ANNÉE 2017

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Biologie – Santé

Ecole doctorale Vie – Agro – Santé

présentée par

Tony MAURO

Préparée à l'unité de recherche INSERM U1230
Laboratoire de Biochimie Pharmaceutique
Faculté de Pharmacie

Étude de la régulation transcriptionnelle de deux ARN régulateurs de *Staphylococcus aureus* : Implication d'un facteur de transcription de la famille SarA

Thèse soutenue à RENNES
le 9 Mars 2017

devant le jury composé de :

Dr. Maude GUILLIER
Chargé de recherche - CNRS UMR8261 (Paris) / *rapporteur*

Dr. Nara FIGUEROA-BOSSI
Directeur de recherche - I2BC (Gif sur Yvette) / *rapporteur*

Pr. Peter REDDER
Professeur - CNRS UMR 5100 (Toulouse) / *examinateur*

Pr. Luc PAILLARD
Professeur - IGDR (Rennes) / *examinateur*

Pr. Brice FELDEN
Professeur - INSERM U1230 (Rennes) / *directeur de thèse*

Dr. Astrid ROUILLON
MCU - INSERM U1230 (Rennes) / *co-directeur de thèse*

« Quand on vous demande si vous êtes capable de faire un travail répondez : "bien sûr, je peux !" Puis débrouillez-vous pour y arriver. »

Theodore Roosevelt



REMERCIEMENTS

La thèse a la particularité d'être à la fois l'aboutissement d'un travail et le début d'une aventure. Cette période est loin d'être solitaire et je dirai même plus : j'ai été entouré d'une équipe formidable. La fin de ma thèse approchant, arrive donc le moment de vous remercier.

Je tiens à remercier les membres de mon jury de thèse, Maude GUILLIER, Nara FIGUEROA-BOSSI, Peter REDDER et Luc PAILLARD d'avoir accepté d'évaluer ce travail. Je remercie également la Direction Générale de l'Armement, la région Bretagne, la Fondation pour la Recherche médicale et l'Agence Nationale de la Recherche pour m'avoir financé aux cours de ces dernières années. Merci à Emmanuelle GUILLOT-COMBE (tutrice DGA) pour son soutien.

Un merci particulier à Brice FELDEN pour m'avoir accueilli au sein de son laboratoire, pionnier dans la découverte des ARN régulateurs chez *S. aureus*. Depuis mon arrivée dans le laboratoire, le staphylocoque doré est passé pour moi du statut de bactérie comme les autres à un pathogène à abattre, et cela, je vous le dois ! Mon intérêt pour cette petite vermine n'est pas prêt de disparaître.

Autres choses qui ne sont pas prêtes de disparaître, c'est toute la reconnaissance, le respect, l'amitié et la considération que j'ai pour toi Astrid. La première fois que nous nous sommes rencontrés, c'était le 10 octobre 2012. Le 10 octobre est donc une fois de plus un jour pas comme les autres. Au-delà d'être le jour de mon anniversaire, cela a été le jour où tu m'as accepté comme stagiaire. SprC, SarA et Star ont jalonné ma thèse. Tout au long de ces 4 années avec toi, tu m'as soutenu dans les moments heureux ou difficiles. Tu as toujours été disponible, à l'écoute et pour tout cela je te remercie du plus profond de mon cœur.

A Lorraine et Régine… je crois que, sans vous, cette thèse n'aurait pas eu la même saveur. Ce n'est pas pour nous jeter des fleurs, mais sans nous, l'ambiance dans le laboratoire n'aurait sensiblement pas été la même. Entre les fous rires, danses et chants, nous avons réussi à finaliser notre thèse dans la bonne humeur. Certes quelques perturbateurs ont essayé de nous miner le moral, mais ensemble nous avons tenu bon ! Vous n'êtes pas des simples collègues, vous êtes devenues de vraies amies.

Ma petite Julie, toi que je martyrise depuis ton arrivée… Mais « qui aime bien, châtie bien » !!! Notre point commun, nous avons la même culture musicale et le bureau des thésards est devenu beaucoup plus animé depuis ton installation (même si à cause de toi, nous sommes plus en tête à tête avec Lorraine…). Je ne me fais aucun souci pour ton avenir dans le

laboratoire. Malgré le départ des trois aînés, tu vas continuer à mettre de la gaieté dans ce monde de science.

Camille, Simon, Guillaume et à tous les nouveaux stagiaires (ça se multiplie à vue d'œil ces bêtes-là), je compte sur vous pour faire à la fois de la bonne science et de la bonne bouffe pour vos titulaires (surtout n'oubliez pas Astrid parce qu'elle ne vous oubliera pas ^^).

Nadège, toi qui a décidé de me quitter pendant ma thèse (ou de quitter le laboratoire plutôt ?), je n'oublierai jamais ces ~~pauses cafés~~ discussions professionnelles le mercredi après-midi, toi, à ton bureau, moi sur mon petit tabouret qui ne verra plus jamais mon postérieur, et toi non plus d'ailleurs.

Noëlla, merci de m'avoir chouchouté, dorloté et materné pendant ces 4 années. Sans nous, la gestion des commandes serait sans doute bancale.

Nathalie, quand on venait le matin de bonnes heures (et pas forcément de bonne humeur), tu étais toujours là, à l'écoute de nos histoires et de nos bêtises aussi.

Valérie, le point commun que nous partageons est un humour qui est, parfois et je dis bien parfois, placé un peu bas mais qu'est-ce qu'on y peut nous ? C'est drôle !!!

Yoann et Gaétan, un immense merci pour toutes les fois où je suis venu vers vous pour des corrections d'anglais et pas que. Un grand merci à Irène, Pierre-Yves et Hélène pour leur aide au cours de l'écriture de ce manuscrit.

Un grand MERCI à tous les autres aussi, Mohamed, Marie-Laure, Marc, Sveta et Annick pour toute l'aide que vous m'avez apportée.

A tous les nouveaux avec qui je n'ai pas eu beaucoup le temps de discuter, je vous souhaite une bonne intégration et que vous apportiez plein de connaissances à notre laboratoire. A tous les anciens qui sont partis du laboratoire, merci pour tout.

Mille mercis à ma famille et mes amis qui ont dû me supporter pendant ces dernières années et que je vais sans doute abandonner pendant quelques temps.

Et merci à tous ceux et celles qui m'ont permis, d'une façon ou d'une autre, d'accéder à cette thèse.

