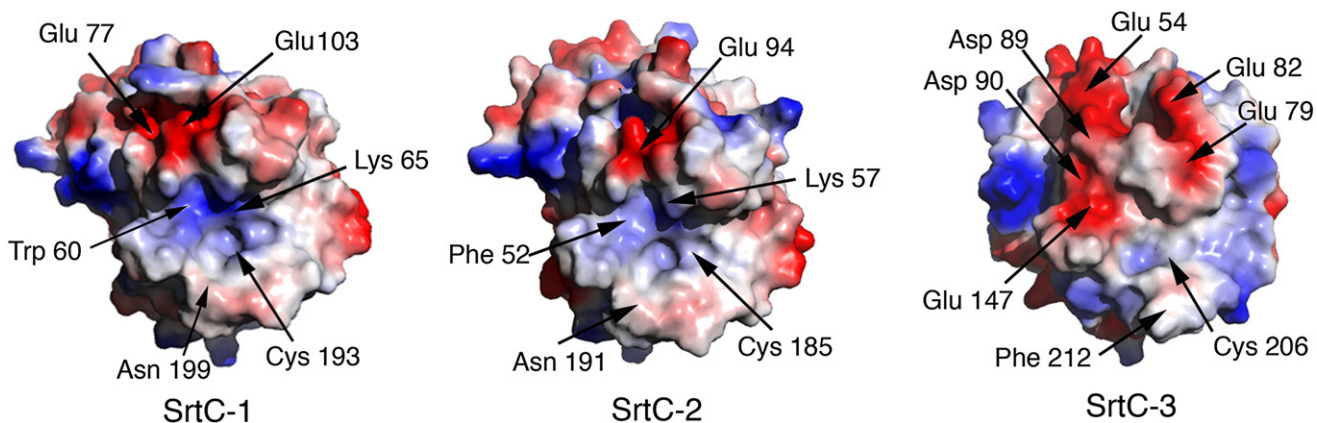
The targeting of pathways essential for bacterial survival by antibiotics has a strong link to the development of drug resistance. Sortases are not essential for bacterial survival, but they allow bacteria to express surface-associated proteins that play important roles in virulence. Developing molecules that challenge a bacterium's capability to successfully employ its pathogenicity

arsenal during infection could be an alternative approach toward antibiotic development (Walsh, 2003). The absolute requirement for sortases for pilus formation in Gram-positive pathogens, coupled to active site features that clearly distinguish them from classic sortases and to the fact that they are surface-exposed and thus potentially ligand-accessible, suggests that these enzymes are tractable, potential targets for novel antibiotic development for a variety of Gram-positive species.

#### EXPERIMENTAL PROCEDURES

##### Cloning, Expression, and Purification of RrgB, SrtC-1, SrtC-2, and SrtC-3

The regions of the *S. pneumoniae* TIGR4 coding for amino acids 30–633 (RrgB), 17–228 (SrtC-1), 9–220 (SrtC-2), and 32–254 (SrtC-3) were amplified using conventional PCR methodologies and cloned into vector pLIM (Protein Expert, Grenoble). Resulting vectors were transformed into *E. coli* RIL cells (Invitrogen). In all four cases, protein expression was induced in Terrific Broth with 1 mM IPTG at 37°C during 3 hr. Cells were harvested by centrifugation



**Figure 5. Surface Potential Representations of All Three Pilus-Forming Sortases, with a Direct View onto the Active Site**

The structure of SrtC-2 was generated by building a model based on the structure of SrtC-1, with which it shares 55.7% sequence identity. Striking features include the highly basic character of the SrtC-1 cleft and the vast acidic region in the vicinity of the SrtC-3 active site (same orientation as in Figure 3).

and lysed by sonication in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM imidazole, 1 mM PMSF, 0.1  $\mu$ M aprotinin, and 1  $\mu$ M pepstatin. The lysates were clarified by centrifugation and applied onto HisTrap<sup>TM</sup>HP columns (Amersham Biosciences) pre-equilibrated in lysis buffer. Protein elution was performed using linear gradients of 20–500 mM imidazole over 20 ml. Protein-containing fractions were dialyzed 2 hr into 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 20 mM imidazole at 4°C and subsequently incubated with 1:10 (w/w) Tev protease overnight at 15°C. The proteins were reloaded onto HisTrap<sup>TM</sup>HP columns and the eluted, cleaved products were loaded onto HiLoad 16/60 Superdex 200 (Amersham Biosciences) in 50 mM HEPES (pH 7.5) and 150 mM NaCl. Pooled, concentrated fractions were employed in crystallization trials and RrgB fiber formation experiments as described below.

Selenomethionylated SrtC-1 was overexpressed in *E. coli* RIL cells in minimal medium supplemented with thiamine (0.2 mg/ml), leucine (50 mg/l), valine (50 mg/l), isoleucine (50 mg/l), lysine (100 mg/l), phenylalanine (100 mg/l), threonine (100 mg/l), and selenomethionine (60 mg/l). Expression and purification was performed as described above for the native SrtC-1.

#### In Vitro RrgB Fiber Formation and Electron Microscopy Analyses

Recombinant, purified RrgB and SrtC-1 were mixed at a ratio of 1:2 (w/w) and incubated overnight at 37°C. RrgB was also incubated with SrtC-2 and SrtC-3 at the same ratio and under the same conditions, in a total volume of 40  $\mu$ l. Reactions were stopped by boiling for 10 min in the presence of Coomassie blue dye and 1% SDS. Aliquots of the mixtures were subsequently analyzed on 4%–12% SDS-PAGE gradients and then subjected to immunoblotting using a 1:5000 dilution of mouse antibodies raised against RrgB. After washing, membranes were incubated with an anti-mouse antibody conjugated to horseradish peroxidase at a 1:10,000 dilution. An immunoreactive signal was detected by chemiluminescence (ECL, Biorad).

Samples destined to electron microscopy analyses were prepared as described above and subsequently loaded onto a Superdex 200 gel filtration column (GE Healthcare). All eluted fractions were tested by immunoblotting with anti-RrgB antibodies, and those that corresponded to the high-molecular-weight forms (Figure 1A) were submitted to negative staining electron microscopy analyses.

Four microliters of the protein sample at approximately 0.05 mg/ml were adsorbed onto the clean face of a carbon film on a mica sheet (carbon/mica interface) and negatively stained with 2% (w/v) neutral sodium silicotungstate. Micrographs were taken under low-dose conditions with a CM12 LaB6 electron microscope working at 120 kV and with a nominal magnification of 60,000X. Electron micrographs were digitized using a Zeiss scanner with a pixel size of 7 nm (1.17 Å at the sample scale).

Volumes of individual protein fibers were calculated by assuming that 1 pixel = 1.17 Å (and thus 50 Å = 43 pixels), and that the average density of a protein is 0.86 Da/Å<sup>3</sup>. The calculated volume will thus correspond to the length of each individual fiber (L) multiplied by the surface of the section ( $\pi \times r \times r$ , where r represents the fiber at half-height). Thus:

for panel (a), 290 Å  $\times$  50 Å; L = 290 Å; r = 25 Å  $\rightarrow$  480 kDa

for panel (d), 140 Å  $\times$  54 Å; L = 140 Å; r = 27 Å  $\rightarrow$  275 kDa

for panel (h, bottom), 120 Å  $\times$  42 Å; L = 120 Å; r = 21 Å  $\rightarrow$  140 kDa

The fibers shown in the EM images thus have masses that range from 140 kDa to approximately 480 kDa.

#### Construction and Analysis of the SP0466 Deletion Mutant

The SP0466-inactivated TIGR4 strain was constructed by PCR insertion of the *cat* cassette directly into the genome. The PCR product was transformed into the wild-type TIGR4 strain, as previously described (Bricker and Camilli, 1999) and then plated onto Columbia blood agar plates containing chloramphenicol (4  $\mu$ g/ml). Transcription of the downstream genes (SP0467, SP0468) was verified by analysis of the total RNA of the mutant, demonstrating no polar effect as a consequence of the SP0466 deletion (data not shown). Wild-type TIGR4 and deletion strains were grown in Todd-Hewitt broth (BD) supplemented by 0.5% yeast extract (Sigma) at 37°C, 5% CO<sub>2</sub>, until an OD<sub>600nm</sub> of 0.45 A.U. Cultures were centrifuged and the bacterial pellet, after being washed with PBS,

was treated with 200 U/ml mutanolysin (Sigma), 1 mg/ml lysozyme (Sigma), and a protease inhibitor cocktail (Complete, Roche) for 3 hr at 37°C. Cellular debris were removed by centrifugation at 11 000 rpm for 15 min.

Thirty microliters of each sample was mixed with XT Sample Buffer and XT Reducing Agent (Bio-Rad), boiled at 100°C for 10 min and loaded onto 4%–12% Criterion XT Precast gels (Bio-Rad). Gels were run for approximately 3 hr and subsequently electrotransferred in Trans-Blot Transfer Medium (Bio-Rad). Incubation times of 1 hr were successively performed using anti-RrgB polyclonal mouse antibodies (diluted 1:5000) and anti-mouse HRP conjugate antibodies (Sigma, diluted 1:120,000) before detection with a chemiluminescent substrate (Pierce).

#### Crystallization, Data Collection, Structure Solution, and SrtC-2 Modeling

Crystals of native SrtC-1 were obtained by the hanging-drop vapor diffusion method in 100 mM Tris-HCl (pH 8.5), 200 mM MgCl<sub>2</sub> hexahydrate, and 30% w/v polyethylene glycol 4000 at 20°C. Crystals were cryoprotected by brief incubation in mother liquor containing increasing concentrations of glycerol (up to 15%), and subsequently were flash-cooled in the nitrogen stream. A first native data set (1.24 Å) was collected at the European Synchrotron Radiation Facility (ESRF) beamline ID14-EH3 (Grenoble, France). A four-wavelength MAD experiment was performed on the Se edge on the ESRF ID23-EH1 beamline.

Diffraction images were indexed and scaled with XDS (Kabsch, 1993) and merged with the CCP4 6.0.2 program suite (CCP4, 1994). Identification of selenium atom positions, refinement, and phasing were performed with AutoSHARP (Bricogne et al., 2003; de la Fortelle and Bricogne, 1997) and automatic model building was performed with ARP/wARP 7.0.1 (Perrakis et al., 1999). PHASER (Storoni et al., 2004) was subsequently used to perform molecular replacement using the model generated by the MAD experiment in order to phase data collected to 1.24 Å, and the model was automatically completed and refined using Lafire 2.6 (Yao et al., 2006). Cycles of manual model building were performed with COOT 0.4.1 (Emsley and Cowtan, 2004), whereas cycles of restrained refinement employing TLS motion determination (Painter and Merritt, 2006) were performed with REFMAC 5.4 (Murshudov et al., 1997) as implemented in the CCP4 program suite. Data collection and refinement statistics are shown in Tables 1 and 2. Residues 20–213 of SrtC-1 are included in the final model. There is an rmsd value of 0.211 Å between the two monomers that form the asymmetric unit dimer.

SrtC-3 crystals were obtained in 100 mM MES (pH 5.0), 0.5 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 1.0 M LiSO<sub>4</sub>, and cryocooled directly on the nitrogen stream. A native data set was collected to 2.14 Å using an in-house Rigaku rotating anode generator equipped with osmic multilayer mirrors and a R-Axis IV<sup>2+</sup> detector. The structure was solved by molecular replacement using the program PHASER (Storoni et al., 2004) and employing a search model consisting of the core of SrtC-1 (residues 82–209) in which all amino acids that differed between SrtC-1/SrtC-3 were replaced by alanines. Manual model building and refinement were performed as for SrtC-1, as described above. For both structures, stereochemical verification was performed by PROCHECK (Laskowski et al., 1993) and secondary structure assignment was performed by DSSP (Kabsch and Sander, 1983) and STRIDE (Heinig and Frishman, 2004). The final SrtC-3 model includes residues 35–242, and there is an rmsd of 0.045 Å between the two monomers that generate the dimer seen in the asymmetric unit. Data collection and structure refinement statistics are shown in Tables 1 and 2. The SrtC-2 model was generated through homology modeling based on the structure of SrtC-1 followed by refinement, using the Modeler 9.3 program (Sali and Blundell, 1993) as implemented in CCP4. Figures were generated with PyMol (<http://www.pymol.org>).

#### ACCESSION NUMBERS

The coordinates of SrtC-1 and SrtC-3 have been deposited in the Protein Data Bank under the ID codes 2w1j and 2w1k, respectively.

#### SUPPLEMENTAL DATA

Supplemental Data include four figures, one table, Supplemental Methods, and Supplemental References and can be found with this article online at [http://www.cell.com/structure/supplemental/S0969-2126\(08\)00385-7](http://www.cell.com/structure/supplemental/S0969-2126(08)00385-7).

## ACKNOWLEDGMENTS

The authors thank Xavier Vernede (LCCP, IBS) and the ESRF ID14-EH3 beam-line staff for help with data collection, the IBS mass spectroscopy facility (LSMP, IBS) for analyses, Alexandra Kraut (Lab. d'Etude de la Dynamique des Protéomes, iRTSV Grenoble) for LCMS/MS experiments, Jean-Pierre Andrieu (LBM, IBS) for N-terminal sequencing, Claire Durmort (LIM, IBS) for help with RNA analysis, Véronique Frachet and the RioBioMol facility (LIM, IBS) for gene cloning, and J. Marquez and the HTX Lab team (Partnership for Structural Biology) for access to and help with high-throughput crystallization. This work was partly supported by Young Investigator (ANR Jeunes Chercheurs) grants to A.M.D.G. (n° 05-JCJC-0049) and G.S. (n° 06-JCJC-0126), as well as European Commission grant LSHM-CT-2004-512138 (to A.D. and T.V.).

Received: July 11, 2008

Revised: October 3, 2008

Accepted: October 9, 2008

Published: December 9, 2008

## REFERENCES

- Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., et al. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. USA* **103**, 2857–2862.
- Bentley, M.L., Gaweska, H., Klele, J.M., and McCafferty, D.G. (2007). Engineering the substrate specificity of *Staphylococcus aureus* sortase A. *J. Biol. Chem.* **282**, 6571–6581.
- Bogaert, D., Hermans, P.W., Adrian, P.V., Rumke, H.C., and De Groot, R. (2004). Pneumococcal vaccines: an update on current strategies. *Vaccine* **22**, 2209–2220.
- Bricker, A.L., and Camilli, A. (1999). Transformation of a type 4 encapsulated strain of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* **172**, 131–135.
- Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030.
- Bronzwaer, S., Cars, O., Buchholz, U., Molstad, S., Goettsch, W., Veldhuijzen, I.K., Kool, J.L., Sprenger, M.J.W., and Degener, J.E. (2002). A European study on the relationship between antimicrobial use and antimicrobial resistance. *Emerg. Infect. Dis.* **8**, 278–282.
- Budzik, J.M., Marraffini, L.A., Souda, P., Whitelegge, J.P., Faull, K.F., and Schneewind, O. (2008). Amide bonds assemble pili on the surface of bacilli. *Proc. Natl. Acad. Sci. USA* **105**, 10215–10220.
- Burrows, L.L. (2005). Weapons of mass retraction. *Mol. Microbiol.* **57**, 878–888.
- CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763.
- de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- Frankel, B.A., Tong, Y., Bentley, M.L., Fitzgerald, M.C., and McCafferty, D.G. (2007). Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* **46**, 7269–7278.
- Gianfaldoni, C., Censini, S., Hillerlingmann, M., Moschioni, M., Facciotti, C., Pansegrau, W., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ruggiero, P. (2007). *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. *Infect. Immun.* **75**, 1059–1062.
- Hava, D.L., and Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol. Microbiol.* **45**, 1389–1405.
- Heinig, M., and Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502.
- Hilleringmann, M., Giusti, F., Baudner, B.C., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ferlenghi, I. (2008). Pneumococcal pili are composed of protofilaments exposing adhesive clusters of RrgA. *PLoS Pathog.* **4**, e1000026.
- Ilangovan, U., Ton-That, H., Iwajara, J., Schneewind, O., and Clubb, R.T. (2001). Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. USA* **98**, 6056–6061.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
- Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**, 1625–1628.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereo chemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
- LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006). RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet. *Infect. Immun.* **74**, 2453–2456.
- LeMieux, J., Woody, S., and Camilli, A. (2008). The roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. *J. Bacteriol.* **190**, 6002–6013.
- Levine, O.S., O'Brien, K.L., Knoll, M., Adegbola, R.A., Black, S., Cherian, T., Dagan, R., Goldblatt, D., Grange, A., Greenwood, B.M., et al. (2006). Pneumococcal vaccination in developing countries. *Lancet* **367**, 1880–1882.
- Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E.T., Rosini, R., et al. (2005). Identification of a universal group B *Streptococcus* vaccine by multiple genome screen. *Science* **309**, 148–150.
- Mandlik, A., Das, A., and Ton-That, H. (2008). The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc. Natl. Acad. Sci. USA* **105**, 14147–14152.
- Manetti, A.G.O., Zingaretti, C., Falugi, F., Capo, S., Bombaci, M., Bagnoli, F., Gambellini, G., Bensi, G., Mora, M., Edwards, E.A., et al. (2007). *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Mol. Microbiol.* **64**, 968–983.
- Marraffini, L.A., and Schneewind, O. (2006). Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol. Microbiol.* **62**, 1402–1417.
- Marraffini, L.A., DeDent, A.C., and Schneewind, O. (2006). Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192–221.
- Mattick, J.S. (2002). Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **56**, 289–314.
- Mazmanian, S.K., Skaar, E.P., Gaspar, A.H., Humayun, M., Gornicki, P., Jelenska, J., Joachimiak, A., Missiakas, D.M., and Schneewind, O. (2003). Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science* **299**, 906–909.
- Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G.O., 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. USA* **102**, 15641–15646.
- Murshudov, G., Vagin, A., and Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255.
- Naik, M.T., Suree, N., Ilangovan, U., Liew, C.K., Thieu, W., Campbell, D.O., Clemens, J.J., Jung, M.E., and Clubb, R.T. (2006). *Staphylococcus aureus* sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* **281**, 1817–1826.
- Nallapareddy, S.R., Singh, K.V., Sillanpaa, J., Garsin, D.A., Hook, M., Erlandsen, S.L., and Murray, B.E. (2006). Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. *J. Clin. Invest.* **116**, 2799–2807.

- Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450.
- Perrakis, A., Morris, R.M., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
- Rosini, R., Rinaudo, C.D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzi, C., Brettoni, C., et al. (2006). Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. *Mol. Microbiol.* **61**, 126–141.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
- Sauer, F.G., Pinkner, J.S., Waksman, G., and Hultgren, S.J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543–551.
- Schneewind, O., Model, P., and Fischetti, V.A. (1992). Sorting of protein A to the staphylococcal cell wall. *Cell* **70**, 267–281.
- Scott, J.R., and Zahner, D. (2006). Pili with strong attachments: Gram-positive bacteria do it differently. *Mol. Microbiol.* **62**, 320–330.
- Spratt, B.G., and Greenwood, B.M. (2000). Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet* **356**, 1210–1211.
- Storoni, L., McCoy, A., and Read, R. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1373–1382.
- Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., and Grandi, G. (2006). Pili in Gram-positive pathogens. *Nat. Rev. Microbiol.* **4**, 509–519.
- Ton-That, H., and Schneewind, O. (2003). Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **50**, 1429–1438.
- Ton-That, H., Mazmanian, S.K., Alksne, L., and Schneewind, O. (2002). Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. *J. Biol. Chem.* **277**, 7447–7452.
- Ton-That, H., Marraffini, L.A., and Schneewind, O. (2004). Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **53**, 251–261.
- Vetsch, M., Puorger, C., Spirig, T., Grauschopf, U., Weber-Ban, E.U., and Glockshuber, R. (2004). Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, 329–332.
- Walsh, C. (2003). Where will new antibiotics come from? *Nat. Rev. Microbiol.* **1**, 65–70.
- Weerkamp, A.H., Handley, P., Baars, A., and Slot, J.W. (1986). Negative staining and immunoelectron microscopy of adhesion-deficient mutants of *Streptococcus salivarius* reveal that the adhesive protein antigens are separate classes of cell surface fibril. *J. Bacteriol.* **165**, 746–755.
- Yao, M., Zhou, Y., and Tanaka, I. (2006). LAFIRE: software for automating the refinement process of protein structure analysis. *Acta Crystallogr. Sect. D* **62**, 189–196.
- Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S.K., Missiakas, D.M., Gornicki, P., Schneewind, O., and Joachimiak, A. (2004). Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* **12**, 1147–1156.
- Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O., and Narayana, S.V.L. (2004a). Crystal structure of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem.* **279**, 31383–31389.
- Zong, Y., Mazmanian, S.K., Schneewind, O., and Narayana, S.V.L. (2004b). The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* **12**, 105–112.



# Structural Basis of Host Cell Recognition by the Pilus Adhesin from *Streptococcus pneumoniae*

Thierry Izoré,<sup>1</sup> Carlos Contreras-Martel,<sup>1</sup> Lamy El Mortaji,<sup>1</sup> Clothilde Manzano,<sup>1</sup> Rémy Terrasse,<sup>1</sup> Thierry Vernet,<sup>1</sup> Anne Marie Di Guilmi,<sup>1,\*</sup> and Andréa Dessen<sup>1,\*</sup>

<sup>1</sup>Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 (CEA, CNRS, UJF); 41 rue Jules Horowitz, F-38027 Grenoble, France

\*Correspondence: diguilmi@ibs.fr (A.M.D.G.), dessen@ibs.fr (A.D.)

DOI 10.1016/j.str.2009.10.019

## SUMMARY

Pili are fibrous virulence factors associated directly to the bacterial surface that play critical roles in adhesion and recognition of host cell receptors. The human pathogen *Streptococcus pneumoniae* carries a single pilus-related adhesin (RrgA) that is key for infection establishment and provides protection from bacterial challenge in animal infection models, but details of these roles remain unclear. Here we report the high-resolution crystal structure of RrgA, a 893-residue elongated macromolecule whose fold contains four domains presenting both eukaryotic and prokaryotic origins. RrgA harbors an integrin I collagen-recognition domain decorated with two inserted “arms” that fold into a positively charged cradle, as well as three “stalk-forming” domains. We show by site-specific mutagenesis, mass spectrometry, and thermal shift assays that intradomain isopeptide bonds play key roles in stabilizing RrgA’s stalk. The high sequence similarity between RrgA and its homologs in other Gram-positive microorganisms suggests common strategies for ECM recognition and immune evasion.

## INTRODUCTION

Pili are elongated, flexible appendages directly associated to the bacterial surface that have been shown to participate in essential events for infection establishment, including adhesion to the extracellular matrix, DNA transfer, and biofilm formation. The well-studied Gram-negative pili consist of thin fibers formed by non-covalently associated building blocks, and binding to host cells is performed by adhesins, which are often localized on the fimbrial tip. Adhesin recognition of oligosaccharides of target cell glycoproteins or glycolipid receptors determines tissue tropism and is often essential for the initial phases of the infection process (Proft and Baker, 2009; Soto and Hultgren, 1999; Telford et al., 2006). Pili in Gram-positive pathogens have only been recently described, and consist of a backbone formed by covalently linked copies of a major pilin to which minor pilins are associated. Although the minor subunits are not required for formation of the pilus shaft, they have been shown to be essential for host cell recognition and adhesion of a variety of Gram-positive microorganisms (Abbot et al., 2007; Barocchi et al., 2006;

Dramsai et al., 2006; Maisey et al., 2007; Mandlik et al., 2007; Nal-lapareddy et al., 2006; Nelson et al., 2007). In addition, pilus-related adhesins of Gram-positive bacteria confer protection against a potentially lethal bacterial challenge in animal disease models (Gianfaldoni et al., 2007; Maione et al., 2005; Mora et al., 2005), underlining the importance of these molecules in pathogenesis and as well as of their potential application as antigens in vaccination strategies (Proft and Baker, 2009; Telford et al., 2006).

The Gram-positive human pathogen *Streptococcus pneumoniae* accounts for a large number of cases of pneumonia, otitis, and meningitis, and is one of the major causes of community-acquired illnesses. Despite the availability of antibiotic treatment and vaccine strategies, it still claims yearly over 1.5 million victims worldwide, mostly due to the spread of antibiotic-resistant strains and the limited efficacy of vaccines (Bronzwaer et al., 2002; Levine et al., 2006). Thirty percent of *S. pneumoniae* strains isolated in clinical environments are pilated and carry the *rfa* genetic element, which codes for all seven genes required for pilus formation: the *rfa* regulating element, the backbone-forming pilin (RrgB), an adhesin (RrgA), a minor pilin (RrgC), and three sortases (SrtC-1, SrtC-2, SrtC-3) (Barocchi et al., 2006; Dramsai et al., 2008; LeMieux et al., 2006). Sortases are cysteine transpeptidases that recognize cell-wall sorting signals (CWSS) such as LPXTG-like motifs on target pilins and catalyze both the covalent association of pilus backbone subunits and the association of minor pilins to the fibrillar shaft (Mandlik et al., 2008a; Scott and Zahner, 2006; Ton-That and Schneewind, 2004). Although sortases have been shown to recognize preferentially one Rrg substrate, it is now known that their specificities can be interchangeable up to a certain extent, a fact that is potentially linked to the similarity of their active sites (Fälker et al., 2008; Manzano et al., 2008; Neiers et al., 2009; Manzano et al., 2009). Immunoelectron microscopy imaging has shown that the pneumococcal backbone fiber is formed by RrgB, whereas RrgA and RrgC are associated along the filament (Barocchi et al., 2006; Hilleringmann et al., 2008; LeMieux et al., 2006).

The only adhesin of the pneumococcal pilus, RrgA, plays a key role in bacterial adhesion both to respiratory and to epithelial cells (Nelson et al., 2007); in addition, it has been shown to interact with different elements of the extracellular matrix (ECM), namely fibronectin, collagen, and laminin (Hilleringmann et al., 2008). Interestingly, however, cellular recognition by pneumococcal pili also occurs in  $\Delta rrgB\Delta rrgC$  strains, in which the RrgB backbone is notably absent, suggesting that the adhesive properties of RrgA are effective even when it is associated directly onto the bacterial surface (Nelson et al., 2007). However,



**Table 1. Data Collection, Phasing, Molecular Replacement, and Structure Refinement Statistics**

Data Collection				
Data set	Peak	Inflection Point	Low Remote	Native
Wavelength (Å)	0.9791	0.9793	0.9751	0.9810
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
a (Å)	51.66	51.78	51.85	51.09
b (Å)	83.66	83.98	84.13	83.73
c (Å)	180.27	180.91	181.24	299.43
β (°)	89.79	89.82	89.83	
Resolution (Å)	2.41	2.41	2.41	1.90
No. observed/unique reflections	193,352/102,274	187,186/100,908	176,800/97,859	582,427/97,681
Completeness (%)	81 (60)	76.4 (50.7)	70.6 (41.8)	88.7 (64.4)
R <sub>sym</sub> (last shell)	7.9 (36.3)	8.6 (59.4)	9.9 (82.4)	5.3 (35.9)
//σ(I) (last shell)	6.97 (2.13)	6.4 (1.41)	5.75 (0.97)	38.29 (5.96)
Phasing				
Phasing power ANO	0.828	0.606	0.256	
Phasing power ISO (acentric/centric)		0.104/0.109	0.192/0.189	
FOM <sub>obs</sub>	0.3433			
FOM <sub>DM</sub>	0.8759			
Molecular Replacement				
Phaser RFZ/TFZ				22.2/31.2
LLG				968/1090
Refinement				
Resolution (Å)				1.90
R <sub>work</sub> (%)				19.80
R <sub>free</sub> (%)				23.80
No. of protein atoms				6486
No. of solvent atoms				819
No. of HEPES molecules				2
Rmsd, bond lengths (Å)				0.024
Rmsd, bond angles (°)				1.890
Mean B factor (Å <sup>2</sup> )				18.56
Residues in most favored/allowed region of Ramachandran plot (%)				98.23

neither the identity of RrgA's cellular receptor(s) nor details regarding its multifaceted mode of ECM recognition are known.

In order to address this question, we solved the high-resolution crystal structure of RrgA from the highly infectious TIGR4 pneumococcal strain. In contrast to all adhesin structures solved to date, which display compact folds of limited size, RrgA is an 893-residue, elongated molecule that harbors four domains associated semilinearly as “beads on a string.” The largest domain harbors a region that displays high similarity to the collagen-binding integrin I domain, whereas the three others display folds present in other bacterial systems, namely a region reminiscent of the CnaB adhesin of *S. aureus* and two immunoglobulin (IgG-like) domains. Both CnaB-like and IgG domains are stabilized by intermolecular isopeptide bonds, which we show, by mutagenesis and thermal shift assays, to be key for RrgA stabilization. The high sequence similarity between RrgA and other streptococcal pilus adhesins suggests that its complex, multifaceted fold, which involves elements from both

eukaryotic and prokaryotic systems, will provide information regarding pilus-ECM interactions in a variety of Gram-positive pathogens.

## RESULTS

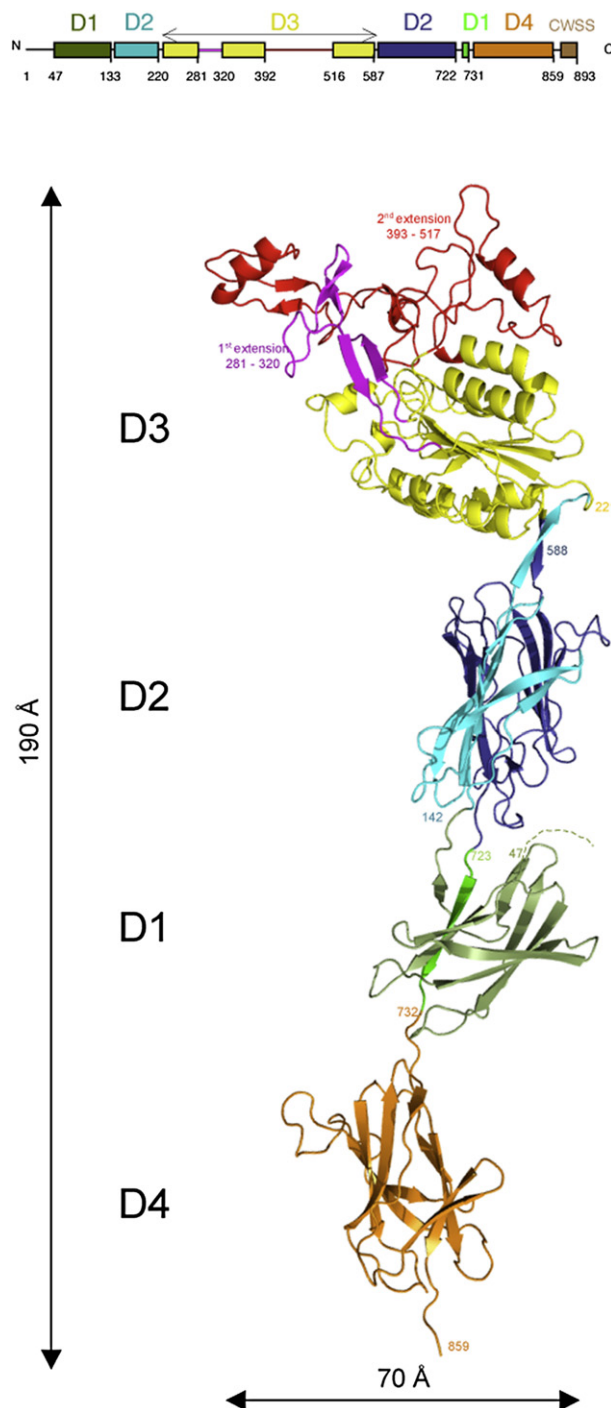
### RrgA, A Four-Domain Adhesin

RrgA is an 893-residue protein whose primary structure analysis predicts the presence of a von Willebrand factor-like domain and reveals an LPXTG-like CWSS at its C terminus; a construct that lacks the first 38 amino acids, predicted to be disordered, as well as the CWSS was used for crystallization. RrgA (39–868) crystallized as a monomer in space group P2<sub>1</sub>, and the structure was solved by performing a three-wavelength anomalous dispersion (MAD) experiment on a selenomethionylated crystal at the ESRF synchrotron in Grenoble; this model was subsequently used to phase native data to 1.9 Å. Data collection, phasing, and refinement statistics are included in Table 1.

RrgA (residues 47-815 Figure 1; see Figure S1 available online) is an unusual four-domain (D1-D4), elongated molecule that is approximately 195 Å long and 70 Å wide; domains D1 and D2 carry elements from both N-terminal and C-terminal regions of RrgA into which other sequences have been inserted (domain D3 is inserted into D2; D2/D3 are inserted into D1). The four domains make few contacts with each other, being mostly associated through short linkers and hairpins, which suggests that the full-length protein, once associated to the RrgB backbone fiber, could display a considerable level of domain flexibility.

Domain D1 is highly reminiscent of an IgG domain (Figure S2). Five  $\beta$  strands are formed from the N terminus of the sequence (residues 47-141), whereas the last strand, which connects to D4, is formed by C-terminal residues 723-731 (dark green in Figure 1B). Domain D2 is formed by N-terminal residues 144-218 and C-terminal residues 593-722 that clasp into an 11- $\beta$ -stranded sandwich (cyan and blue, Figure 1B). The two "halves" of the sandwich are locked together by an isopeptide bond between Lys191 and Asn695. Notably, the fold of D2 is reminiscent of that of the B domain of the collagen-binding adhesin Cna from *S. aureus* (Deivanayagam et al., 2000) (Figure S2); a comparison of the two structures reveals that it is mostly the core  $\beta$  strands that are structurally comparable, and RrgA harbors additional hairpins and loops located away from the core. The D2 domain is linked to a third, major domain (D3) through a short  $\beta$ -hairpin structure, itself also composed of residues from the N-terminal and C-terminal regions of the molecule (Thr214-Gln218, and Val587-Lys591).