SOMMAIRE

LISTE DES FIGURES 5

LISTE DES TABLEAUX 11

LISTE DES ABREVIATIONS 13

INTRODUCTION 15

I. *STAPHYLOCOCCUS AUREUS* 17

A. Généralités 17

B. Infections à S. aureus 18

C. Génétiques de S. aureus 22

II. LA TRANSCRIPTION CHEZ LES PROCARYOTES 31

A. Généralités 31

B. Les étapes de la transcription 36

III. LA REGULATION DE LA TRANSCRIPTION 38

A. Régulations intrinsèques de la transcription : 38

B. Régulations extrinsèques de la transcription : 42

IV. LES FACTEURS DE TRANSCRIPTION DE *STAPHYLOCOCCUS AUREUS* 45

A. Les systèmes à deux composants 45

B. Les facteurs de transcription 48

V. REGULATION TRANSCRIPTIONNELLE D'ARN REGULATEURS 57

CONTEXTE DE MA THESE 59

RESULTATS ET DISCUSSIONS 61

PARTIE I : 63

RESULTATS ET DISCUSSIONS 65

I. CARACTERISATION DE L'ARN SRN_3610_SPRC 65

A. Localisation de srn_3610_sprC 65

B. Séquence de srn_3610_sprC et de son promoteur 66

C. Prédictions de structures secondaire de l'ARN 67

II. RECHERCHE DE REGULATEURS DE LA TRANSCRIPTION DE *SRN_3610_SPRC* 67

A. Expression de srn_3610_sprC dans différents contextes mutants 68

B. Action de SarA sur le promoteur de Srn_3610_SprC 71

C. Lien entre SarA et le promoteur de srn_3610_sprC 75

III. IDENTIFICATION D'UN ARN REGULATEUR : SRN_9340 80

A. Recherche bio-informatique 80

B. Caractérisation de Srn_9340 ..... *82*
C. Séquence de srn_9340 et de son promoteur *84*
D. Structures des transcrits de Srn_9340 ..... *86*
IV. RECHERCHE DE REGULATEURS DE LA TRANSCRIPTION DE *SRN_9340* 86
A. Expression de srn_9340 dans différents contextes mutants *86*
B. Action de SarA sur la transcription de srn_9340 ..... *89*
C. Fixation de SarA sur le promoteur de srn_9340 ..... *92*
D. Détermination de la zone de fixation de SarA sur srn_9340 ..... *94*
E. Analyse de la séquence de srn_9340 fixée par SarA *95*
V. ETUDE DE LA FIXATION DE SARA *IN VIVO* 98
A. SarA fixe les promoteurs de srn_3610_sprC et srn_9340 in vivo *98*
B. SarA empêche la fixation de l'ARN polymérase in vivo *100*
C. L'ARN polymérase et SarA se fixent sur les promoteurs de srn_3610_sprC et srn_9340 de manière mutuellement exclusive *102*
D. SarA et la transcription in vitro de srn_3610_sprC et srn_9340 ..... *104*
VI. SCHEMA GENERAL DE LA REGULATION DE L'EXPRESSION DE *SRN_3610_SPRC* ET *SRN_9340* 105

PARTIE II : **107**

RESULTATS ET DISCUSSIONS **109**

I. ETUDE DE LA LEVEE DE REPRESSION DE SARA 109
A. Expression de srn_3610_sprC et srn_9340 au cours du temps *109*
B. Utilisation du gène rapporteur gfp, témoin de l'expression de srn_3610_sprC et srn_9340 ..... *110*
C. Expression de P_{sprC} dans différents milieux de culture *113*
D. Expression de P_{sprC} à différentes températures *115*
E. Expression de P_{sprC} en condition de stress osmotique *117*
F. Expression de P_{sprC} dans des conditions favorisant le biofilm *119*
G. Expression de P_{sprC} lors d'un stress d'autolyse *120*
H. Expression de P_{srn_9340} lors d'un stress d'autolyse *124*
II. DETERMINATION DE CIBLES 126
A. Construction de vecteur de surexpression de srn_3610_sprC et srn_9340 ..... *126*

PARTIE III : **131**

RESULTATS ET DISCUSSIONS **133**

I. RECHERCHE DE SEQUENCES CONSENSUS DE LA FIXATION DE SARA A PARTIR DE CIBLES IDENTIFIEES 133

II. UTILISATION DES SEQUENCES CONSENSUS POUR LA RECHERCHE DE CIBLES 141
III. DETERMINATION DE CIBLES DE SARA PAR UNE METHODE A HAUT DEBIT 147
IV. SITE DE FIXATION DE SARA ET REGULATION 148
V. DISCUSSIONS ET MISE EN PLACE DE MODELE DE REGULATION 151

CONCLUSION ET PERSPECTIVES 155

MATERIELS ET METHODES 167

I. SOUCHES BACTERIENNES ET CONDITIONS DE CULTURES 169
A. Conditions de cultures 169
B. Purification d'ADN plasmidique et Transformations 169
II. ETUDES DES ARN 170
A. Extraction d'ARN 170
B. Détermination des extrémités : RACE (Rapid Amplification cDNA ends) 170
C. Northern Blot 171
D. RT-PCRq 171
E. Transcription in vitro 172
III. ETUDES DES PROTEINES 172
A. Extraction de protéines 172
B. Western Blot 172
C. Expérience de gènes rapporteurs : ß-lactamase 173
D. Expérience de gènes rapporteurs : Green Fluorescent Protéin (GFP) 173
IV. ETUDES DE L'INTERACTION PROTEINES-ADN 173
A. Production de SarA purifiée 173
B. Retards sur gel 174
1. Préparation des sondes 174
2. Réaction de fixation 174
C. Empreinte à la DNase 175
1. Préparation des sondes 175
2. Réaction de fixation 175
D. Immunoprécipitation de la chromatine : ChIP 176