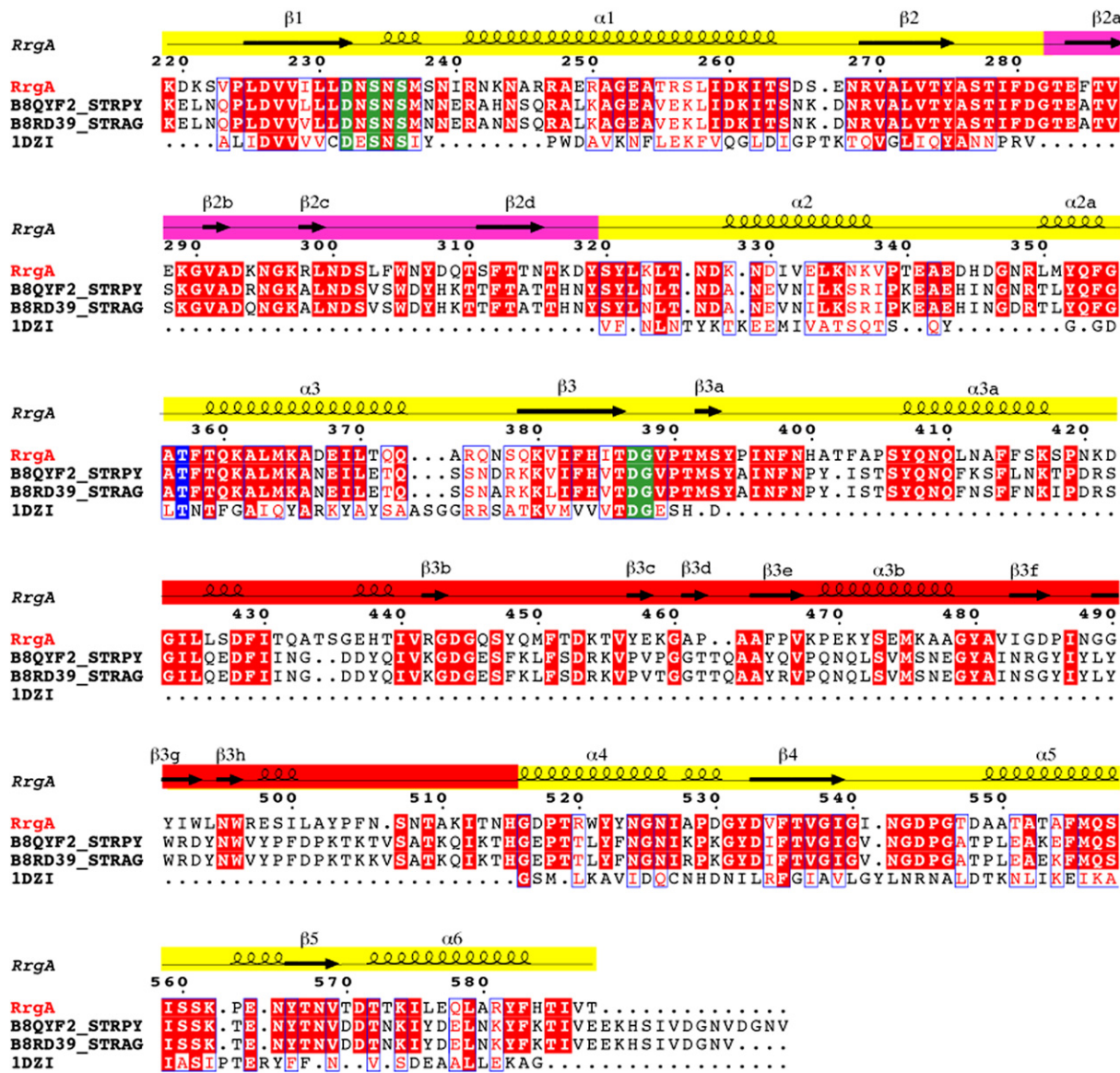
RrgA's D3 domain (Lys219-Ile586) contains a region (yellow in Figure 1) whose fold highly resembles that of the human A3 domain of von Willebrand factor (VWA), a molecule shown to interact with collagens I and III (Bienkowska et al., 1997; Cruz et al., 1995). The VWA fold is found in many eukaryotic cell surface proteins involved in interactions with the extracellular matrix (Springer, 2006), including within the  $\beta$  domain of integrins, where it also has been shown to play a role in collagen interaction (Emsley et al., 2000; Kamata and Takada, 1994; Tuckwell et al., 1995). The classical VWA/integrin I fold consists of a central, six-stranded  $\beta$  sheet surrounded by  $\alpha$  helices, resembling a dinucleotide-binding domain (Bienkowska et al., 1997; Emsley et al., 2000; Huizinga et al., 1997; Lee et al., 1995b; Qu and Leahy, 1995; Vorup-Jensen et al., 2003) and superposition of the structure of the  $\alpha 2\beta 1$  integrin onto the analogous domain in RrgA reveals high structural similarity (Figure S2) (albeit a sequence identity level that is below 10%). However, RrgA's integrin I-like domain displays a five-stranded parallel sheet and carries two main inserted "arms" (Figure S3). The first inserted region immediately follows  $\beta 2$  and involves residues Phe280-Asp318 (magenta in Figure 1); it folds into an elongated region carrying two  $\beta$  hairpins. The second, larger insertion region ranges from Met392 to Gly516 (red in Figure 1); although it is mostly composed of loops, it carries one short hairpin region and two helices. These two inserted arms extrude from the core of the integrin I-like domain and stretch out, away from the body of the molecule, making a large number of interactions throughout. Interestingly, sequence analyses and structural predictions of RrgA homologs in different pathogenic streptococcal species (Figure 2) reveal that the full-length molecules display a sequence identity level of approximately 50% (with



**Figure 1. Schematic Arrangement and Tertiary Fold of RrgA**

(Top) RrgA is composed of four domains, where D3 is inserted into D2, and D2/D3 are inserted into D1.

(Bottom) In 3D, RrgA's four domains are aligned much like beads on a string; D1 is shown in dark green (N-terminal region) and light green (C-terminal strand); the Cna-B-like domain of D2 is formed by residues from the N terminus (light blue) and C terminus (dark blue) of the molecule. The integrin I-like domain is shown in yellow, with the extended arms in red and magenta. The C terminus of RrgA is composed of domain D4 (orange).



**Figure 2. Streptococcal Pilus Adhesins Show Multiple Similar Structural Features**

Sequence alignment of the D3 domain of RrgA with the analogous domain of the major pilus adhesins (RrgA homologs) from *Streptococcus agalactiae* (B8RD39) and *Streptococcus pyogenes* (B8QYF2), as well as the I domain of eukaryotic integrin  $\alpha 2\beta 1$ . Identical residues are shown in white with a red background. Residues involved in the MIDAS motif are highlighted in green. Thr357, which could potentially bind the metal ion directly upon a ligand-induced conformational change, is highlighted in blue. Secondary structure elements corresponding to inserted arms are shown in magenta and red; note their clear absence in the sequence of integrin  $\alpha 2\beta 1$ . The red arm region shows the highest level of sequence divergence.

similarity levels beyond 70%), and also harbor integrin I-like domains decorated with arms. However, it is precisely in the region of the inserted arms that the sequence identity level is the lowest (in the range of 40%); this is the most evident in the analysis of Figure 2, and suggests that this region could be employed for different roles in pathogenesis (see below).

Domain D4 (orange in Figure 1) precedes the CWSS C-terminal sequence and is separated from domain D3 through a 13-residue stretched linker region with high temperature factors (and thus flexibility). D4, like D1, is highly reminiscent of an IgG domain (Figure S2), and is composed uniquely of

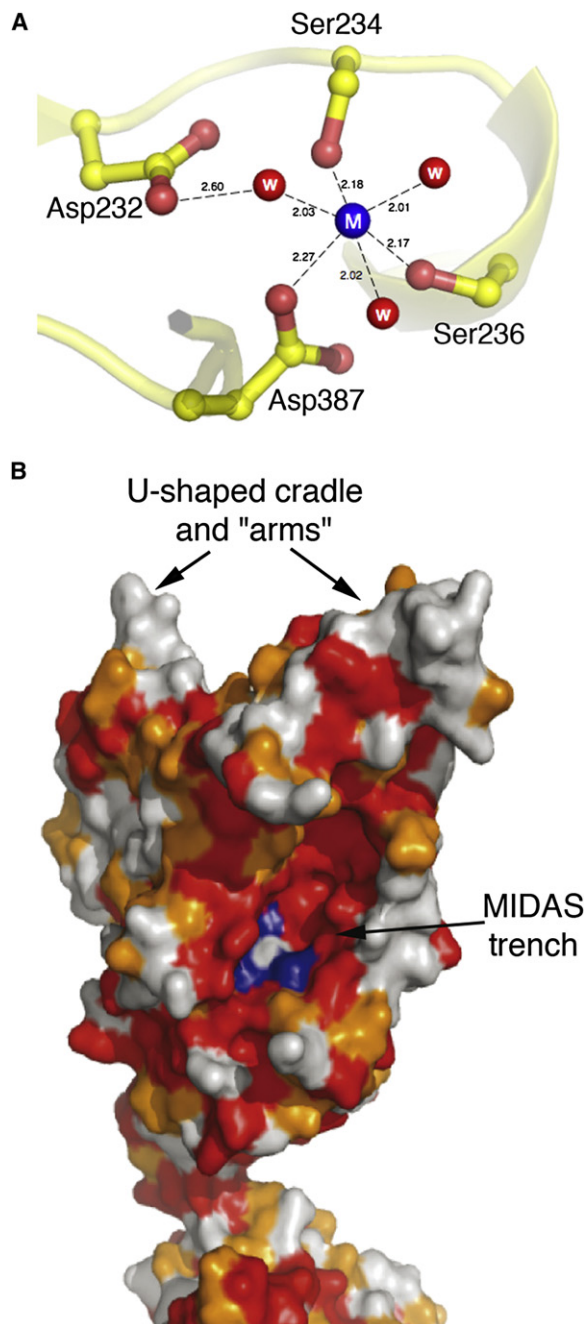
C-terminal residues 736-859. IgG domains have been identified within the structure of the minor adhesin GBS52 of the pilus of *S. agalactiae*, where they have been shown to be crucial for bacterial adhesion (Krishnan et al., 2007), and is notably present in structures of adhesins associated to Gram-negative pili (Soto and Hultgren, 1999). In addition, a second isopeptide bond is present between Lys742 and Asn854 (see below).

**Multiple Extracellular Matrix-Binding Platforms**

Integrin I domains carry a MIDAS (metal ion-dependent adhesion site) motif located largely on the C-termini of the central  $\beta$  sheet.

This site is formed by an aspartate and 2 serine residues organized within a Asp-X-Ser-X-Ser sequence, as well as a Thr and a second Asp residue, and has been shown to play a key role in the recognition of collagen (Emsley et al., 2000; Kamata and Takada, 1994). The analogous region in RrgA, located within domain D3, also carries a MIDAS motif composed of the Asp232-X-Ser234-X-Ser236 constellation of residues on loop1 and one acidic residue (Asp 387) on loop 3 (Figures 2 and 3A). In addition, as is the case for integrins, the MIDAS motif of RrgA coordinates a metal ion that we have modeled here as an  $Mg^{+2}$  (although no metal was included in the crystallization solution). Ser234, Ser236, and Asp387 form direct bonds to the metal through their hydroxyl oxygen atoms, whereas Asp232 and Thr357 contribute water-mediated contacts. It is of note that integrin I domains undergo large conformational changes upon ligand binding (Springer, 2006), and the conformation of the RrgA MIDAS domain, presented here in the absence of ligand, is closed. It is of interest, however, that the high structural similarity between the D3 domain of RrgA and beta-I integrin domains, including the presence of metal within the MIDAS motif, clearly points to collagen recognition as a potential function for RrgA; interestingly, this is confirmed by the studies of Hilleringmann and colleagues (Hilleringmann et al., 2008) who identified collagen as a major target for RrgA. One notable difference between the MIDAS motif of eukaryotic integrins and pneumococcal RrgA is that in the former, the motif is located superficially, on a shallow region of the protein surface, whereas in RrgA it is positioned within an a trench-like region cradled by the two inserted arms (Figure S4). Superposition of the structure of integrin  $\alpha 2\beta 1$  in complex with a triple-stranded peptide from collagen III (PDB 1dzt) onto RrgA generated a large number of clashes (not shown), suggesting that accommodation of the collagen peptide into the trench would require localized movement of MIDAS residues, as shown for "open" conformations of many integrin I domains (Springer, 2006). Superposition of the structure of the RrgA MIDAS domain with the open and "closed" forms of the CR3  $\beta 2$ -integrin (Lee et al., 1995a) suggests that cleft opening could include movement of the Thr357 side chain, which could potentially interact directly with the metal ion in the liganded form (Figure S5). In addition, conformational changes could include movement or rotation of the arms. Such an entrenched recognition between a pilus adhesin and an ECM component could provide accrued stability to the bacterium on the matrix surface.

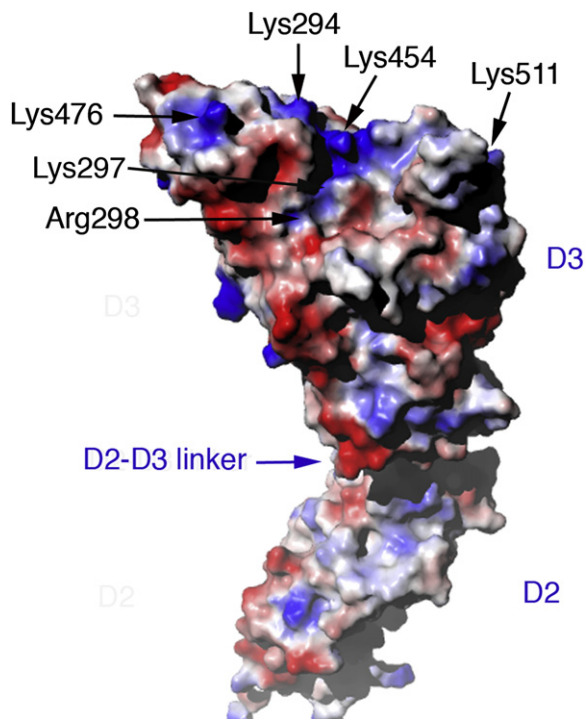
In addition to recognizing collagen, RrgA has also been shown to be important for *S. pneumoniae*'s recognition of laminin and fibronectin, common components of the eukaryotic extracellular matrix (ECM) (Hilleringmann et al., 2008). Glycosaminoglycans (GAGs) are linear polysaccharides attached covalently to proteins of the proteoglycan family, and are ubiquitously present on all animal cell surfaces. Interestingly, several proteins that recognize GAGs do so through elongated crevices rich in basic residues (Imberty et al., 2007). Notably, analysis of a surface potential map of RrgA reveals that the two insertion arms of D3 join on the region that is the farthest from the N and C termini of RrgA in order to form a large, cradle-shaped surface of highly basic character (Figure 4). The concentration of basic residues present in the U-shaped cradle suggests that the cradle could provide the binding site for negatively charged molecules (such



**Figure 3. The MIDAS Motif of RrgA**

(A) The metal ion, refined here as a  $Mg^{2+}$ , is stabilized by a classical constellation of residues also seen in MIDAS motifs of eukaryotic proteins. (B) Mapping of identical (red) and similar (orange) residues in RrgA-like adhesins in Streptococcal spp. (*S. pneumoniae*, *S. agalactiae*, *S. pyogenes*) onto the structure of the D3 domain of RrgA. Residues involved in forming the MIDAS motif are in dark blue. The central region of the domain, which harbors the MIDAS motif, is the most conserved, while the arms of the structure display patches of nonconserved sequence.

as GAGs that are directly associated to fibronectin and laminin) and thus bridge the interaction between the bacterium and the ECM. Interestingly, only few of the basic residues identified in the U-shaped cradle of RrgA are strictly conserved (as Lys or



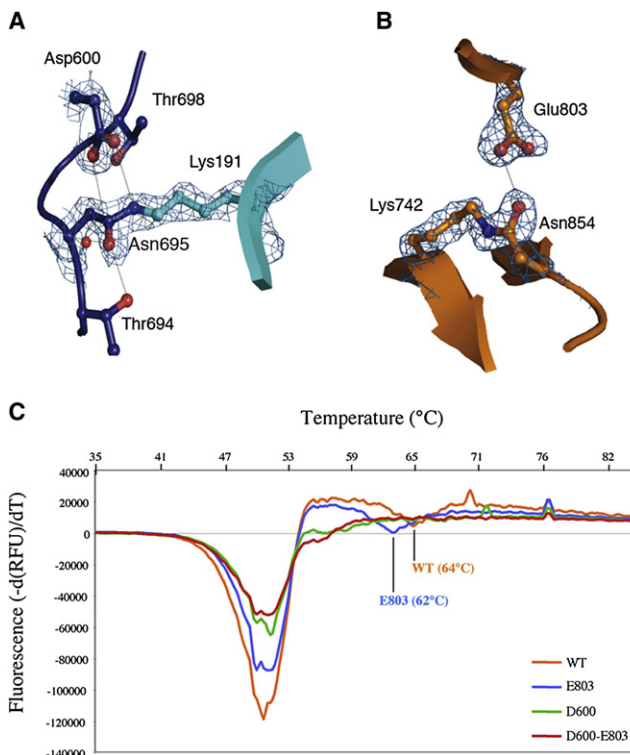
**Figure 4. Another ECM-Binding Region Is Formed by the Inserted Arms**

Surface potential diagram of the D3 domain of RrgA that shows the basic U-shaped cradle. This view is 180° away from the one shown in Figure 3B.

Arg) among the sequences of the pilus adhesins of *S. agalactiae* and *S. pyogenes* (Figures 2 and 3B), suggesting that the crevice formed by the arms in these adhesins could potentially recognize other host cell molecules. This observation could provide a partial explanation for the distinct tissue tropism observed for pilated Gram-positive species.

#### The “Stalk” Is Stabilized by Intramolecular Isopeptide Bonds

The pilus-forming biosynthetic machinery in Gram-positive organisms requires the formation of intermolecular isopeptide bonds, notably between the sortases and the LPXTG-like motif on their target proteins (Mandlik et al., 2008b; Scott and Zahner, 2006). In addition, isopeptide bonds formed within a single domain have also been identified in the structure of Sp0128, the major pilin subunit from *S. pyogenes* (Kang et al., 2007) and of a minor accessory pilus molecule from the *S. agalactiae* pilin (Krishnan et al., 2007). These bonds have been shown to be the product of an intramolecular reaction between the side chains of a Lys and an adjacent Asp residue, and the NZ and OD1 atoms of the resulting bond are both stabilized by hydrogen bonding with an acidic residue in the immediate vicinity (Kang et al., 2007). RrgA harbors two intramolecular isopeptide bonds, one located in the Cna-B-like D2 domain, and a second one in D4. The bond within D2 is formed between the side chain of Lys191 (in the N-terminal region of the RrgA sequence) and Asn695 (in the C terminus of the structure). Although an acidic group (Asp600) is located in its vicinity, the side chains of Thr694 and Thr698 also provide additional stabilization for the



**Figure 5. Domains D2 and D4 Are Stabilized by Intramolecular Isopeptide Bonds**

(A) Domain D2 carries a bond between the side chains of Lys191 and Asn695 that, in addition to Asp600, is also stabilized by two Thr hydroxyl groups.

(B) The isopeptide bond in domain D4 is stabilized by nearby Glu803. In both (A) and (B), Fo-Fc simulated annealing omit maps are shown.

(C) Thermal shift analysis results for RrgA isopeptide bond mutants. Wild-type RrgA (orange) displays maxima at 51°C and 64°C, while mutants involving the D2 domain isopeptide bond have lost the 64°C peak, indicating that the RrgA stalk could be compromised.

isopeptide linkage (Figure 5A). The isopeptide bond present in D4 is formed by proximal elements of the C-terminal sequence, namely Lys742 and Asn854; the hydrogen bonding stabilization profile is more classical, being provided by Glu803 (Figure 5B). As is the case for other isopeptide linkages, the intramolecular isopeptide bonds in RrgA are also surrounded by hydrophobic residues: Ile206, Leu208, and Phe713 for the D2 bond and Leu751, Phe756, and Ile852 for the D4 bond.

We sought to confirm the importance of the two isopeptide bonds for RrgA's stability using site-specific mutagenesis, mass spectrometry, and thermal shift analyses. The molecular mass of wild-type RrgA, measured by mass spectrometry, was consistent with the absence of two NH<sub>3</sub> groups (thus confirming the formation of two isopeptide bonds; see Table 2). We initially studied the isopeptide bond in D2, and generated single mutants in which Lys191, Asn695, and Asp600 were mutated into alanines. For each mutant, the measured mass was compatible with the loss of one isopeptide bond; thus, mutants K191A and N695A were no longer capable of generating the isopeptide bond in D2 (uniquely in D4). Interestingly, an RrgA variant carrying a D600A mutation behaved similarly to the K191A and N695A mutants, also displaying the inability to generate one

**Table 2. Mass Measurements of RrgA Variants by Electrospray Ionization Mass Spectrometry Time-of-Flight Mass**

	CalculatedMass (Da)	MeasuredMass (Da)	Mass Difference (Calculated – Measured)	Number of Isopeptide Bonds Present
WT RrgA	94 197.5	94,163.7	33.8	2
E803A	94 139.5	94,123.8	15.7	1
K191A	94 140.4	94,124.8	15.6	1
N695A	94 154.5	94,139.3	15.2	1
D600A	94 153.5	94,139.3	14.2	1
D600A/E803A	94 095.5	94,096.0	–0.5	0

The presence of one isopeptide bond corresponds to the loss of one NH<sub>3</sub> group (17Da).

isopeptide bond even in the presence of intact Lys191 and Asn695 side chains. These results suggest that the function of the acidic side chain in isopeptide bond formation is dual: it provides hydrogen bonding stabilization to the bond once it is formed, but must also play a key role in its formation during protein folding. In light of these results and in order to study the isopeptide bond located within domain D4, we generated an RrgA variant carrying a single E803A mutation. Mass spectrometric measurements of RrgA-E803A confirmed that one isopeptide bond was also missing in this form, thus confirming that the presence of an acidic residue in the immediate environment of an isopeptide bond is an essential factor for its formation. Notably, no isopeptide bonds could be identified in an RrgA-D600A-E803A double mutant (Table 2).

We then performed thermal shift analyses of all RrgA variants. Proteins were gradually heated in an IQ5 Bio-Rad apparatus and thermal unfolding curves were monitored through the detection of changes in fluorescence of a Sypro Orange probe. This analysis (Figure 5c) reveals that wild-type RrgA displays two unfolding maxima: the first at approximately 51°C (±0.5), which is shared with all other RrgA variants, and a second, less pronounced peak, whose T<sub>m</sub> value is approximately 64°C (±0.2). Because the major peak is present in experiments with all RrgA forms independent of the presence or absence of isopeptide bonds, it could represent the unfolding of the D3 domain. In the RrgA-E803A variant, which lacks the isopeptide bond in domain D4, the T<sub>m</sub> of the second peak is 2°C lower than for the wild-type form; this observation suggests that the minor peak corresponds to the unfolding of the smaller domains. Notably, in the cases of both RrgA-D600A and RrgA-D600A-E803A, the second peak is not observed; the absence of this second transition point suggests that absence of the isopeptide bond in domain D2 (Cna-B-like) engenders a destabilization of the entire RrgA stalk. Recently, it has been shown that the isopeptide bonds carried by *S. pyogenes* Sp0128 also play a key role in stabilization of the fiber-forming molecule (Kang and Baker, 2009) whereas isopeptide bonds in the analogous molecule of *B. cereus*, BcpA, are crucial for pilus assembly (Budzik et al., 2008). These results confirm that such bonds play a universal stabilizing role in the pilus-formation processes of Gram-positive species.

## DISCUSSION

Adhesion to target host cells by surface-exposed virulence factors, such as pili, is a key step in infection establishment by

a large number of bacterial pathogens. The well-studied pili from Gram-negative pathogens have been shown to display specific adhesins on their tips that play important roles in adherence to both the extracellular matrix and to glycoprotein or glycolipid receptors, often determining tissue tropism (Proft and Baker, 2009; Soto and Hultgren, 1999). In Gram-positive pathogens, however, clarification of the details of the pilus-mediated adhesion process has only recently started. In this work, we report the structural characterization of a totally novel adhesin that is essential for the infectivity process of a major human pathogen and whose tertiary fold is composed of linearly aligned domains observed in both eukaryotic and prokaryotic organisms.

RrgA is composed of four independent domains linked by flexible regions. This organization can potentially allow for interdomain flexibility, which could be important for the recognition of different host targets during the infection process, while the presence of two isopeptide bonds guarantees stability of the individual domains. The composite nature of RrgA includes two IgG domains (D1/D4), and their potential function could include enhancement of binding of the adhesin to host tissue receptors and/or providing stability for the macromolecule. These functions are observed in a number of bacterial adhesins; PapG, for example, is located at the distal end of the P pilus of uropathogenic *E. coli* (UPEC) and folds into two IgG-like domains, one of which is key for receptor recognition (Dodson et al., 2001). The GBS52 pilus adhesin of *S. agalactiae* also harbors two IgG-like regions, one of which is key in promoting adherence to lung tissue, while the second one provides structural support (Krishnan et al., 2007). In RrgA, the latter function could be shared by its Cna-B-like domain (D2), because the B domain of collagen-binding protein Cna of *S. aureus* has also been suggested as being the stalk that projects the A collagen-binding domain away from the bacterial surface and toward the ECM (Deivanayagam et al., 2000). In RrgA, domains D1, D2, and D4 are associated linearly, with D3 located at the edge (Figures 1; Figure S1). This arrangement suggests that one of the roles of domains D1, D2, and D4 could be to project the ECM-recognition D3 domain away from the pilus fiber in order to reach out toward the host cell as much as possible. Association of RrgA to the RrgB pilus backbone is catalyzed through recognition of dedicated sortases of the LPXTG-like CWSS motif on the C terminus of D4, lending support to this hypothesis.

The D3 domain of RrgA harbors two different potential ECM-recognition platforms: the integrin I domain, shown to participate in collagen binding in many eukaryotic proteins, and the basic

U-shaped cradle, located at the very edge of the adhesin, which could potentially be the binding site for negatively charged glucosaminoglycans or other acidic molecules of the ECM. The high sequence similarity between RrgA and pilus adhesins from *S. pyogenes* and *S. agalactiae* points to the possibility that the three-dimensional fold of these proteins could be highly similar (notably with the strict conservation of the MIDAS motif in all three molecules, Figures 2 and 3B); the identification of collagen-binding capabilities by a *S. pyogenes* strain (Kreikemeyer et al., 2005) provides support for this proposition. In addition, the presence of the integrin I domain, which is widespread in a number of eukaryotic proteins, in pilus-associated adhesins of all three pathogenic species suggests that this could be a strategy that allows the pathogen to evade host defense mechanisms. Albeit the structural similarity of the core regions, there are notable charge differences within the arms (notably within residues forming the cradle) that could be an indication of the differential interaction of such adhesins with the host, which is to be expected from molecules that target different tissues.

RrgA and its counterparts in *S. pyogenes* and *S. agalactiae* have been shown to provide protection in animal models of acute disease and could thus represent interesting potential antigens for vaccine development (Abbot et al., 2007; Nelson et al., 2007; Dramsi et al., 2006; Maisey et al., 2007). Notably, strategies for the generation of novel vaccines include the rational engineering of antigenic targets through the understanding of domain boundaries and epitope structures (Dormitzer et al., 2008). The identification of individual domains with distinct host cell recognition functions in the structure of the pilus adhesin reported here is an essential step in this process, and should be exploited in the development of novel anti-streptococcal vaccines. In addition, the predicted structural similarity between the adhesins of these three major pilated microorganisms suggests that the RrgA structure could provide a basis for the understanding of ECM-recognition features by different Gram-positive pathogens.

## EXPERIMENTAL PROCEDURES

### Expression and Purification of Recombinant Proteins

The region of the *S. pneumoniae* TIGR4 *rrgA* gene (SP0462) coding for amino acids 39 to 868 was amplified using conventional polymerase chain reaction (PCR) methodologies and cloned into vector pLIM (Protein Expert, Grenoble). The resulting vector was transformed into *E. coli* RIL cells (Invitrogen) to produce recombinant protein with an N-terminal hexahistidine tag. Protein expression was induced in Terrific Broth with 1 mM IPTG at 37°C for 3 hr. Bacteria were collected by centrifugation, resuspended in a minimal volume of buffer (50 mM Tris-HCl, 150 mM NaCl [pH 8.0], 20 mM Imidazole, 1 mM PMSF, 0.1 μM aprotinin, and 1 μM pepstatin) and lysed by sonication. The lysate was clarified by centrifugation and applied onto an immobilized metal chelate affinity chromatography column (HisTrapHP; GE HealthCare) pre-equilibrated in lysis buffer. Protein elution was performed by an imidazole step-gradient. Fractions containing RrgA were pooled and dialyzed 3 hr into 50 mM Tris-HCl, 150 mM NaCl (pH 8.0) at 4°C, and subsequently incubated with 1/5 (w:w) TEV protease at 15°C. The protein was reloaded onto an HisTrapHP column and the eluted cleaved products dialyzed 3 hr into the same buffer as above. Subsequently, RrgA was loaded onto a MonoQ (5/50GL GE HealthCare) column in 50 mM Tris-HCl (pH 8.0) and eluted in the flow through; fractions were pooled, concentrated, and injected onto a HiLoad 16/60 Superdex 200 (GE HealthCare) column in 50 mM Tris-HCl, 150 mM NaCl (pH 8.0). Pooled and concentrated fractions were used for crystallization tests. All RrgA variants were expressed using the same methodology, except that the MonoQ step was not employed.

Selenomethionylated RrgA was expressed in *E. coli* RIL cells in minimal medium supplemented with thiamine (0.2 mg/ml), leucine (50 mg/ml), valine (50 mg/ml), isoleucine (50 mg/ml), lysine (100 mg/ml), phenylalanine (100 mg/ml), threonine (100 mg/ml), and selenomethionine (60 mg/ml). Expression and purification was performed as for the native protein.

The pLIM-RrgA expression plasmid served as a template for the introduction of single amino or double acid substitutions by PCR using the QuikChange XL Site Directed Mutagenesis Kit (Stratagene). All the mutant constructs were subsequently sequenced; the results showed that only expected mutations were introduced during PCR. Expression and purification of all RrgA mutants were performed as for the wild-type protein.

### Crystallization, Data Collection, and Structure Solution

Crystals of native RrgA were grown by the hanging-drop vapor diffusion method at 20°C in 20% PEG 3350, 0.2 M disodium phosphate dihydrate, while selenomethionine crystals were produced in 50 mM MES (pH 6.5), 5% PEG 4000, 10% glycerol, and 100mM NaCl. Both native and selenomethionylated crystals were cryoprotected by successive brief incubations in mother liquor containing increasing concentrations of glycerol (up to 20%), and were subsequently flash-cooled in liquid nitrogen until data collection. A three-wavelength MAD experiment was performed on the Se edge on the ID-29 beamline (2.4 Å, space group P2<sub>1</sub>). The three data sets from the MAD experiment were processed and scaled with XDS (Kabsch, 1993); identification of selenium atom positions, refinement, and phasing were performed with AutoSHARP (Bricogne et al., 2003; de la Fortelle and Bricogne, 1997) and automatic model building was performed with ARP/wARP 7.0.1 (Perrakis et al., 1999). PHASER (Storoni et al., 2004) was subsequently used to perform molecular replacement using the model generated by the MAD experiment in order to phase a native data set (1.9 Å, space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>) collected at the European Synchrotron Radiation Facility (ESRF) ID23-EH1 beamline (Grenoble, France). Cycles of automatic and manual building were performed using ARP/wARP and Coot (Emsley and Cowtan, 2004), respectively, coupled with refinement cycles by REFMAC 5.4 (Murshudov et al., 1997).

### Fluorescence-Based Assays

Assays were conducted in an IQ5 96-well format real-time PCR instrument (Bio-Rad) in the presence of a Sypro Orange (Molecular Probes) probe. The total volume was to 25 μl. Samples were heat-denatured from 20°C to 100°C at a rate of 1°C per minute. At each step, excitation was performed at 470 nm, while emission of Sypro Orange fluorescence was monitored at 570 nm. Plotting of the fluorescence versus temperature curves, followed by the calculation of the first derivative at each point allowed the identification of each inflection point; the minima were referred to as the melting temperatures ( $T_m$ ). Data result from two independent experiments performed in triplicate.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at [doi:10.1016/j.str.2009.10.019](https://doi.org/10.1016/j.str.2009.10.019).

## ACKNOWLEDGMENTS

The coordinates of RrgA have been deposited at the Protein Data Bank with code 2WW8 and are available prerelease from the authors. The authors wish to thank Richard Kahn (IBS) and the staff of ESRF beamlines ID29 and ID23 for help with data collection, Izabel Bérard and Eric Forest from the IBS mass spectrometry facility (LSMP, IBS) for analyses, J. Marquez and the HTX Lab team (Partnership for Structural Biology) for access to and help with high-throughput crystallization, and Hugues Lortat-Jacob (IBS) for many helpful discussions. This work was partly supported by a Young Investigator (ANR Jeunes Chercheurs) grant to A.M.D.G. (n° 05-JCJC-0049) and by EC grant LSHM-CT-2004-512138 (to A.D.).

Received: August 21, 2009

Revised: September 29, 2009

Accepted: October 22, 2009

Published: January 12, 2010

## REFERENCES

- Abbot, E.L., Smith, W.D., Siou, G.P.S., Chiriboga, C., Smith, R.J., Wilson, J.A., Hirst, B.H., and Kehoe, M.A. (2007). Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. *Cell. Microbiol.* 9, 1822–1833.
- Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernero, J., Moschioni, M., Masignani, V., et al. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. USA* 103, 2857–2862.
- Bienkowska, J., Cruz, M., Atiemo, A., Handin, R., and Liddington, R. (1997). The von Willebrand Factor A3 domain does not contain a metal ion-dependent adhesion site motif. *J. Biol. Chem.* 272, 25162–25167.
- Bricogne, G., Vonnrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2023–2030.
- Bronzwaer, S.L., Cars, O., Buchholz, U., Molstad, S., Goettsch, W., Veldhuijzen, I.K., Kool, J.L., Sprenger, M.J.W., and Degener, J.E. (2002). A European study on the relationship between antimicrobial use and antimicrobial resistance. *Emerg. Infect. Dis.* 8, 278–282.
- Budzik, J.M., Marraffini, L.A., Souda, P., Whitelegge, J.P., Faull, K.F., and Schneewind, O. (2008). Amide bonds assemble pili on the surface of bacilli. *Proc. Natl. Acad. Sci. USA* 105, 10215–10220.
- Cruz, M.A., Yuan, H., Lee, J.R., Wise, R.J., and Handin, R.I. (1995). Interaction of the von Willebrand factor (vWF) with collagen. *J. Biol. Chem.* 270, 10822–10827.
- de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 276, 472–494.
- Deivanayagam, C.C.S., Rich, R.L., Carson, M., Owens, R.T., Danthuluri, S., Bice, T.W., Hook, M., and Narayana, S.V.L. (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8, 67–78.
- Dodson, K.W., Pinkner, J.S., Rose, T., Magnusson, G., Hultgren, S.J., and Waksman, G. (2001). Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor. *Cell* 105, 733–743.
- Dormitzer, P.R., Ulmer, J.B., and Rappuoli, R. (2008). Structure-based antigen design: a strategy for next generation vaccines. *Trends Biotechnol.* 26, 659–667.
- Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M.-C., Kojadinovic, M., Lalioui, L., Poyart, C., and Trieu-Cout, P. (2006). Assembly and role of pili in group B streptococci. *Mol. Microbiol.* 60, 1401–1413.
- Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008). Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol. Rev.* 32, 307–320.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J., and Liddington, R. (2000). Structural basis of collagen recognition by integrin  $\alpha 2\beta 1$ . *Cell* 101, 47–56.
- Fälker, S., Nelson, A.L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008). Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol. Microbiol.* 70, 595–607.
- Gianfaldoni, C., Censini, S., Hilleringmann, M., Moschioni, M., Facciotti, C., Pansegrau, W., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ruggiero, P. (2007). *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. *Infect. Immun.* 75, 1059–1062.
- Hilleringmann, M., Giusti, F., Baudner, B.C., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M.A., and Ferlenghi, I. (2008). Pneumococcal pili are composed of protofilaments exposing adhesive clusters of RrgA. *PLoS Pathog.* 4, e1000026.
- Huizinga, E.G., van der Plas, R.M., Kroon, J., Sixma, J.J., and Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure* 5, 1147–1156.
- Imberty, A., Lortat-Jacob, H., and Perez, S. (2007). Structural view of glycosaminoglycan-protein interactions. *Carbohydr. Res.* 342, 430–439.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800.
- Kamata, T., and Takada, Y. (1994). Direct binding of collagen to the I domain of integrin  $\alpha 2\beta 1$  (VLA-2, CD49b/CD29) in a divalent cation-independent manner. *J. Biol. Chem.* 269, 26006–26010.
- Kang, H.J., and Baker, E.N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *J. Biol. Chem.* 284, 20729–20737.
- Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* 318, 1625–1628.
- Kreikemeyer, B., Nakata, M., Oehmcke, S., Gschwendtner, C., Normann, J., and Podbielski, A. (2005). *Streptococcus pyogenes* collagen type I-binding Cpa surface protein. Expression profile, binding characteristics, biological function, and potential clinical impact. *J. Biol. Chem.* 280, 33228–33239.
- Krishnan, V., Gaspar, A.H., Ye, N., Mandlik, A., Ton-That, H., and Narayana, S.V.L. (2007). An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion. *Structure* 15, 893–903.
- Lee, J.-O., Bankston, L.A., Arnaout, M.A., and Liddington, R.C. (1995a). Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure* 3, 1333–1340.
- Lee, J.-O., Rieu, P., Arnaout, M.A., and Liddington, R. (1995b). Crystal structure of the A domain from the  $\alpha$  subunit of integrin CR3 (CD11b/CD18). *Cell* 80, 631–638.
- LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006). RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet. *Infect. Immun.* 74, 2453–2456.
- Levine, O.S., O'Brien, K.L., Knoll, M., Adegbola, R.A., Black, S., Cherian, T., Dagan, R., Goldblatt, D., Grange, A., Greenwood, B.M., et al. (2006). Pneumococcal vaccination in developing countries. *Lancet* 367, 1880–1882.
- Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E.T., Rosini, R., et al. (2005). Identification of a universal group B *Streptococcus* vaccine by multiple genome screen. *Science* 309, 148–150.
- Maisey, H.C., Hensler, M., Nizet, V., and Doran, K.S. (2007). Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. *J. Bacteriol.* 189, 1464–1467.
- Mandlik, A., Swierczynski, A., Das, A., and Ton-that, H. (2007). Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* 64, 111–124.
- Mandlik, A., Das, A., and Ton-That, H. (2008a). The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc. Natl. Acad. Sci. USA* 105, 14147–14152.
- Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2008b). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* 16, 33–40.
- Manzano, C., Contreras-Martel, C., El Mortaji, L., Izore, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A.M., and Dessen, A. (2008). Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* 16, 1838–1848.
- Manzano, C., Izoré, T., Di Guilmi, A.M., and Dessen, A. (2009). Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. *Biochemistry* 48, 10549–10557.
- Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G.O., 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. USA* 102, 15641–15646.
- Murshudov, G., Vagin, A., and Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.