ANNEXES 185

REFERENCES 220

LISTE DES FIGURES

Figure 1 : Morphologie de *S. aureus*.....18
Figure 2 : Localisation des zones de portages de *S. aureus* chez l'Homme. 18
Figure 3 : Localisation des infections dues à *S. aureus*. 19
Figure 4 : Principaux modes d'action des antibiotiques 20
Figure 5 : Acquisition du génome accessoire par transfert horizontaux. 23
Figure 6 : Représentation circulaire des quatre génomes de *S. aureus* 25
Figure 7 : Les différents types de l'îlot génomique vSaβ.....26
Figure 8 : Relations génétiques entre les différentes souches apparentées à NCTC8325....26
Figure 9 : Exemple de mécanismes d'action des ARN régulateurs.....28
Figure 10 : Les différentes classes de facteurs σ.....34
Figure 11 : Composition d'un promoteur bactérien. 36
Figure 12 : Positionnement de l'ARN polymérase sur la région promotrice.....36
Figure 13 Les différents étapes de l'initiation de la transcription. 37
Figure 14 : Différents types de signaux de terminaison de la transcription.....38
Figure 15 : Interférence transcriptionnelle par l'organisation chromosomique des promoteurs 40
Figure 16 : Régulation protéique du facteur σ^B par RsbUVW. 41
Figure 17 : Régulation transcriptionnelle en cas de carence nutritif. 45
Figure 18 : Comparaison des facteurs de transcription de la souche USA300-FPR avec d'autres souches de *S. aureus* (présence du gène en vert, absence en rouge).....46
Figure 19 : Implication des TCS dans différentes fonctions cellulaires.....47
Figure 20 : Mécanisme du système de « quorum-sensing » par *agr*.....48
Figure 21 : Protéines membres de la famille SarA.....49
Figure 22 : Présence des protéines dans la souche Newman selon la phase de croissance de la bactérie.....49
Figure 23 : Différents mécanismes d'action des facteurs de transcription de la famille SarA 50
Figure 24 : Inter-régulations des membres de la famille SarA.....51
Figure 25 : Organisation génétique du gène *sarA* sous contrôle de ses trois promoteurs.....52
Figure 26 : Localisation génomique du gène *srn_3610_sprC* dans la souche *S. aureus* NCTC8325. 65
Figure 27 : Représentation des séquences de *srn_3610_sprC* des 79 souches de *S. aureus*.66
Figure 28: Séquences nucléotidiques du promoteur (P_{sprC}) et du gène *srn_3610_sprC*.....67
Figure 29 : Structure de l'ARN Srn_3610_SprC.....67
Figure 30 : Expression du gène *srn_3610_sprC* dans différentes souches mutantes pour l'un des facteurs de transcription de la famille SarA. 68

Figure 31 : Expression du gène *srn_3610_sprC* dans les souches mutantes HG003Δ*arnIII* ou HG003Δ*sigB*. 69
Figure 32 : Quantité d'ARN Srn_3610_SprC dans les souches doublement mutées pour *sarA* et un autre facteur de transcription. 70
Figure 33 : Expression de *srn_3610_sprC* en absence de SarA et/ou l'ARNIII. 71
Figure 34 : Récapitulatif des combinaisons plasmidiques transformées dans les souches HG003 et HG003Δ*sarA* pour l'expérience de complémentation fonctionnelle. 72
Figure 35 : Expression de *srn_3610_sprC* et *sarA* dans les différentes souches utilisées pour les expériences de complémentation. 72
Figure 36 : Protocole de l'expérience de gène rapporteur (ß-lactamase) 74
Figure 37 : Mesure de l'activité du promoteur P_{sprC} en fonction de la quantité de protéine SarA présente. 74
Figure 38 : SarA est capable de fixer *in vitro* le promoteur de *srn_3610_sprC* (P_{sprC}). 75
Figure 39 : SarA chevauche la promoteur et le gène de *srn_3610_sprC*. 76
Figure 40 : SarA fixe *in vitro* la sonde $sprC_{P267}$. 77
Figure 41 : SarA fixe *in vitro* la sonde $sprC_{P267}$ de manière spécifique.. 78
Figure 42 : la séquence $sprC_{47}$ est nécessaire à la fixation de SarA. 78
Figure 43 : la sonde $sprC_{47}$ est suffisante à la fixation de SarA. 79
Figure 44 : Identification de 22 nucléotides fixés par SarA. 79
Figure 45 : Les 22 nt sont nécessaires à la fixation de SarA. 80
Figure 46 : Localisation du résultat n° 2 (A) et n°5 (B) du BLASTN. 82
Figure 47 : *srn_9340* produit deux transcrits. 83
Figure 48 : Détermination des extrémités 5' et 3' des deux transcrits de *srn_9340*. 83
Figure 49 : Localisation génomique de *srn_9340*. 83
Figure 50 : Séquences nucléotidique du promoteur (P_{srn_9340}) et du gène *srn_9340*. 84
Figure 51 : Séquence de *srn_9340* de la souche *S. aureus* NCTC 8325. 84
Figure 52 : Structure des deux transcrit de *srn_9340* prédites par Mfold. 86
Figure 53 : Expression de *srn_9340* dans différents contextes mutants pour l'un des facteurs de transcription de la famille SarA.. 87
Figure 54 : Expression de *srn_9340* en absence d'expression de l'ARNIII. 88
Figure 55 : Expression de *srn_9340* dans des souches ne produisant ni SarA ni l'un des autres facteurs de transcription étudiés. 88
Figure 56 : Expression de *srn_9340* dans les différentes souches utilisées lors des expériences de complémentation fonctionnelle de SarA. 90
Figure 57 : Activité du promoteur P_{srn_9340} en absence de SarA. 91
Figure 58 : Détermination de la demi-vie des ARN Srn_3610_SprC et Srn_9340 dans la souche *S. aureus* Newman. 91