- Nallapareddy, S.R., Singh, K.V., Sillanpaa, J., Garsin, D.A., Hook, M., Erlandson, S.L., and Murray, B.E. (2006). Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. *J. Clin. Invest.* *116*, 2799–2807.
- Neiers, F., Madhurantakam, C., Falker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009). Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J. Mol. Biol.* *393*, 704–716.
- Nelson, A.L., Ries, J., Bagnoli, F., Dahlberg, S., Falkner, S., Rounioja, S., Tschop, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., et al. (2007). RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. *Mol. Microbiol.* *66*, 329–340.
- Perrakis, A., Morris, R.M., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* *6*, 458–463.
- Proft, T., and Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria—structure, assembly, and their role in disease. *Cell. Mol. Life Sci.* *66*, 613–635.
- Qu, A., and Leahy, D.J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1,  $\alpha$ L $\beta$ 2) integrin. *Proc. Natl. Acad. Sci. USA* *92*, 10277–10281.
- Scott, J.R., and Zahner, D. (2006). Pili with strong attachments: Gram-positive bacteria do it differently. *Mol. Microbiol.* *62*, 320–330.
- Soto, G.E., and Hultgren, S.J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.* *181*, 1059–1071.
- Springer, T.A. (2006). Complement and multifaceted functions of VWA and integrin I domains. *Structure* *14*, 1611–1616.
- Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* *60*, 432–438.
- Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., and Grandi, G. (2006). Pili in Gram-positive pathogens. *Nat. Rev. Microbiol.* *4*, 509–519.
- Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* *12*, 228–234.
- Tuckwell, D., Calderwood, D.A., Green, L.J., and Humphries, M.J. (1995). Integrin  $\alpha$ 2 I-domain is a binding site for collagens. *J. Cell Sci.* *108*, 1629–1637.
- Vorup-Jensen, T., Ostermeier, C., Shimaoka, M., Hommel, U., and Springer, T. (2003). Structure and allosteric regulation of the  $\alpha$ x $\beta$ 2 integrin I domain. *Proc. Natl. Acad. Sci. USA* *100*, 1873–1878.

**Note Added in Proof**

As this manuscript was going to press, Hilleringman and coworkers published work describing the electron microscopy analysis of streptococcal pili, where RrgA is clearly located at the tip of the pilus fibril (Hilleringman et al. (2009) *EMBO J.* *28*, 3921–3930.).

# Stability and Assembly of Pilus Subunits of *Streptococcus pneumoniae*\*<sup>§</sup>

Received for publication, November 6, 2009, and in revised form, January 20, 2010. Published, JBC Papers in Press, February 10, 2010, DOI 10.1074/jbc.M109.082776

Lamya El Mortaji<sup>‡</sup>, Remi Terrasse<sup>‡</sup>, Andrea Dessen<sup>§</sup>, Thierry Vernet<sup>‡</sup>, and Anne Marie Di Guilmi<sup>†1</sup>

From the <sup>‡</sup>Laboratoire d'Ingénierie des Macromolécules and the <sup>§</sup>Bacterial Pathogenesis Group, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 (CEA, CNRS, UJF), 41 rue Jules Horowitz, F-38027 Grenoble, France

Pili are surface-exposed virulence factors involved in bacterial adhesion to host cells. The *Streptococcus pneumoniae* pilus is composed of three structural proteins, RrgA, RrgB, and RrgC and three transpeptidase enzymes, sortases SrtC-1, SrtC-2, and SrtC-3. To gain insights into the mechanism of pilus formation we have exploited biochemical approaches using recombinant proteins expressed in *Escherichia coli*. Using site-directed mutagenesis, mass spectrometry, limited proteolysis, and thermal stability measurements, we have identified isopeptide bonds in RrgB and RrgC and demonstrate their role in protein stabilization. Co-expression in *E. coli* of RrgB together with RrgC and SrtC-1 leads to the formation of a covalent RrgB-RrgC complex. Inactivation of SrtC-1 by mutation of the active site cysteine impairs RrgB-RrgC complex formation, indicating that the association between RrgB and RrgC is specifically catalyzed by SrtC-1. Mass spectrometry analyses performed on purified samples of the RrgB-RrgC complex show that the complex has 1:1 stoichiometry. The deletion of the IPQTG RrgB sorting signal, but not the corresponding sequence in RrgC, abolishes complex formation, indicating that SrtC-1 recognizes exclusively the sorting motif of RrgB. Finally, we show that the intramolecular bonds that stabilize RrgB may play a role in its efficient recognition by SrtC-1. The development of a methodology to generate covalent pilin complexes *in vitro*, facilitating the study of sortase specificity and the importance of isopeptide bond formation for pilus biogenesis, provide key information toward the understanding of this complex macromolecular process.

Attachment of pathogenic bacteria to host cells is mediated by bacterial surface-exposed adhesins which, in some instances, are located at the tip of a filamentous surface appendages known as pili. Pili of the Gram-negative bacteria have been extensively described (1). Despite the diversity in their assembly pathways, they share a common structural feature involving the noncovalent association of pilin subunits. Even though a first description of Gram-positive pili was published in 1968 (2), the first studies regarding their assembly mechanism were published more recently for the *Corynebacteria diphtheriae* pilus (3). Since then, pili have been identified in many other Gram-positive bacteria, including *Actinomyces*, *Enterococcus*, *Bacil-*

*lus*, and several Streptococcal species, *S. pyogenes*, *S. agalactiae*, and *S. pneumoniae* (4–10). In contrast to Gram-negative bacteria, both major pilin polymerization and minor pilin association to the Gram-positive pilus backbone occur through the formation of covalent bonds catalyzed by specialized transpeptidases. These membrane-associated sortase enzymes recognize a LPXTG-like sequence (sorting motif) within their protein substrate (11, 12). The enzymatic reaction involves a nucleophilic cysteine residue, which is essential for the covalent intermolecular association between pilus subunits (13–15). In addition, intramolecular bonds within pilin subunits, formed between lysine and asparagine side chains, have been identified in the high resolution structures of major pilins Spy0128 from *S. pyogenes* (16) and SpaA from *C. diphtheriae* (17) as well as in the main adhesin of the *Streptococcus pneumoniae* pilus, RrgA (18). The role of these unusual intramolecular cross-links in pilin subunit stabilization has been shown by site-specific mutagenesis and both proteolytic and thermal stability studies (19, 20). Notably, Lys-Asn intramolecular bonds were also identified in the Cna collagen-binding domains A and B (CnaA and CnaB, respectively) of *Staphylococcus aureus* (20–22) as well as the N2CnaB domain of the *S. agalactiae* minor pilin subunit GBS52 (23) and the CnaA-similar ACE protein from *Enterococcus faecalis* (24). The CnaB domains are repetitive regions whose fold is reminiscent of the IgG fold (22).

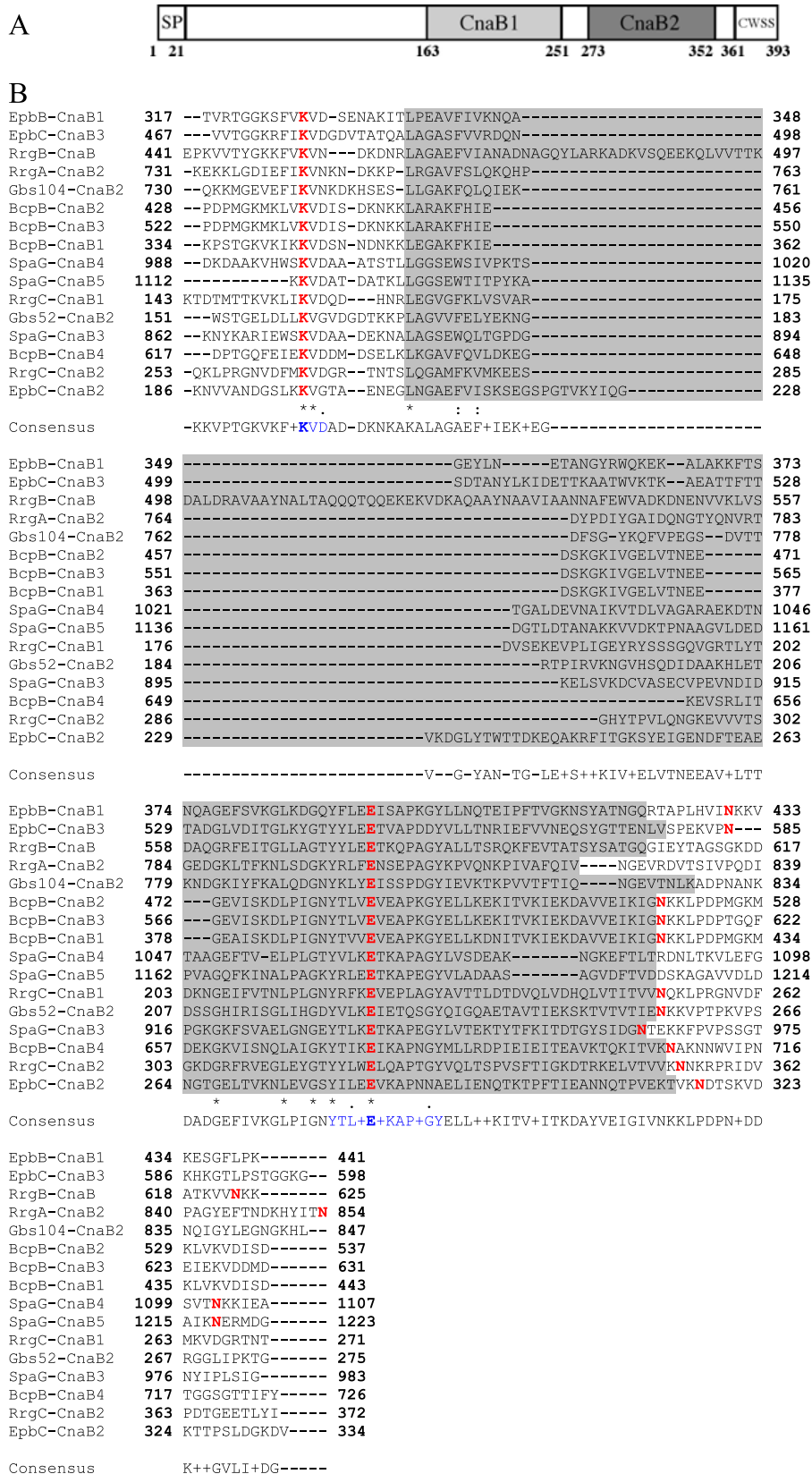
*S. pneumoniae* is an important human pathogen causing a large number of respiratory tract infections, such as pneumonia and sinusitis, but also invasive diseases like septicemia and meningitis. Among the surface-exposed virulence factors, the pneumococcal pilus has been recently shown to play a role in host cell adhesion (10). Pilus formation requires the expression of seven genes encoded by the *rlrA* pathogenicity islet, including a RofA-like transcriptional regulator (RlrA), three sortases (SrtC-1, SrtC-2, and SrtC-3), and three structural proteins (RrgA, RrgB, and RrgC) (10). Electron microscopy observations of immunogold-labeled pneumococcus allowed investigations on the structural composition of the pilus and showed that RrgB is the major pilin subunit forming the elongated fiber shaft, but data concerning the precise localization of the minor RrgA and RrgC pilin subunits remained controversial (10, 25–27). Recently, an elegant structural analysis of the native pneumococcal pilus using a combination of electron microscopy techniques provided evidence that the RrgB-shaft is composed of monomeric covalently linked subunits oriented head-to-tail (28). Furthermore, RrgA and RrgC do not localize together in the pili, but are present at the distal and the proximal end, respectively, in accordance with the host adhesion

\* This work was supported in part by EC Grant LSHM-CT-2004-512138 and ANR Grant 05-JC-JC-0049.

<sup>§</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1 and S2 and Tables S1–S3.

<sup>†1</sup> To whom correspondence should be addressed. Tel.: 33-4-38-78-56-34; Fax: 33-4-38-78-54-94; E-mail: diguilmi@ibs.fr.

## Pneumococcal Pilin Assembly



**FIGURE 1. Intramolecular bonds in RrgC.** *A*, predicted topology of RrgC. *SP*, signal peptide; the *CWSS* comprises a LPXTG-like motif followed by a hydrophobic membrane-spanning domain and a cytoplasmic positively charged tail. *B*, sequence alignments of pilin subunits from *S. pneumoniae* (RrgA, RrgC, and RrgB), *S. agalactiae* (Gbs52 and Gbs104), *B. cereus* (BcpB), *E. faecalis* (EpbB and EpbC), and *C. diptheriae* (SpaG). The CnaB domain is highlighted in gray; lysine, asparagine, and glutamate residues involved in the intramolecular bond are shown in red, and the consensus motifs including the Lys and the Glu residues are indicated in blue.

function of RrgA and the putative cell wall anchoring role of RrgC (28).

Investigation of the roles and relative contributions of the three sortases to the pilus formation process by genetic mutational approaches revealed that although deletion of individual sortases does not compromise RrgB polymerization, indicating a redundancy in sortase functionality, deletion of all three sortases abolishes RrgB fiber formation (27, 29). The role played by the three different sortases in incorporation of the minor pilin subunits onto the fiber remains largely unresolved: both SrtC-1 and SrtC-2 have been suggested as playing a role in incorporation of RrgA, and both SrtC-1 and SrtC-3 have been proposed as participating in association of RrgC to the RrgB backbone (27, 29).

Here, we provide insights into the molecular mechanisms of *S. pneumoniae* pilus biogenesis and of the specificity of cognate sortases by combining biochemical and biophysical approaches. In this work, we identify the intramolecular bonds present in RrgB and RrgC as well as their stabilization properties. We address the question of sortase specificity by employing a recombinant protein expression platform in which pilins and sortases are co-expressed. We show that co-expression of RrgB, together with RrgC and SrtC-1, leads to the formation of a covalent RrgB-RrgC complex whose association is abrogated when SrtC-1 carries a mutation in the nucleophilic cysteine. The deletion of the IPQTG sequence of RrgB, but not of the corresponding sequence in RrgC, impairs complex formation, indicating that SrtC-1 recognizes the sorting motif of RrgB prior to the final transpeptidation reaction with an amine group of RrgC. Finally, we provide evidence that the intramolecular bonds stabilizing RrgB may play a role in its recognition by SrtC-1 and subsequent association to RrgC.

### EXPERIMENTAL PROCEDURES

**Cloning Procedures**—Sequences corresponding to coding regions of RrgB (30–633) and RrgC (22–368),

**TABLE 1**  
ESI-TOF mass spectral analyses of the native and mutant proteins of RrgC

RrgC	$M_{\text{average}}^a$		$\Delta(M_{\text{expected}} - M_{\text{observed}})^b$	NH3 units lost
	Expected	Observed by ESI-TOF		
	<i>Da</i>	<i>Da</i>	<i>Da</i>	
Native	40,983.07	40,948.19	34.88	2
K155A	40,925.98	40,907.71	18.27	1
E222A	40,925.04	40,906.89	18.15	1
N252A	40,940.05	40,921.88	18.17	1
K264A	40,925.98	40,907.79	18.19	1
E322A	40,925.04	40,907.09	17.95	1
N354A	40,940.05	40,921.83	18.22	1
E222A/E322A	40,867.00	40,867.49	-0.49	0

<sup>a</sup> Average molecular mass.<sup>b</sup> Difference between expected and observed molecular mass.

lacking signal peptides, transmembrane anchors, and C-terminal tails, were amplified and cloned into a pLIM vector using the LIC method (P'X Therapeutics, Grenoble). Primers (listed in supplemental Table S1) and *S. pneumoniae* TIGR4 chromosomal DNA were used for PCR amplification. DNA fragments were also cloned into pETDuet and pACYCDuet bi-cistronic plasmids (Novagen), leading to constructs expressing combinations of proteins carrying histidine tags, S-tags, or devoid of tags. Resulting plasmids were then singly transformed or co-transformed into *Escherichia coli* STAR (Invitrogen) to obtain strains expressing all protein combinations (supplemental Table S2).

**Site-directed Mutagenesis**—Mutations were introduced by PCR-based site-directed mutagenesis (Stratagene QuikChange® II XL site-directed mutagenesis kit). Mutant plasmids were verified by DNA sequencing (Cogenics, Grenoble) and transformed into *E. coli* STAR for protein expression.

**Expression and Purification of RrgB, RrgC, and Mutants**—For both RrgB and RrgC, protein expression was induced in Terrific Broth with 0.5 mM IPTG<sup>2</sup> at 37 °C during 3 h. Cells were harvested by centrifugation and lysed by sonication in 50 mM HEPES, pH 7.5, 200 mM NaCl, 20 mM imidazole (buffer A), and a protease inhibitor mixture (Complete EDTA-free, Roche). The lysates were clarified by centrifugation and applied onto HisTrap™ HP columns (Amersham Biosciences) pre-equilibrated in buffer A. Protein elution was performed using linear gradients of 20–500 mM imidazole over 20 column volumes. Eluted proteins were dialyzed and purified by gel filtration chromatography. These proteins were used to immunize mice to produce polyclonal antibodies.

**Co-expression of Pilin Subunits, Sortase, and Complex Purification**—Bacterial cultures of Duet vectors transformed into *E. coli* STAR cells were performed, and protein expression was induced in 4 ml of Terrific Broth with 0.5 mM IPTG at 37 °C during 3 h. Cells were harvested by centrifugation and chemically lysed for 20 min in 1 ml of BugBuster® 1× (Novagen) diluted in buffer A and supplemented with 2 μl of r-lysozyme (Novagen) and 0.8 μl benzonase nuclease (Novagen). The

lysates were clarified by centrifugation and applied onto His MultiTrap™ HP (GE Healthcare) pre-equilibrated in buffer A. Protein were eluted in 50 mM HEPES, pH 7.5, 200 mM NaCl, 500 mM imidazole.

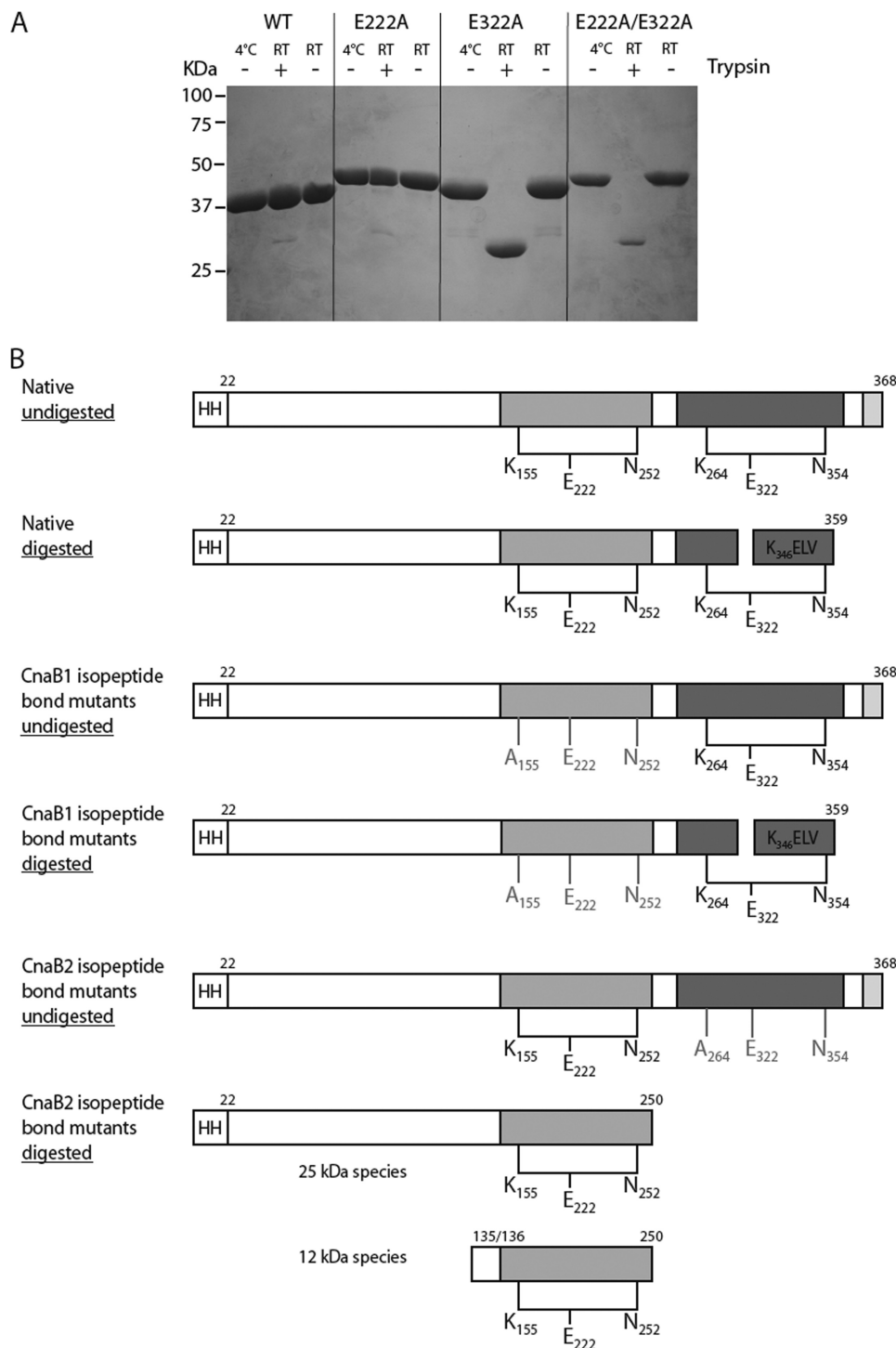
**Purification of the RrgB-RrgC Complex**—*E. coli* strain co-transformed with pETDuet-RrgB and pACYCDuet-RrgC-SrtC-1 vectors was cultured in 2 liters of Terrific Broth, and protein expression was induced with 0.5 mM IPTG at 37 °C during 3 h. Cells were harvested by centrifugation and lysed by sonication in buffer A containing protease inhibitor mixture (Complete EDTA-free, Roche). The lysates were clarified by centrifugation and applied onto a NiNTA resin column (Qiagen) pre-equilibrated in buffer A. Protein elution was performed using imidazole steps of 60, 100, and 300 mM. Fractions containing the RrgB-RrgC complex were pooled, dialyzed against 50 mM MES pH 6.0, 50 mM NaCl, and loaded onto an ion-exchange RQ-1 ml column (GE Healthcare) before elution by a linear gradient of NaCl (50–250 mM). A gel filtration chromatography S200 (GE Healthcare) equilibrated in 50 mM MES pH 6.0, 150 mM NaCl was the final purification step. The purification yield of the RrgB-RrgC complex was ~500 μg per liter of culture.

**Western Blotting**—Samples were loaded onto 4–12% Criterion XT Precast gels (Bio-Rad) and were run and subsequently electrotransferred in Trans-Blot Transfer Medium (Bio-Rad) for 30 min at 100 V. Incubation times of 1 h were successively performed using the appropriate antibody (polyclonal mouse, diluted 1:5 000) and anti-mouse horseradish peroxidase conjugate (Sigma) was diluted 1:120,000 before detection with a chemiluminescent substrate (Pierce).

**Thermal Shift Assay**—TSA experiments were carried out using an IQ5 96-well format real-time PCR instrument (Bio-Rad). Briefly, 25 ng of recombinant RrgC or RrgB proteins, native or mutant, were mixed with 2 μl of 5000× Sypro Orange (Molecular Probes) previously diluted 1:100 in water. Samples were heat-denatured from 20 to 95 °C at a rate of 1 °C per minute. Protein thermal unfolding curves were monitored by detection of changes in fluorescence of the SyproOrange. The inflection point of the fluorescence *versus* temperature curves was identified by plotting the first derivative over the temperature and the minima were referred to as the melting temperatures ( $T_m$ ). The fluorescence of buffers and salt solutions were checked as controls.

<sup>2</sup> The abbreviations used are: IPTG, isopropyl-1-thio-β-D-galactopyranoside; MES, 4-morpholineethanesulfonic acid; WT, wild type; CWSS, cell wall sorting signal.

## Pneumococcal Pilin Assembly



**FIGURE 2. Proteolysis stability of RrgC.** *A*, trypsin digestion products were separated on 12.5% SDS-PAGE. The molecular weight markers are shown on the left side of the gel. Whether trypsin was added or not is noted on top of the gel as well as the reaction temperatures. *B*, summary of the trypsin digestion data.

**Mass Spectrometry Analyses**—Accurate molecular masses of proteins were determined by ESI-TOF mass spectrometry performed on a LC/MSD-TOF mass spectrometry system (Applied Biosystems) in 95% acetonitrile, 5% water, 0.03% formic acid. 3  $\mu$ l of each sample were diluted in 97  $\mu$ l of formic acid 0.1% and 100  $\mu$ l of each dilution were desalted on line. LC/MS-MS experiments were carried out on the RrgB-RrgC

complex. The protein complex band was excised from the gel and subsequently washed with 50 mM acetonitrile/ammonium bicarbonate 50/50 (v/v) for 30 min, before dehydration with acetonitrile. Subsequently, the sample was treated with 7% H<sub>2</sub>O<sub>2</sub>. After drying was complete, the band was rehydrated in 15  $\mu$ l of digestion buffer (50 mM ammonium bicarbonate, pH 8.1) containing 150 ng of trypsin and incubated at 4 °C for 15 min. 30  $\mu$ l of digestion buffer were then added, and the digestion reaction was carried out at 37 °C overnight. Peptides were extracted from the gel by diffusion for 15 min with agitation, followed by three sequential 5 min sonication steps in 50% acetonitrile, 5% formic acid, and 100% acetonitrile. Digestion and extraction solutions were pooled and dried under vacuum. Peptide mixtures were re-dissolved in 25  $\mu$ l of water/acetonitrile 95/5 (v/v) containing 0.2% trifluoroacetic acid prior to LCMS/MS analysis. All experiments were performed in a 96-well system coordinated by an EVO15 (Tecan) robot.

**Proteolysis Assay and N-terminal Sequencing**—RrgC, RrgB, and mutants were diluted in a buffer composed of 50 mM HEPES pH 7.5 and 200 mM NaCl. Subsequently, 20  $\mu$ g of each Rrg protein was mixed with 50 ng of trypsin. The reaction was carried out at room temperature for 16 h. The samples were analyzed by SDS-PAGE. The proteins were either stained with a Coomassie Blue solution or transferred to a Problot membrane (Applied Biosystems) and stained according to supplier's instructions. The protein bands were cut from the membrane and sequenced using automated Edman degradation on an Applied Biosystem gas-phase sequencer model 477A with on-line analysis of the phenyl thiohydantoin derivatives.

## RESULTS

**Identification of the Residues Involved in the Formation of Two Intramolecular Bonds in the RrgC Minor Pilin**—Pfam analysis of the RrgC amino acid sequence predicted the presence of a signal peptide, the cell wall sorting signal (CWSS) and two

TABLE 2

## Molecular masses and N-terminal sequencing of the trypsin digestion of wild type and mutant RrgC variants

The calculated masses take into account the presence of one or two isopeptide bonds, depending on the RrgC variant considered.

	Undigested			Digested		
	Calculated mass	Measured mass	N-terminal sequence	Measured mass	N-terminal sequence	
Native	40,949.07	40,948.19	GSSHHHH	40,009.86	GSSHHHH KELVTVV	
K155A	40,908.98	40,907.71	GSSHHHH	39,969.72	GSSHHHH KELVTVV	
E222A	40,908.04	40,906.89	GSSHHHH	39,968.94	GSSHHHH KELVTVV VEPLAGY ELVTVVK	
N252A	40,923.05	40,921.88	GSSHHHH	39,983.95	GSSHHHH KELVTVV	
K264A	40,908.98	40,907.79	GSSHHHH	28,551.07 12,893.41 12,765.19	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>	
E322A	40,908.04	40,907.09	GSSHHHH	28,551.21 12,893.43 12,765.23	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>	
N354A	40,923.05	40,921.83	GSSHHHH	28,551.32 12,893.37 12,765.24	GSSHHHH ND <sup>a</sup> ND <sup>a</sup>	

<sup>a</sup> ND, not determined.

CnaB domains, CnaB1 and CnaB2 (Fig. 1A). The crystal structure of the pneumococcal pilin adhesin RrgA reveals the presence of two intramolecular Lys-Asn cross-links in an IgG-fold domain and in a CnaB domain (18). To test for the presence of such bonds in RrgC, we first searched its sequence for the presence of the putative conserved residues involved in isopeptide bond formation through multiple sequence alignments (Fig. 1B). This analysis clearly highlights a pattern of highly conserved lysine and glutamate residues which, in the crystal structures of GBS52 (23) and RrgA (18), are involved in the formation of the isopeptide bonds. The lysine and glutamate residues are included in KVD and GXYXLXEXAPXGY (E-box) motifs, respectively, while the asparagine residue is located ~25–30 residues downstream from the glutamate. We thus predicted that the intramolecular isopeptide bond in the CnaB1 domain of RrgC could be formed by Lys-155, Glu-222, and Asn-252, while the CnaB2 domain could harbor a bond formed by Lys-264, Glu-322, and Asn-354 (Fig. 1B).

A His-tagged recombinant form of RrgC (residues 22–368) was produced and mass spectrometry confirmed the presence of two isopeptide bonds in the native form (Table 1). The observed molecular mass (40,948 Da) was consistent with the loss of two NH<sub>3</sub> units (when compared with the calculated mass, 40,983 Da). Each bond-forming residue was mutated into alanine; single mutants as well as the double E222A/E322A mutant were produced and purified as the native molecule. Mass spectrometry measurements confirmed the loss of one NH<sub>3</sub> unit in each single mutant and the loss of two NH<sub>3</sub> units in the double mutant, supporting the role of Lys-155, Glu-222, Asn-252, Lys-264, Glu-322, and Asn-354 in the formation of the isopeptide bonds in RrgC.

*The Two Internal Covalent Bonds Stabilize RrgC*—Both native and mutant RrgC forms were tested for resistance to proteolytic digestion (Fig. 2A). Neither the native nor the CnaB1-E222A mutant seemed to be affected by trypsin digestion while the CnaB2-E322A and the double E222A-E322A mutant proteins were degraded into a proteolytic product of ~30kDa (Fig. 2A, lanes 8 and 11). It is interesting to note that in

the absence of the protease, all three mutant proteins, lacking either one or two isopeptide bonds, migrated on SDS-PAGE slightly slower than the native protein, suggesting that the intramolecular bonds could aid in the formation of a more compact protein fold (Fig. 2A).

Further characterization of the proteolytic products were carried out by mass spectrometry and N-terminal sequencing (Table 2 and Fig. 2B). In the native RrgC protein as well as in the three mutant variants which lack the CnaB1 isopeptide bond, an internal tryptic cleavage occurred in the CnaB2 domain at position <sup>346</sup>KELVT without dissociation of the polypeptide chains (Fig. 2A), indicating that the CnaB2 intramolecular bond is capable of holding these two chains together (Fig. 2B). In the mutants lacking the CnaB2 isopeptide bond, proteolysis induces the total digestion of the CnaB2 domain (25-kDa species) and of the RrgC N-terminal region (12-kDa species) suggesting that the isopeptide bond is key for the stabilization of this domain but also influences the overall stability of RrgC (Fig. 2C).

To further investigate the role played by the isopeptide bonds in RrgC stability, we performed thermal shift analyses of the different RrgC variants (Fig. 3). Native RrgC showed two distinct melting temperature ( $T_m$ ) values of 63.7 °C (±0.6) and 77.7 °C (±0.5). The CnaB1 bond mutants, namely K155A, E222A, and N252A, each presented a unique  $T_m$  value of 68.3 °C (±0.6), 57.4 °C (±1.3), and 61 °C (±1), respectively (Fig. 3A). The proteins mutated in the CnaB2 bond, namely K264A, E322A, and N354A also showed unique  $T_m$  values of 66, 65, and 67 °C, respectively (Fig. 3B). The double mutant E222A-E322A, deleted in both isopeptide bonds showed a drastically diminished  $T_m$  value of 47.9 °C (±2.1) (Fig. 3C). Taken together, these data show that the isopeptide bonds influence the stability of RrgC.

*Identification of Residues Involved in the Formation of Three Intramolecular Isopeptide Bonds in the Major RrgB Pilin*—We investigated the presence of intramolecular bonds in RrgB by using the same approach as for RrgC. A His-tagged recombinant form of RrgB (residues 30–633) was produced and purified from *E. coli*

## Pneumococcal Pilin Assembly

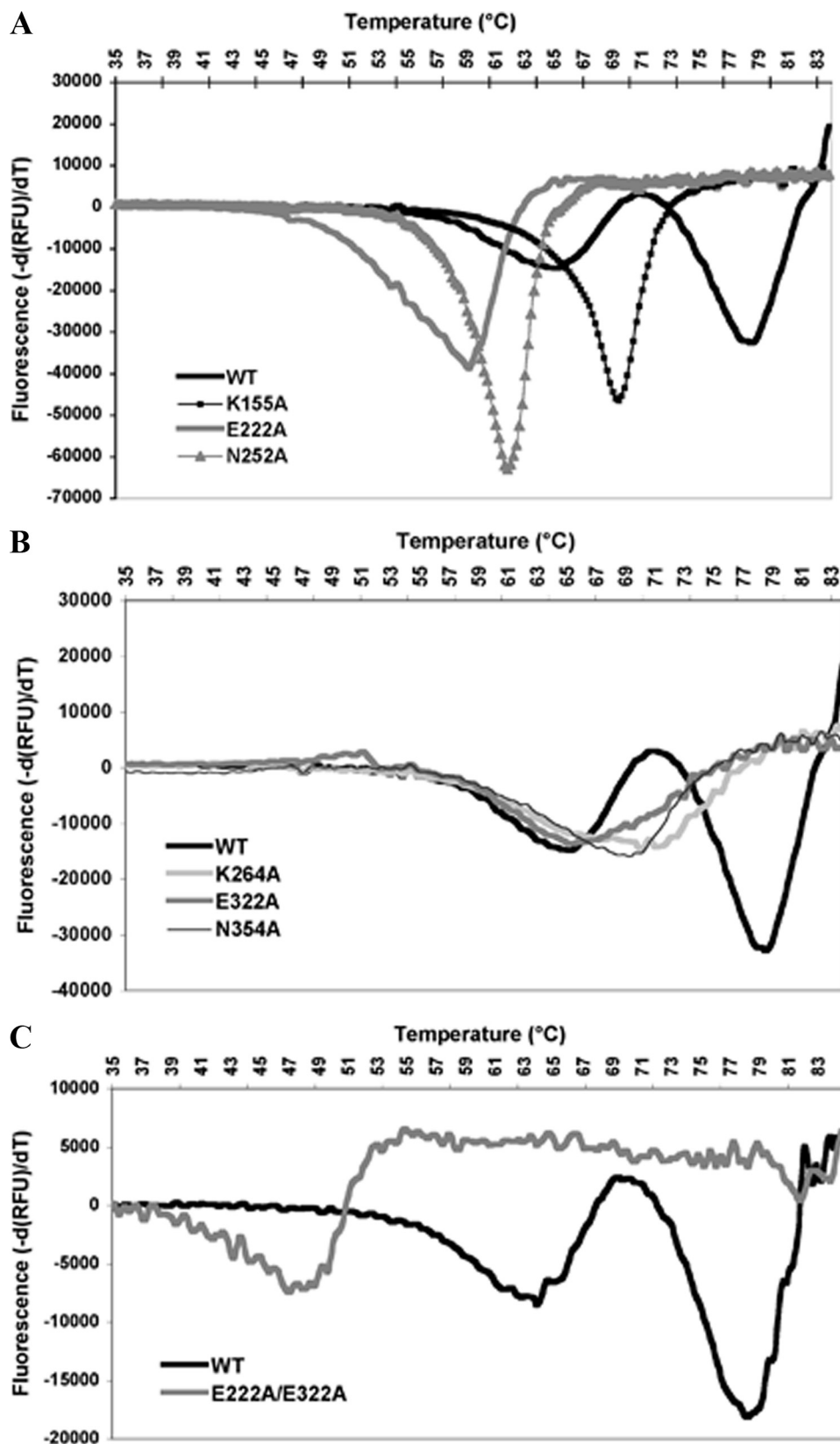


FIGURE 3. Thermal shift assay of RrgC and variants. The wild-type curve has been included in each figure to facilitate the comparison with the different RrgC variants. A, WT, and CnaB1 mutants. B, WT and CnaB2 mutants. C, WT and the double glutamate mutant.

cells. Mass spectrometry provided evidence of the presence of three isopeptide bonds since the molecular mass observed (67,400 Da) is consistent with the loss of three  $\text{NH}_3$  units (when compared with the expected mass, 67,453 Da) (Table 3).