Figure 59 : SarA fixe la sonde *srn_9340*P225 *in vitro*. 92
Figure 60 : SarA fixe de manière spécifique la sonde *srn_9340*$_{P225}$ *in vitro*. 94
Figure 61 : SarA fixe le promoteur de *srn_9340*. 94
Figure 62 : Les 51 nucléotides sont nécessaires et suffisants à la fixation de SarA sur le promoteur de *srn_9340*. 96
Figure 63 : Identification de 23 nucléotides fixés par SarA. 96
Figure 64 : Les 23 nt sont nécessaires à la fixation de SarA. 97
Figure 65 : Protocole d'immunoprécipitation de la chromatine. 99
Figure 66 : SarA fixe *in vivo* les promoteurs de *srn_3610_sprC, srn_9340* et *hla*. 99
Figure 67 : SarA gêne la fixation de l'ARN polymérase pour le promoteur *srn_3610_sprC*. 100
Figure 68 : Représentation des différentes possibilités de fixation de SarA ou de l'ARN polymérase sur le promoteur de *srn_9340*. 101
Figure 69 : L'ARN polymérase d'*E. coli* est capable de fixer les promoteurs de *srn_3610_sprC* (A) et *srn_9340* (B). 102
Figure 70 : SarA et l'ARN polymérase fixent de manière mutuellement exclusive les promoteurs de *srn_3610_sprC* et *srn_9340*. 103
Figure 71 : SarA réprime la transcription *in vitro* de *srn_3610_sprC* et *srn_9340*. 104
Figure 72 : Interrégulation de *srn_3610_sprC* et *srn_9340* par SarA, SarR et ARNIII. 105
Figure 73 : Schéma général de la répression de la transcription de *srn_3610_sprC* et *srn_9340* par le facteur de transcription SarA. 106
Figure 74 : Expression de *srn_3610_sprC* (A) et *srn_9340* (B) dans les souches HG003 et HG003Δ*sarA* au cours du temps. 109
Figure 75 : Représentation des trois plasmides utilisés pour l'expérience de dgène rapporteur *gfp*. 110
Figure 76 : Protocole d'expérience de gène rapporteur *gfp*. 111
Figure 77 : Expression des promoteurs de P_{sprC} et de P_{srn_9340} dans la souche HG003 et de HG003ΔsarA. 111
Figure 78 : Représentation graphique des hypothèses de régulation transcriptionnelle de *srn_3610_sprC* et *srn_9340*. 112
Figure 79 : Protocole d'expérience de gène *gfp* dans différents milieux de culture 113
Figure 80 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* dans différents milieux de culture. 114
Figure 81 : Protocole d'expérience de gène rapporteur *gfp* à différentes températures de culture. 115
Figure 82 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* à différentes températures de croissance (20°C, 37°C et 42°C). 116
Figure 83 : Protocole d'expérience de gène *gfp* en condition de stress osmotique 117

Figure 84 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* en condition de stress osmotique (+ NaCl 1,5M). 118
Figure 85 : Protocole d'expérience de gène rapporteur *gfp* en condition favorisant la création de biofilm. 119
Figure 86 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* dans les conditions favorisant le biofilm. 120
Figure 87 : Protocole d'expérience de gène rapporteur *gfp* lors d'un stress d'autolyse 121
Figure 88 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. 121
Figure 89 : Protocole d'expérience de gène rapporteur *blaZ* (ß-lactamase) lors d'un stress d'autolyse 122
Figure 90 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. 123
Figure 91 : Activité du promoteur *srn_9340* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse 124
Figure 92 : Activité du promoteur *srn_9340* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse 125
Figure 93 : Séquence du promoteur P_{veg} 126
Figure 94 : Représentation des plasmides utilisés pour l'expérience de surexpression de *srn_3610_sprC* et *srn_9340*. 127
Figure 95 : Validation de la surexpression de *srn_3610_sprC* (A) et *srn_9340* (B) par l'activité du promoteur P_{veg}. 127
Figure 96 : Schéma général des conditions de régulation de *srn_3610_sprC* et *srn_9340* selon (A) si l'expression de ces gènes est réprimée ou (B) si elle est activée. 129
Figure 97 : Détermination d'une région commune entre les séquences consensus de Chan *et al.* et Chien *et al.* 134
Figure 98 : Alignement de la séquence consensus déterminée par Sterba *et al.* et la séquence consensus1. (alignement réalisé par Custal Oméga). 134
Figure 99 : Motif commun (consensus5) entre les différents sites de fixation de SarA de gènes réprimés par la protéine 136
Figure 100 : Motif consensus (consensus6) entre les différents sites de fixation de SarA de gènes activés par la protéine. 137
Figure 101 : Alignements des séquences $sprC_{47}$ (bleue) et srn_9340_{51} (verte) 138
Figure 102 : Alignement de la séquence consensus1 avec les séquences $sprC_{47}$ et srn_9340_{51} 138
Figure 103 : Alignement de la séquence consensus déterminée par Sterba *et al.* avec les séquences $sprC_{47}$ et srn_9340_{51}. 139

Figure 104 : Séquences de *srn_3610_sprC*, *srn_9340* et *sarV* (promoteur et gène). 139
Figure 105 : Motif commun (consensus7) entre les différentes séquences de *srn_3610_sprC*, *srn_9340* et *sarV*. 140
Figure 106 : Séquences de *srn_3610_sprC*, *srn_9340* et *sarV* (promoteur et gène).. 140
Figure 107 : Expression des cibles identifiées par la recherche du motif dans le génome de la souche HG003. 144
Figure 108 : Expression des cibles identifiées par la recherche de la séquence consensus2 dans le génome de la souche HG003.....147
Figure 109 : Séquences de *scpA, atlR* et *saouhsc_02943* (promoteur et gène).....149
Figure 110 : Séquences de *saouhsc_02943, srn_1490* et *srn_4940* (promoteur et gène). .150
Figure 111 : Modèle de répression (A) et d'activation (B) par SarA de l'expression de gènes selon la position du dimère sur le promoteur. 152
Figure 112 : Représentation des différentes perspectives de ce travail de thèse.....158
Figure 113 : Stratégie expérimentale pour la détermination du régulon de SarA 164
Figure 114 : Description des trois étapes de PCR pour la délétion au sein des sondes.....175

LISTE DES TABLEAUX

Tableau 1 : Caractéristiques des MRSA associées aux soins (HA-MRSA) ou aux communautés (CA-MRSA). 22
Tableau 2 : Résultats de la recherche de *sprC*$_{47}$ dans le génome de *S. aureus* NCTC8325.81
Tableau 3 : Liste de souches selon la présence ou l'absence de *srn_3610_sprC* ou *srn_9340*. 85
Tableau 4 : Sites de protection par SarA de gènes dont l'expression est réprimée.....135
Tableau 5 : Sites de protection par SarA de gènes dont l'expression est activée. 136
Tableau 6 : Résultat de la recherche de la séquence consensus4 dans le génome de la souche NCTC8325 141
Tableau 7 : Résultat de la recherche du motif consensus7 dans le génome de la souche NCTC8325 143
Tableau 8 : Résultat de la recherche de la séquence consensus8 dans le génome de la souche NCTC8325. 146
Tableau 9 : Récapitulatif des différentes séquences consensus. 151
Tableau 10 : Liste des souches utilisées pour l'étude 177
Tableau 11 : Liste des oligonucléotides utilisés pour l'étude.....179
Tableau 12 : Liste des plasmides utilisés pour l'étude 182