In contrast to RrgC, the topology of RrgB presents a single predicted CnaB domain at the C terminus (Fig. 4A). The alignment of this RrgB domain with other pilins carrying CnaB predicted domains allowed us to identify that a potential isopeptide bond could involve residues Lys-453, Glu-577, and Asn-623, with the lysine and glutamate residues being included in KVD-like and GXYXLX-EXAPXGY (E-box) motifs (Figs. 1B and 4A). The E405, also present in an E-box motif, could play a role in the formation of another intramolecular bond in RrgB (Fig. 4A). Because no additional glutamate residue could be identified to explain the formation of the third bond, we looked for an aspartate residue which might play such a role, as observed in the RrgA crystal structure (18). Sequence alignments of RrgA and RrgB homologs suggested that Asp-241, defined in a D-box, might be involved in the formation of an intramolecular bond in RrgB (Fig. 4B). To assess the role of these various acidic residues in isopeptide bond formation, they were replaced by Ala either individually or together. Mass spectrometry confirmed that all single mutants lacked one  $\text{NH}_3$  unit, indicating that each mutated residue is involved in the formation of an isopeptide bond (Table 3). No isopeptide bonds were detected in the triple mutant (Table 3).

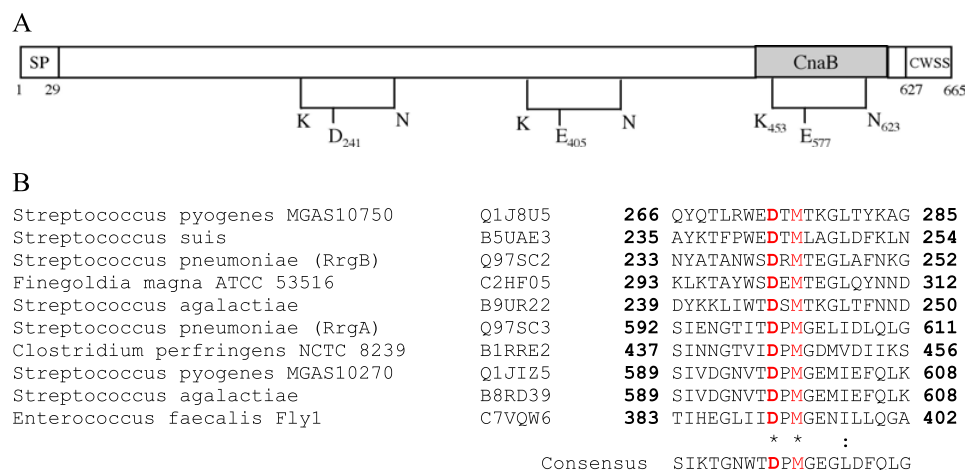
*The Three Internal Covalent Bonds Stabilize RrgB*—We investigated the role played by the three isopeptide bonds in RrgB stabilization by analyzing the proteolytic susceptibility and the thermal stability of the different RrgB variants (Fig. 5). Wild-type (WT) and RrgB variants were tested for resistance to proteolytic digestion (Fig. 5A). Neither native, E405A nor E577A mutant proteins were affected by trypsin digestion while the D241A

**TABLE 3**  
**ESI-TOF mass spectral analyses of the native and mutant proteins of RrgB**

RrgB	$M_{\text{average}}^a$		$\Delta(M_{\text{expected}} - M_{\text{observed}})^b$	NH <sub>3</sub> units lost
	Expected	Observed by ESI-TOF		
	<i>Da</i>	<i>Da</i>	<i>Da</i>	
Native	67,453.11	67,400.51	52.60	3
D241A	67,409.10	67,375.51	33.59	2
E405A	67,395.07	67,361.38	33.69	2
E577A	67,395.07	67,361.84	33.23	2
D241A/E405A/E577A	67,293.03	67,293.78	-0.75	0

<sup>a</sup> Average molecular mass.

<sup>b</sup> Difference between expected and observed molecular mass.



**FIGURE 4. Intramolecular bonds in RrgB.** *A*, topology of RrgB indicating the positions of the intramolecular bonds. The predicted CnaB domain is comprised between residues 547 and 605. The lysine and asparagine residues forming the bond-containing Glu-405 and Asp-241 have not been identified. *SP*, signal peptide; the cell wall sorting signal (*CWSS*) comprises a LPXTG-like motif followed by a hydrophobic membrane-spanning domain and a cytoplasmic positively charged tail. *B*, sequences alignment showing the D-box motif.

and the triple mutant D241A/E505A/E577A were cleaved in multiple fragments ranging from 20 to 50 kDa (Fig. 5*A*, lanes 5 and 14). It is interesting to note that in the absence of the protease, the D241A, E405A, and the triple mutant proteins migrate on SDS-PAGE slightly slower than the native protein, indicating that, as in the case of RrgC, isopeptide bond formation enhances protein compactness (Fig. 5*A*).

No significant thermal denaturation of the native RrgB was observed, an indication of the stable fold of the protein (Fig. 5*B*) while the D241A and E405A mutants displayed  $T_m$  values of 58.5 °C (±0.5) and 58.9 °C (±2.5), respectively (Fig. 5*B*). The E577A mutant presents an even more pronounced defect in stability as a three-state unfolding pattern is observed, characterized by the  $T_m$  values of 48.6 °C (±2.1), 59.2 °C (±2.0), and 68.8 °C (±1.9) (Fig. 5*B*). The triple RrgB mutant that does not contain any intramolecular bonds has a deeply decreased thermal stability reflected by the single low  $T_m$  value of 43.1 °C (±1.1) (Fig. 5*B*).

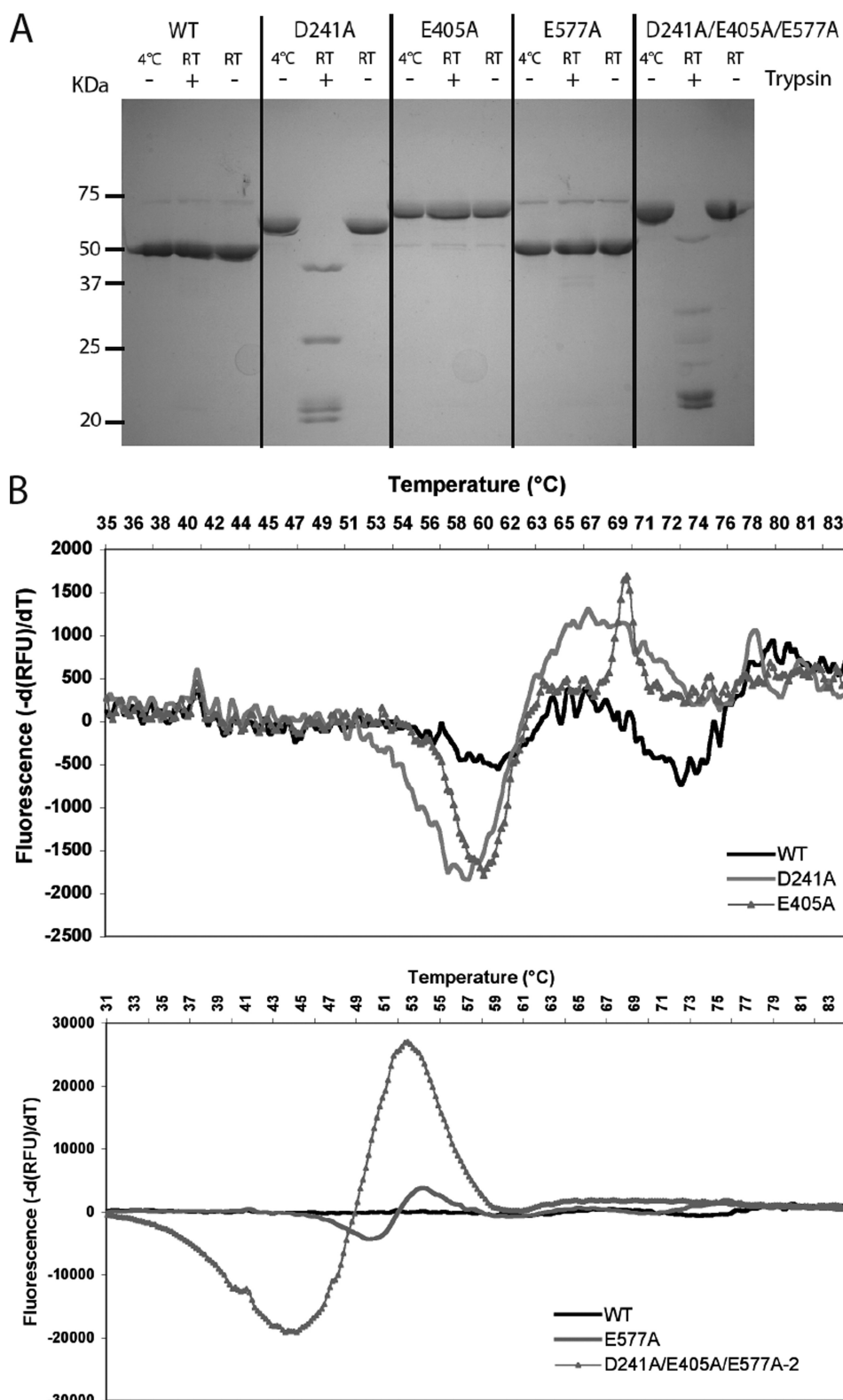
**SrtC-1 Covalently Associates RrgB and RrgC**—SrtC-1 is the main sortase responsible for RrgB polymerization (13). To determine if SrtC-1 displays restrictive specificity toward RrgB or is also able to recognize RrgC and RrgA, we designed a co-expression system which allows testing of sortase-pilin interactions (Fig. 6). Genes encoding the soluble forms of RrgA, RrgB, RrgC and SrtC-1 proteins were cloned in the bi-cistronic plasmids pETDuet and pACYCDuet, and the corresponding proteins were expressed in *E. coli* (Fig. 6). The co-expression

products were purified with His-Trap<sup>TM</sup>HP columns to which only His-tagged RrgA and RrgC, associated or not to RrgB, could be retained; eluted proteins were separated onto gradient polyacrylamide gels and immunoblotted with mouse polyclonal antibodies specific for RrgB (Fig. 7). As expected, no RrgB was immunodetected in the eluted fractions of strains expressing only RrgA or RrgC and SrtC-1 (Fig. 7, lanes 1, 3, 4). Despite the absence of a His tag on RrgB, two weak RrgB bands (60 and 70 kDa) are detected in the RrgB-containing eluted fractions which could be due to nonspecific binding to the HisTrap column (Fig. 7, lanes 2, 5, 7, 8, 16). However, strains that co-express SrtC-1, RrgB, and RrgC, present an intense band which migrates with a molecular apparent mass of 120 kDa and is recognized by both anti-RrgB (Fig. 7, lane 14) and anti-RrgC antibodies (data not shown). This finding suggests that SrtC-1 is able to covalently link the pilin subunits RrgB and RrgC. This same product was identified when RrgA was also co-expressed (Fig. 7, lane 15), although this pilin subunit is not involved in the formation of the RrgB-RrgC complex. In contrast, co-expression of SrtC-1 with RrgA and RrgB did not yield a complex between the two pilins (Fig. 7, lane 12) demonstrating a substrate specificity of sortase SrtC-1 for the association of RrgB with RrgC exclusively. To verify the catalytic nature of RrgB-RrgC complex formation, we mutated the SrtC-1 active site cysteine 193 into alanine and performed co-expression experiments as described above. No RrgB-RrgC complex could be identified, indicating that the association between RrgB and RrgC is specifically catalyzed by SrtC-1 (Fig. 7, lanes 16 and 17).

**SrtC-1 Recognizes the RrgB Sorting Signal**—To further characterize the covalent complex, we purified it to homogeneity and performed mass spectrometry analysis, which yielded a mass of 106,565.51 Da. This value corresponds precisely to the sum of the mass of one His<sub>6</sub>-tagged RrgC subunit associated to an IPQT-cleaved-S-tag RrgB subunit. The purified RrgB-RrgC complex was submitted to MS/MS experiments and peptides from both proteins were identified (supplemental Table S3). To verify which cell wall sorting signal is recognized by SrtC-1,



## Pneumococcal Pilin Assembly



**FIGURE 5. Proteolysis and thermal stability of RrgB and variants.** *A*, trypsin digestion products were separated on 12.5% SDS-PAGE. The molecular weight markers are shown on the left side of the gel. Whether trypsin was added or not is noted on top of the gel as well as the reaction temperatures. *B*, wild-type curve has been included in each figure to facilitate the comparison with the different RrgB variants.

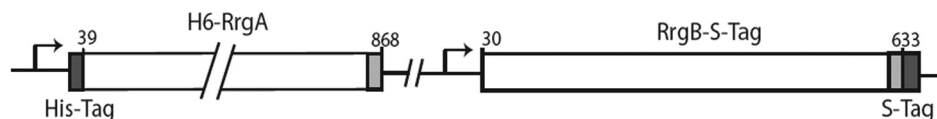
both RrgC and RrgB motifs were deleted, and the mutated proteins were tested in the SrtC-1 co-expression system (Fig. 8). The RrgB-RrgC complex could be successfully purified in the

140kDa and 150kDa (Fig. 9, lane 13). Finally, the absence of all the isopeptide bonds in both RrgC and RrgB lead to inefficient complex formation, since only a minor amount of RrgB-RrgC

absence of the RrgC VPDTG motif, while complex production was totally abrogated when RrgB lacks its IPQTG motif (Fig. 8). This observation demonstrates that catalysis by SrtC-1 proceeds through recognition of the RrgB IPQTG motif.

*Role of the Isopeptide Bonds in the RrgB-RrgC Complex Formation—* Isopeptide bonds stabilize RrgB and RrgC subunits, which are covalently associated by SrtC-1. Consequently, we investigated if the compact fold induced by the lysine-asparagine intramolecular bonds could influence the formation of the covalent RrgB-RrgC complex. The acidic residues shown to be involved in the formation and in the stabilization of the isopeptide bonds in both RrgB and RrgC were mutated into alanines in the co-expression system described above and the RrgB-RrgC complexes bearing the point mutations were purified and detected by the anti-RrgB antibodies as previously described (Fig. 9). The Glu-222 and Glu-322 mutations in RrgC and the single Asp-241, Glu-405, and Glu-577 mutations in RrgB did not affect complex formation (Fig. 9). However, when Asp-241 in RrgB was mutated into alanine, a significant quantity of monomeric RrgB was present in the HisTrap<sup>TM</sup>HP column eluate, indicating that a large amount of RrgB remained bound to RrgC through noncovalent interactions (Fig. 9, lane 10). The co-elution of the mutated RrgB variant is not due to artifactual binding to the resin as the recombinant RrgB mutants behave as the wild-type RrgB (Fig. 9). Furthermore, the noncovalent association of RrgB-D241A to RrgC is not dependent on the activity of the SrtC-1 as the RrgB-D241A is still co-eluted with RrgC and present in the elution fraction when the SrtC-1 is mutated in the catalytic Cys-193 (Fig. 9). When all three isopeptide bonds in RrgB were deleted, the RrgB-RrgC complex migrates with two apparent molecular masses of

pETDuet RrgA\_RrgB (Amp<sup>R</sup>)



pACYCDuet RrgC\_SrtC-1 (Cm<sup>R</sup>)

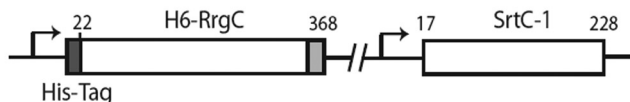


FIGURE 6. **Constructs for the co-expression platform.** Schematic representation of the pETDuet RrgA-RrgB and pACYCDuet RrgC-SrtC-1 constructs.

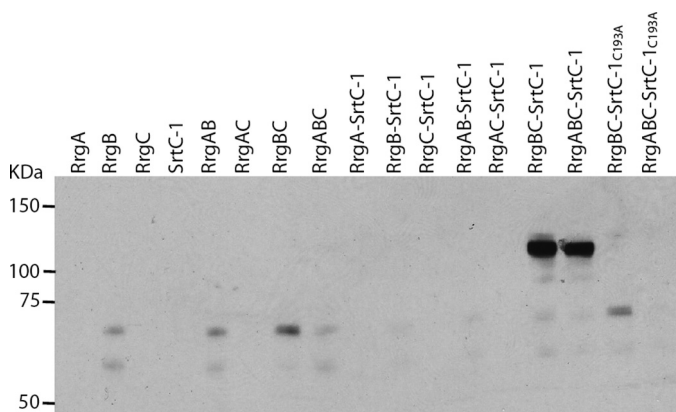


FIGURE 7. **Formation of the covalent complex RrgB-RrgC by SrtC-1.** Co-expression of RrgA, RrgB, RrgC, and SrtC-1 in *E. coli*. The purification was performed on HisTrap<sup>TM</sup> HP columns and only RrgA and RrgC are fused to His tag. The eluted proteins were separated on 4–12.5% SDS-PAGE, transferred on nitrocellulose membrane and immunodetected with an anti-RrgB polyclonal antibody.

complex could be formed (Fig. 9, lane 14). In conclusion, the stabilization of the fold of both RrgB and RrgC by intramolecular bond formation plays a role in the efficient recognition of the pilin subunits by SrtC-1 to catalyze the formation of the RrgB-RrgC covalent complex.