LISTE DES ABREVIATIONS

A : Adénine

ADN : Acide désoxyribonucléique

Agr : Accessory gene regulator

AIP : auto-inducing peptide

ARN : Acide ribonucléique

ARNm : ARN messager

ARNrég : ARN régulateur

ARNt : ARN de transfert

ARNtm : ARN transfert-message

Atl : Autolysine

BHI : Brain heart infusion

BSA : Bovine Serum Albumine

C : Cytosine

CSPD : Disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.13,7]decan}-4-yl) phenyl phosphate

C-terminale : carboxy-terminale

DIG : digoxigénine

DO : densité optique

E. coli : Escherichia coli

ECF : fontion extracytoplasmique

G : Guanine

HRP : peroxidase de raifort

Kb : kilo bases

K_D : constante de dissociation

LB : Luria Broth

nM : nano molaire

nm : nano mètre

Nt : nucléotide

NTP : Nucléoside triphosphate

N-terminale : amino-terminale

Pb ; paire de bases

PCRq : Quantitative polymerase chain reaction

PNAG : poly-*N*-acteyl ß-1-6-glucosamine

RACE : Rapid amplification of cDNA ends

RBS : Ribosome binding site

RNase : Ribonucléase

Rut : Rho utilization site

RV : Région variable

Sae : staphylococcal accessory element locus

Sar : Staphylococcal accessory regulator

SARM : *S. aureus* résistant à la méthicilline

S. aureus : Staphylococcus aureus

SCC : Staphyloccocal chromosome cassette

SDS : Sodium dodecyl sulfate

Spr : Small pathogenicity island RNA

T : Thymine

TBE : Tris-Borate-EDTA

TCS : Système à deux composants

TSB : Tryptic soy broth

UA : Unité arbitraire

µM : micro molaire

µm : micro mètre

°C : degré Celsius

6His : étiquette 6 Histidine

INTRODUCTION

INTRODUCTION

I. *Staphylococcus aureus*

A. Généralités

1. Mise en évidence de *Staphylococcus aureus*

L'attribution de la mise en évidence de *Staphylococcus aureus* (*S. aureus*) est légèrement différente selon le côté de la Manche où l'on se situe. A qui d'Alexander Ogston ou de Louis Pasteur attribuer la paternité de la découverte des staphylocoques ? Pour répondre à cette question, retournons 130 ans en arrière. Avec l'émergence des techniques de culture et d'observation de micro-organismes, Louis Pasteur isola, à partir de pus de furoncles, une bactérie capable de s'agréger en amas (1880). Parallèlement, Alexander Ogston identifia dans des abcès, des bactéries qui s'agençaient en « grappes de raisins » (1881). Il nomma le genre *Staphylococcus* (du grec *staphyle* : grain de raisin et *kokkos* : coque) (Newsom, 2008). En 1884, Friedrich Julius Rosenbach distingua deux types de staphylocoques selon leur couleur. L'un est blanc (*Staphylococcus pyogenes albus* nommé aujourd'hui *S. epidermidis*) et l'autre de couleur doré (*S. aureus)*. Dès lors, l'intérêt pour ce micro-organisme issu d'infections humaines n'a cessé de croître.

2. Place de *S. aureus* dans le monde du vivant

S. aureus appartient à la division des Firmicutes dans le règne des Procaryotes, cette division regroupant principalement des bactéries positives à coloration de Gram (Gram +). Ce pathogène fait partie de la classe des Bacilli, de l'ordre des Bacillales et enfin de la famille *Staphylococcaceae*. Cette famille est composée de 5 genres dont *Staphylococcus* est le genre majoritaire. A l'heure actuelle, il a été répertorié plus d'une cinquantaine d'espèces de *Staphylococcus* dont la plupart est portée par les membres du règne Animal. Chez l'Homme, les espèces les plus couramment retrouvées sont *S. saprophyticus, S. epidermidis* et *S. aureus*, tous trois responsables d'infections.

3. Caractéristiques de *S. aureus*

S. aureus est une bactérie à Gram positif anaérobie facultative dont la taille est d'environ 1µm (Figure 1B). Comme observées initialement, les bactéries se regroupent en amas de type « grappes de raisins » (Figure 1A) et sur boîtes de culture elles se présentent en colonie lisse. Les staphylocoques dorés sont caractérisés par une couleur orangée laquelle est le résultat de la production de pigments caroténoïdes. Le temps de doublement de *S. aureus* est d'environ 20 minutes à 37°C. Ces bactéries sont résistantes, elles peuvent croître dans une large gamme de pH (4,5 - 9,5), à des températures extrêmes (de 15°C à

45°C) et dans des milieux dont la concentration en NaCl peut aller jusqu'à 10%. Pour autant, les staphylocoques dorés ne sont pas connus pour pouvoir se mettre sous la forme de spore (moyen de résistance à des conditions extrêmes). Néanmoins, leur grande adaptabilité leur a permis de compter parmi les plus grands pathogènes de l'Homme.

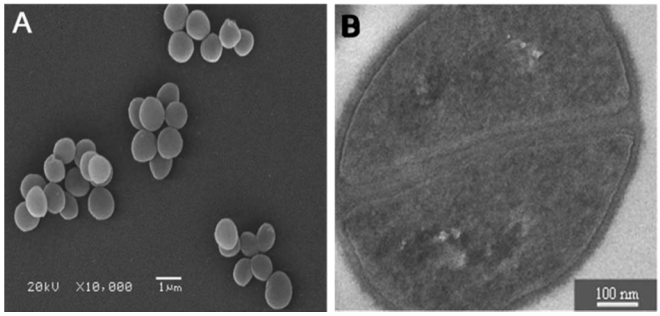


Figure 1 : Morphologie de *S. aureus*.
(A) Visualisation des amas en « grappes de raisins » au microscope électronique à balayage (x10 000).
(B) Visualisation du contact de bactéries de *S. aureus* au microscope électronique en transmission (x70 000).
D'après (Zhang et al., 2013)

B. Infections à *S. aureus*

1. *S. aureus*, bactérie commensale et pathogène

Le staphylocoque doré est une bactérie commensale de l'Homme et des animaux. Ainsi, il est estimé que 30% de la population humaine est porteuse permanente de la bactérie (Kluytmans et al., 1997). Comme présentés dans la Figure 2, les sites majeurs de portage de *S. aureus* sont les fosses nasales antérieures, les aisselles et les mains (Wertheim et al., 2005).