**DISCUSSION**

The recent discovery of pili expressed at the surface of Gram-positive bacteria allowed the publication of large amounts of data concerning pilus assembly, at the genetic and biochemical levels. However, the molecular details of substrate recognition and sortase catalysis remain largely unraveled. The pneumococcal pilus is composed of a fiber formed by RrgB, to which two minor pilins, RrgA and RrgC, are associated. Three sortases are involved in pilus assembly, SrtC-1, SrtC-2, and SrtC-3. One major challenge is to define the specificity of the sorting signal recognition and the molecular basis of the transpeptidation reaction of the pneumococcal-pilus-associated sortases.

The recent crystal structure of the RrgA adhesin showed the presence of two intramolecular bonds (18). In this work, we identified and characterized the intramolecular bonds in RrgC and RrgB. Sequence alignment analyses and mutagenesis data show that the presence of an E-box, containing a glutamate residue, is associated to the formation of the isopeptide bonds. The E-box motif had been described as playing a key role in the

incorporation of minor pilin subunits (30), and here we present experimental evidence that the role of the E-box is related to the stabilization of the intramolecular bonds (see also below). We also describe a new D-box motif in which the aspartate residue plays a role in the formation of the isopeptide bond. The lysine residue forming the intramolecular bond is contained in a small conserved sequence motif, KVD. Recombinant RrgB contains three intramolecular isopeptide

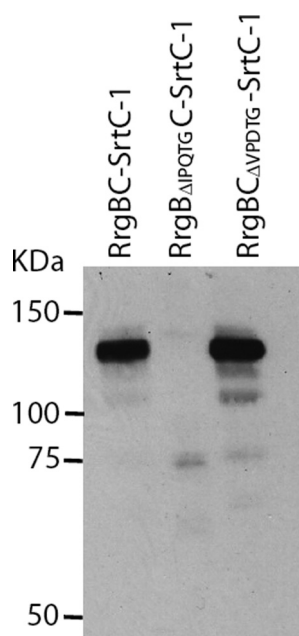
bonds even though sequence analysis revealed a potential fourth site involving residues Lys-41, Asn-184, and Glu-143. It cannot be excluded that this supplementary bond is formed in the native protein. The intramolecular bonds in the CnaB domains of RrgC and RrgB are most probably located in the same position as in the analog domains GBS52-N2 (23) and RrgA-D4 (18): this is suggested by the structural alignments of GBS52-N2 and RrgA-D4 with the RrgC and RrgB CnaB sequences, highlighting the conservation of the lysine and glutamate residues participating in the intramolecular bonds, with the secondary structures elements (supplemental Fig. S1).

The role of the isopeptide bonds in RrgC and RrgB in protein stability was also investigated. The thermostability assay used in this study is based on the modification of the fluorescence signal emitted by the SyproOrange probe upon its binding to hydrophobic residues that are exposed while the protein denatures. Consequently, the protein thermal unfolding assay is performed by monitoring of a mixture of protein with the probe over a range of temperatures. The reported temperature values correspond to the melting temperatures ( $T_m$ ) calculated as the inflection point of the fluorescence *versus* temperature; they correspond to the mid-point temperature values determined in circular dichroism experiments. TSA studies thus are good indicators of the multidomain organization of a protein; if distinct domains unfold independently, this could give rise to multiple  $T_m$  curves, as observed for RrgB.

Both thermostability and protease sensitivity data show that the absence of intramolecular bonds in mutated variants of RrgC and RrgB destabilizes the proteins. However, how these intramolecular bonds cause the protein to be more stable and compact remains an open question. We propose the following hypothesis. The CnaB domains resemble the IgG-like structure with a four- and a three- $\beta$ -stranded barrel, but exhibit a novel IgG reverse fold (23). The Lys-Asn bond cross-links the three- $\beta$ -stranded sheet while the Glu residue points from the four- $\beta$ -stranded sheet (supplemental Fig. S2). Consequently, the intramolecular bonds act as locks which maintain a globally closed conformation of the  $\beta$ -barrel, as observed for the disulfide bonds in IgG, thus leading to an increased compactness and stability of the proteins.

We observed that point mutations of the Lys and Asn residues and of the catalytic Glu or Asp residues impair the isopeptide bond formation. Rearrangements induced by such mutations have been observed recently by Kang and Baker (19) in the

## Pneumococcal Pilin Assembly

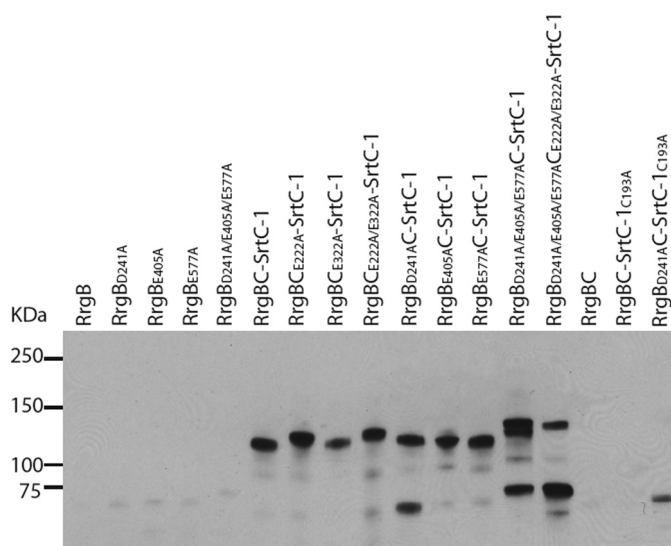


**FIGURE 8. Substrate recognition specificity of SrtC-1.** Both sorting signals in RrgB and RrgC were deleted, and each mutant was analyzed for its ability to be recognized and processed by SrtC-1. The purification of the co-expressed proteins, RrgB, RrgC, and SrtC-1 was performed on HisTrap<sup>TM</sup>HP columns. The eluted proteins were separated on 4–12.5% SDS-PAGE, transferred onto a nitrocellulose membrane, and immunodetected with an anti-RrgB polyclonal antibody.

*S. pyogenes* major pilin. The crystal structure of the E117A variant showed that the side chains of Lys-36 and Asn-168 can adopt alternative orientations due probably to the inappropriate positioning of ionic bonding partners. However, this remains a local effect which does not cause major protein conformational changes. We believe that a similar situation occurs in the mutants described in this work.

Our results also show that the pilin intramolecular bonds influence the formation of the RrgB-RrgC complex. Indeed, a significant amount of the D241A RrgB mutant monomer was co-eluted with RrgC showing that both pilin subunits interact in a stable noncovalent complex whose formation is independent on the enzymatic activity of the SrtC-1. The structural changes caused by the absence of this single intramolecular isopeptide bond in RrgB reduce the ability of SrtC-1 to covalently form RrgB-RrgC. These findings suggest that SrtC-1, in addition to recognizing the IPQTG motif of RrgB, also interacts with other region(s) of the pilin, which could be less accessible when the Asp-241 intramolecular bond does not stabilize the RrgB structure. This result is in accordance with what was observed for BcpA, the shaft forming pilin of *B. cereus*, which contains three intramolecular bonds, one of which plays an important role in pilus formation (31).

To address the question of the pilus biogenesis mechanisms at the molecular level, we used biochemical approaches that mimic the assembly reactions *in vitro*. We developed an expression platform dedicated to the production of recombinant pilins and sortases, in different combinations, in order to reconstitute the pneumococcal pilus assembly. The co-expression in *E. coli* of RrgB, RrgC, and SrtC-1 led to the formation of a covalent RrgB-RrgC complex which was catalyzed by SrtC-1. Fol-



**FIGURE 9. Role of the internal bonds in RrgB and RrgC in the covalent complex formation.** The variants of RrgB and RrgC deleted in the internal bonds were introduced in the co-expression platform. The purification of the different RrgB forms (wild-type, single, and triple mutants) and of the co-expressed proteins, RrgB, RrgC, and SrtC-1 was performed on HisTrap<sup>TM</sup>HP columns. The eluted proteins were separated on 4–12.5% SDS-PAGE, transferred on nitrocellulose membrane, and immunodetected with an anti-RrgB polyclonal antibody.

lowing cleavage of the IPQTG sorting signal of RrgB, SrtC-1 forms an amide bond between the C-terminal threonine and an NH<sub>2</sub> group on the RrgC subunit. The RrgB-RrgC complex is the first sortase-product between two different pneumococcal pilins obtained in *in vitro* conditions. In this work, we do not provide accurate quantitative data about the sortase enzymatic efficiency as the overexpression is performed in a heterologous host. However, an estimation of the pneumococcal sortase efficiency in *E. coli* can be estimated to be around 1% based on the purification of 500 μg of RrgB-RrgC complex from about 30–50 mg of RrgB, RrgC, and SrtC-1 expressed.

The number of pilus-associated sortases is variable: one sortase is present in *Bacillus* and *Corynebacteria* species, two in GAS and GBS species and three in *S. pneumoniae*. Consequently, it can be assumed that one pilus-associated sortase may recognize different sorting motifs and/or accept different nucleophilic amine groups. One example of such redundancy concerns SrtD of *B. cereus* which recognizes both BcpA and BcpB (31, 32). We describe a different situation in *S. pneumoniae*: SrtC-1 only recognizes and cleaves the IPQTG sorting signal of RrgB but catalyzes the formation of an amide bond with either another subunit of RrgB, leading to formation of the fiber polymer (13), or with a RrgC monomer, to form the heterodimeric RrgB-RrgC complex (this work). The mechanism at the basis of the balance between both reactions remains to be described. The *in vitro* RrgB-RrgC complex formation is relevant in regard to the current knowledge of pilus biogenesis: RrgC is associated to the pilus fiber (26, 28) and SrtC-1 has been proposed to incorporate RrgC in the native pilus as observed by mutational genetics in the TIGR4 strain (27). Consequently, we propose that the complex RrgB-RrgC whose association is catalyzed by SrtC-1, may represent a molecular unit onto which pilus elongation and/or its cell wall insertion may take place.

*Acknowledgments*—We thank Izabel Bérard and Eric Forest from the IBS mass spectroscopy facility (LSMP, IBS) for analyses, Alexandra Kraut and Jérôme Garin (Lab. d'Etude de la Dynamique des Protéomes, EDyP Grenoble) for access to the LC-MS/MS platform and experiments, and Jean-Pierre Andrieu for (LEM, IBS) for N-terminal sequencing.

## REFERENCES

1. Proft, T., and Baker, E. N. (2009) *Cell Mol. Life Sci.* **66**, 613–635
2. Yanagawa, R., Otsuki, K., and Tokui, T. (1968) *Jpn J. Vet. Res.* **16**, 31–37
3. Ton-That, H., and Schneewind, O. (2003) *Mol. Microbiol.* **50**, 1429–1438
4. Mishra, A., Das, A., Cisar, J. O., and Ton-That, H. (2007) *J. Bacteriol.* **189**, 3156–3165
5. Kline, K. A., Kau, A. L., Chen, S. L., Lim, A., Pinkner, J. S., Rosch, J., Nallapareddy, S. R., Murray, B. E., Henriques-Normark, B., Beatty, W., Caparon, M. G., and Hultgren, S. J. (2009) *J. Bacteriol.* **191**, 3237–3247
6. Budzik, J. M., Marraffini, L. A., and Schneewind, O. (2007) *Mol. Microbiol.* **66**, 495–510
7. 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) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15641–15646
8. Rosini, R., Rinaudo, C. D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzi, C., Brettoni, C., Buccato, S., Margarit, I., Grandi, G., and Telford, J. L. (2006) *Mol. Microbiol.* **61**, 126–141
9. Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prévost, M. C., Kojadinovic, M., Lalioui, L., Poyart, C., and Trieu-Cuot, P. (2006) *Mol. Microbiol.* **60**, 1401–1413
10. Barocchi, M. A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Massignani, 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) *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2857–2862
11. Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008) *FEMS Microbiol. Rev.* **32**, 307–320
12. Maresso, A. W., and Schneewind, O. (2008) *Pharmacol. Rev.* **60**, 128–141
13. Manzano, C., Contreras-Martel, C., El Mortaji, L., Izoré, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A. M., and Dessen, A. (2008) *Structure* **16**, 1838–1848
14. Manzano, C., Izoré, T., Job, V., Di Guilmi, A. M., and Dessen, A. (2009) *Biochemistry* **48**, 10549–10557
15. Neiers, F., Madhurantakam, C., Fälker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B., and Achour, A. (2009) *J. Mol. Biol.* **393**, 704–716
16. Kang, H. J., Coulibaly, F., Clow, F., Proft, T., and Baker, E. N. (2007) *Science* **318**, 1625–1628
17. Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H., and Baker, E. N. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16967–16971
18. Izoré, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A. M., and Dessen, A. (2010) *Structure* **18**, 106–115
19. Kang, H. J., and Baker, E. N. (2009) *J. Biol. Chem.* **284**, 20729–20737
20. Symersky, J., Patti, J. M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., Jin, L., Schneider, A., DeLucas, L. J., Höök, M., and Narayana, S. V. (1997) *Nat. Struct. Biol.* **4**, 833–838
21. Zong, Y., Xu, Y., Liang, X., Keene, D. R., Höök, A., Gurusiddappa, S., Höök, M., and Narayana, S. V. (2005) *EMBO J.* **24**, 4224–4236
22. Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., Höök, M., and Narayana, S. V. (2000) *Structure* **8**, 67–78
23. Krishnan, V., Gaspar, A. H., Ye, N., Mandlik, A., Ton-That, H., and Narayana, S. V. (2007) *Structure* **15**, 893–903
24. Liu, Q., Ponnuraj, K., Xu, Y., Ganesh, V. K., Sillanpää, J., Murray, B. E., Narayana, S. V., and Höök, M. (2007) *J. Biol. Chem.* **282**, 19629–19637
25. LeMieux, J., Hava, D. L., Basset, A., and Camilli, A. (2006) *Infect Immun.* **74**, 2453–2456
26. Hilleringmann, M., Giusti, F., Baudner, B. C., Massignani, V., Covacci, A., Rappuoli, R., Barocchi, M. A., and Ferlenghi, I. (2008) *PLoS Pathog.* **4**(3)
27. Fälker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008) *Mol. Microbiol.* **70**, 595–607
28. Hilleringmann, M., Ringler, P., Müller, S. A., DeAngelis, G., Rappuoli, R., Ferlenghi, I., and Engel, A. (2009) *EMBO J.* **28**, 3921–3930
29. LeMieux, J., Woody, S., and Camilli, A. (2008) *J. Bacteriol.* **190**, 6002–60013
30. Ton-That, H., Marraffini, L. A., and Schneewind, O. (2004) *Mol. Microbiol.* **53**, 251–261
31. Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F., and Schneewind, O. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10215–10220
32. Budzik, J. M., Oh, S. Y., and Schneewind, O. (2009) *J. Biol. Chem.* **284**, 12989–12997





## **RESUME EN FRANÇAIS :**

*Streptococcus pneumoniae* est un pathogène majeur chez l'homme, responsable d'otites, de pneumonies, de septicémies et de méningites. Récemment des structures de type pilus ont été identifiées à la surface de *S. pneumoniae* et jouent un rôle important dans les étapes initiales de colonisation des tissus hôtes. Six gènes sont impliqués dans la formation de cette structure. Trois d'entre eux codent pour les protéines structurales ou pilines (RrgA, RrgB et RrgC) et trois autres gènes codent pour les enzymes, appelées sortases, qui catalysent l'association covalente des pilines (SrtC-1, SrtC-2 et SrtC-3). Des modèles de formation du pilus ont été proposés suite à des études de délétion génétique, mais aucune donnée biochimique permettant d'expliquer précisément la formation du pilus au niveau biomoléculaire n'est encore disponible.

L'étude individuelle des protéines impliquées dans la formation du pilus a permis la mise en évidence de ponts intramoléculaires Lys-Asn stabilisateurs présents dans chacune des pilines. De plus, la résolution cristallographique de RrgA et RrgB permet de mieux comprendre les propriétés adhésives de cette structure mais également son mécanisme d'assemblage.

Comme le rôle de chacune des sortases reste imprécis, nous avons développé un système de co-expression permettant de tester toutes les combinaisons de pilines et de sortases. Celui-ci nous a permis d'identifier les spécificités de chacune des sortases, de générer des complexes covalents piline/piline mais également piline/sortase et ainsi d'obtenir des éléments clés dans la compréhension de la biogenèse de cette structure.

**TITRE EN ANGLAIS:** Molecular Mechanisms of the pneumococcal pilus biogenesis

## **RESUME EN ANGLAIS**

*Streptococcus pneumoniae* is the most common cause of otitis, sinusitis, pneumonia, sepsis and meningitis. Recently, pili were identified at the surface of *S. pneumoniae* and were proposed to play a role in the initial step of host tissues colonisation. Six genes are involved in the pilus formation. Three of them encode structural proteins named pilins (RrgA, RrgB and RrgC), the three others encode specific enzymes called sortases, which catalyse the covalent association of the pilins (SrtC-1, SrtC-2 and SrtC-3). Pilus formation models have been proposed based on genetic studies, but no biochemical data explaining precisely the molecular process of pilus formation is yet available.

The individual study of each pilin led to the identification of stabilising Lys-Asn intramolecular bonds in each protein. Moreover, the crystallographic structure of RrgA and RrgB provided precious informations regarding the adhesive properties as well as the mechanism of pilus assembly.

Since the role of each sortase remains unclear, our aim was to elucidate, step by step, the molecular mechanisms involved in the pilus biogenesis. Therefore, we have developed a co-expression platform allowing the production of the pilins subunits together with sortases in *E.coli*. This system allowed to decipher the substrate specificity of the sortases, to generate covalent pilin/pilin complexes, as well as pilin/sortase complexes and thus provided key information towards the understanding complex macromolecular process.

## **MOTS-CLES/KEY WORDS**

*S. pneumoniae*, pilus, pilines, sortases, ponts intramoléculaires / *S. pneumoniae*, pilus, pilins, sortases, intramolecular bonds

## **INTITULE ET ADRESSE DU LABORATOIRE**

Institut de Biologie Structurale Jean Pierre Ebel (UMR 5075)

41, rue Jules Horowitz 38027 Grenoble Cedex