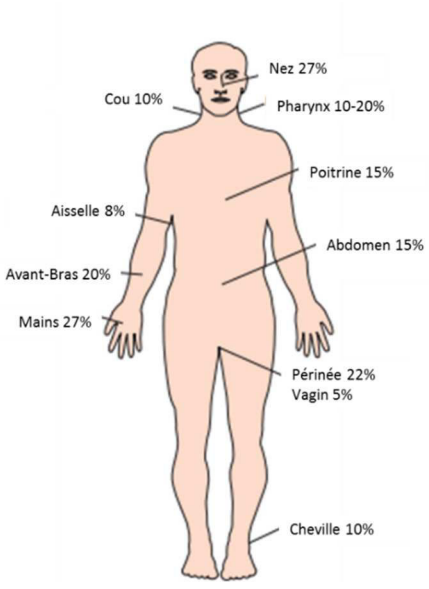


Figure 2 : Localisation des zones de portages de *S. aureus* chez l'Homme.
D'après (Wertheim et al., 2005)

Bien avant la caractérisation de *S. aureus*, des infections pouvant être attribuées à cette bactérie ont été décrites à travers les siècles. Par exemple, la sixième plaie d'Egypte (apparition de furoncles incurables visant les Hommes et le bétail), rapportée dans la Bible, pourrait être une infection massive à *S. aureus*. Depuis, de nombreuses infections aux staphylocoques dorés ont été répertoriées dessinant le portrait d'un pathogène redoutable.

Les infections à *S. aureus* sont diversement localisées dans l'organisme et de gravité variable. Les infections de la peau et des tissus mous regroupent des infections fréquentes, généralement bégnines. Parmi celles-ci, nous avons l'impétigo, maladie la plus retrouvée chez les enfants avec une prédominance de *S. aureus* allant jusqu'à 81% ou les furoncles dont le pus a permis l'identification de *S. aureus*. D'autres infections plus graves peuvent être dues dans une moindre mesure aux staphylocoques dorés (cellulites non purulentes, fasciite nécrosante, pyomyosite, etc.) (Tong et al., 2015).

Dans les pays industrialisés, *S. aureus* est la principale cause des endocardites. Les staphylocoques dorés sont aussi impliqués dans des infections pulmonaires et osseuses (Tong et al., 2015).

L'ensemble des infections présentées peut être la porte d'entrée d'une infection plus globale de l'organisme : la bactériémie. La bactériémie correspond à la présence des bactéries dans le système sanguin et peut engendrer des infections de grandes gravités (syndrome de choc toxique, sepsis ou choc septique) (Tong et al., 2015).

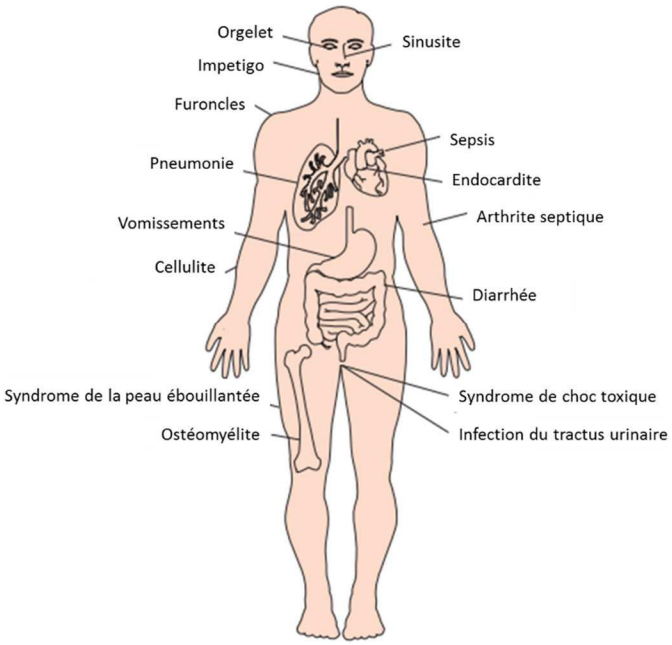


Figure 3 : Localisation des infections dues à *S. aureus*.
D'après (Wertheim et al., 2005)

6% des hospitalisations sont suivies par le développement d'infections nosocomiales. Trois bactéries en sont majoriatirement responsables (*Escherichia coli* : 26% ; *S. aureus* : 16% ; *Pseudomonas aeruginosa* : 8,4%) (données INSERM [lien internet 1]). Selon l'Organisation Modiale de la Santé, il a été recensé en Europe plus de 4 millions de patients atteints par les infections nosocomiales dont 37000 a entraîné la mort. *S. aureus* est donc devenu un problème majeur de santé publique (données OMS [lien internet 2]). Nous n'en parlerons pas ici mais *S. aureus* est également un pathogène de l'Animal provoquant de nombreuses infections qui sont coûteuses aux filières agricoles.

L'ensemble de ces données démontre la nécessité de développer des stratégies pour lutter contre ces infections.

2. Luttes contre les infections à *S. aureus*

De façon générale les stratégies contre les maladies infectieuses reposent sur un trépied constitué par les mesures d'hygiène, l'usage de vaccins et l'emploi des médicaments anti-infectieux pour le traitement. Concernant *S. aureus*, s'il n'existe aucune démarche vaccinale efficace, par contre des mesures d'hygiène individuelles ou collectives peuvent être mises en place en milieu communautaire ou dans un environnement hospitalier. Enfin, les antibiotiques, d'origine naturelle ou synthétique, restent à ce jour le meilleur moyen pour détruire la bactérie (bactéricide) ou pour limiter l'infection (bactériostatique). Ils existent cinq grandes familles d'antibiotiques classées selon leur mode d'action. Ils peuvent agir sur la réplication de l'ADN (fluoroquinolones, novobiocine), la transcription (rifamycines) ou encore sur la voie de synthèse des folates (sulfamides et trimethoprime). Les deux sites privilégiés d'action des antibiotiques sont la synthèse protéique (tétracycline, aminosides) et la synthèse de la paroi (pénicilline, glycopeptides).

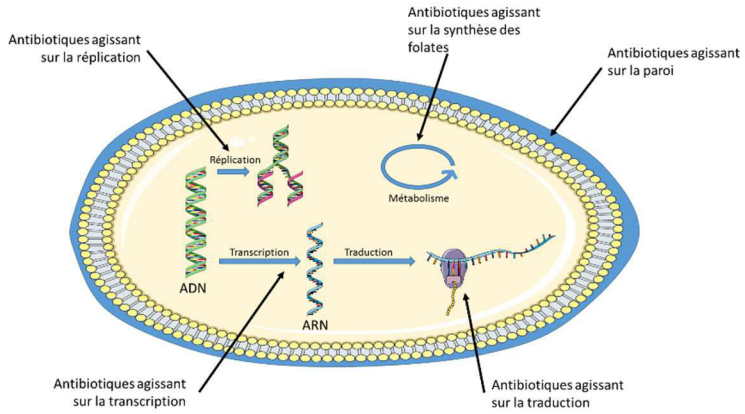


Figure 4 : Principaux modes d'action des antibiotiques

D'autres thérapies ont été utilisées pour lutter contre les microorganismes telles que la phagothérapie. Ils existent, comme chez les Eucaryotes, des virus spécifiques des Procaryotes : les bactériophages. Le cycle d'infection complet des bactériophages entraîne la lyse bactérienne et la libération de nouveaux phages. Ce mécanisme est utilisé à la fois comme outil moléculaire et stratégie thérapeutique. Néanmoins, cette thérapie a certains inconvénients. L'intégralité des bactéries ne peut pas être éliminée et surtout les bactériophages peuvent porter des gènes qui pourraient rendre la bactérie plus virulente (Kropinski, 2006). De plus, des freins psychologiques existent pour l'ingestion d'organismes viraux actifs pour éradiquer des infections bactériennes : cette thérapie reste donc encore peu utilisée.

De nos jours, de nouvelles thérapies se mettent en place en utilisant des « armes » venant directement du monde bactérien. Il a été montré que certains organismes produisent un système de défense contre les microorganismes : les peptides antimicrobiens. Ces peptides forment des pores au niveau de la paroi bactérienne. Notre laboratoire étudie l'un de ses peptides produits par *S. aureus* : PepA1 (30 acides aminés). La production de ce peptide entraîne la lyse de *S. aureus*, des bactéries et des cellules humaines environnantes. Une nouvelle stratégie a alors été imaginée pour modifier ce peptide afin de retirer l'effet hémolytique tout en gardant l'effet bactéricide (Solecki et al., 2015).

3. Résistances aux antibiotiques

Dès l'utilisation massive d'antibiotique, lors de la 2nde Guerre Mondiale, des systèmes de résistances à ces antibiotiques ont été mis en place par les bactéries. Par exemple, la pénicilline G (découverte par Alexander Fleming en 1928) fut administrée massivement en 1940. Deux années plus tard, des souches de *S. aureus* résistantes à cet antibiotique émergèrent. De plus, après les premières utilisations de la méticilline (1960), il ne fallut qu'une année pour que la bactérie parvienne à produire la protéine PBP2a, capable d'inhiber l'action de cet antibiotique. Des souches devenues résistantes à la méticilline (SARM) après avoir acquis la cassette génomique SCCmec se retrouvent partout dans le monde. Parmi ces souches SARM, différentes catégories peuvent être distinguées selon le comportement épidémiologique. Nous avons ainsi les HA-SARM (hospital associated : issu du monde hospitalier), les CA-SARM (community associated : transmis au sein d'une communauté) et les LA-SARM (livestock associated : dans un élevage) (Lindsay, 2014).

Les souches HA-SARM sont prédominantes dans les pays industrialisés et ont une faculté de colonisation importante. Ils touchent principalement les personnes âgées, immunodéprimées ou déjà atteintes d'infection. Les points de dissémination des souches CA-SARM sont beaucoup plus nombreux. La proximité des personnes dans certains lieux favorise les échanges de microorganismes. Parmi ceux-ci, nous retrouvons les casernes militaires, les prisons, les crèches, etc. Les CA-SARM atteignent majoritairement les personnes jeunes provoquant principalement des infections cutanées. Les différences entre les HA- et les CA-SARM ne sont pas qu'épidémiologiques. Il a été montré qu'au point de vue génomique les HA- et les CA-SARM ne portent pas toujours les mêmes cassettes SCCmec (porteur de gène *mecA* produisant la protéine PBP2a contre les antibiotiques de la famille ß-lactames). Les différences globales sont recensées dans le tableau suivant.

Caractéristiques	HA-MRSA	CA-MRSA
Lieu	Hôpital, maison de repos	Gymnase, prison, caserne, école, garderie…
Patients	Personnes âgées	Enfants et jeunes adultes
Facteurs de risque	Conditions sous-jacentes, transfert à l'hôpital, dispositifs médicaux	Manque d'hygiène, blessures cutanées
Maladies	Bactériémie, pneumonie, infections des tissus mous et la peau, infections du tractus urinaire	Infections des tissus mous et la peau, pneumonie nécrosante, ostéomyélite
Cassette SCCmec	Types I, II, III, IV	Types IV et V

Tableau 1 : Caractéristiques des MRSA associées aux soins (HA-MRSA) ou aux communautés (CA-MRSA). D'après Mertens *et al.,* 2012.

C. Génétiques de *S. aureus*

1. Caractéristiques génétiques

En 2001, les génomes des souches de *S. aureus* N315 et Mu50 ont été entièrement séquencés (Kuroda et al., 2001). Depuis, le séquençage génomique de nombreuses souches de *S. aureus* a permis d'identifier les zones conservées à travers les souches (appelées core génome) et les parties variables (génome accessoire). Les génomes de *S. aureus* ont une taille avoisinant 2,9 millions de paires de bases formant un chromosome unique circulaire. Ils font partie des génomes dits « AT riches », c'est-à-dire que la composition nucléotidique est majoritairement constituée des bases adénines et thymines (au détriment donc des guanines et cytosine). Le pourcentage de guanine et cytosine (%GC) est aux alentours de 33%.

a. Le core génome

Le core génome représente les régions communes à travers les différentes souches de *S. aureus*. Il regroupe environ 1500 gènes nécessaires, entre autres, aux différents métabolismes, à la structure, à la survie de la bactérie ainsi que les gènes codant pour les régulateurs de l'expression génique (Bosi et al., 2016). Il représente environ 75% du génome globale avec un %GC proche du génome complet (Lindsay and Holden, 2004).

b. Le génome accessoire

Les régions génomiques non partagées dans l'ensemble des souches sont appelées génome accessoire. Ces éléments accessoires, issus d'une acquisition antérieure, peuvent être de nature différente (prophages, plasmides, îlots de pathogénie). Le génome accessoire diffère du core génome par le %GC et par les codons utilisés pour la traduction. Les protéines produites dans le génome accessoire vont avoir un rôle dans la résistance aux antibiotiques, la production de facteurs de virulence (toxines, adhésines) mais également dans la survie face aux stress environnementaux. Le génome accessoire est composé d'éléments de taille (10 à 200kb) et localisation variables (intra- ou extra-chromosomiques) (Moon et al., 2016).

- Acquisition du génome accessoire

S. aureus peut acquérir un génome accessoire soit par transfert vertical (de cellule mère aux cellules filles) soit par transfert horizontal. Le transfert horizontal de gènes est de trois types. La transformation consiste en l'acquisition d'ADN libre dans l'environnement de la bactérie. La deuxième possibilité est la transduction, correspondant à l'apport d'information génétique par le biais de bactériophages. Enfin, les bactéries peuvent échanger après contact et par conjugaison des éléments accessoires du génome.

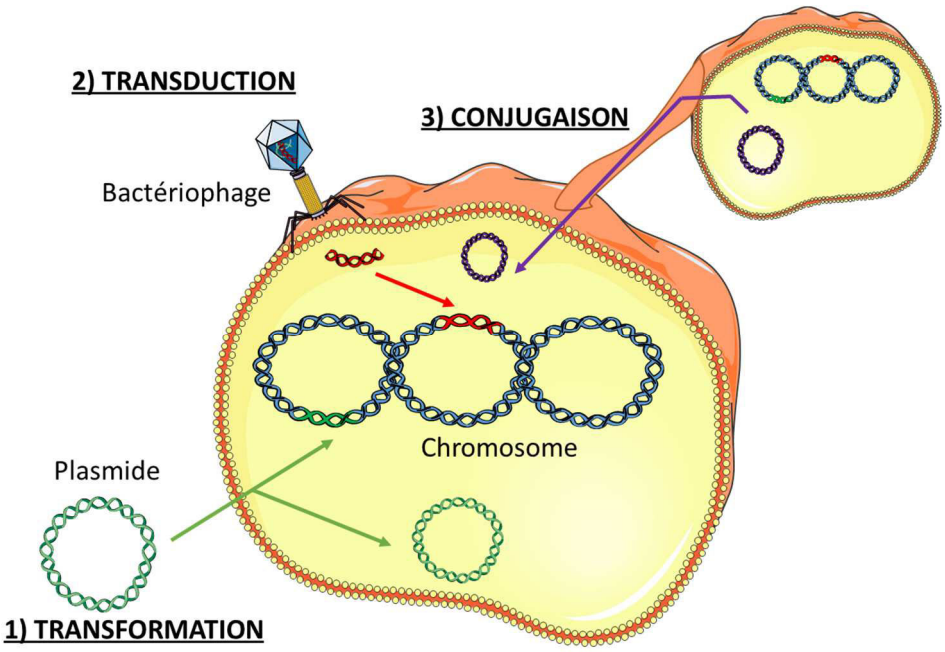


Figure 5 : Acquisition du génome accessoire par transfert horizontaux.

La définition des différents éléments du génome accessoire est très variable et il est difficile de convenir d'une définition commune. C'est pourquoi, les définitions choisies ne sont qu'un choix personnel et j'ai essayé de recenser ici les six catégories d'éléments du génome accessoire selon les critères majoritairement retrouvés dans la littérature.

- Plasmides

Les plasmides sont des morceaux d'ADN extra-chromosomiques circulaires de taille variable, capable de se répliquer de manière autonome par rapport au chromosome. L'acquisition de plasmides par *S. aureus* est réalisée principalement par transduction et par conjugaison. En effet, *S. aureus* a une très faible compétence pour recevoir de l'ADN extracellulaire par transformation (Malachowa and DeLeo, 2010). Chez *S. aureus*, les plasmides portent, entre autres, de nombreux gènes de résistances pour divers antibiotiques (Jensen and Lyon, 2009).

- SCC : Staphylococcal cassette chromosome

Les cassettes chromosomiques de *S. aureus* (SCC) sont caractérisées par la présence de gènes *ccrA* et *ccrB* dont les protéines permettent l'intégration de la cassette à proximité de l'origine de réplication du chromosome (Gill et al., 2005). L'intégration de ces cassettes implique que leur réplication est dépendante de celle du chromosome bactérien. La cassette la plus étudiée est SCC*mec* contenant le gène *mecA* produisant la protéine contrant l'action de la méticilline. Selon la structure de la cassette SCC*mec*, huit différents types ont été caractérisés (2009).

- Transposons

Les transposons sont définis comme des éléments mobiles, codant plusieurs protéines, encadrés par des séquences d'insertion nécessaires à la transposition. Chez *S. aureus*, ces éléments transposables sont de petite taille et portent des gènes codant pour des protéines agissant contre les antibiotiques.

- Prophages

Les prophages sont des régions accessoires appartenant au chromosome de *S. aureus*. Venus de l'infection de la bactérie par des bactériophages, l'ADN viral est intégré dans le génome. Retrouvés dans la plupart des souches de *S. aureus*, les prophages portent de nombreux gènes de virulence comme des entérotoxines, leucocidines ou encore la staphylokinase. (Malachowa and DeLeo, 2010). Il est recensé actuellement cinq prophages : ΦSa1, ΦSa2, ΦSa3, ΦSa4 et ΦSa5 (dont la nomenclature n'est pas homogène selon les souches de *S. aureus*). Ces prophages peuvent être induits par application d'un rayonnement ultraviolet entraînant leur retrait du chromosome (Thompson and Hart, 1981).

- Ilots de pathogénie/îlots génomiques

Les îlots de pathogénie et les îlots génomiques sont des régions d'ADN ayant été intégrées dans le chromosome bactérien (Malachowa and DeLeo, 2010). Ils sont caractérisés par leur structure génomique. Ces îlots contiennent des gènes codant pour des intégrases ou transposases, ils sont souvent encadrés par de petites séquences répétées, associés à la présence de gènes transcrits en ARN de transfert (ARNt). En revanche, contrairement aux îlots de pathogénie, les îlots génomiques ne contiennent que des versions tronquées ou inactives de ces gènes. C'est pourquoi, les îlots génomiques sont qualifiés comme des vestiges d'ADN acquis par transfert horizontal et intégrés dans le génome bactérien.

Dix îlots de pathogénie ont été identifiés dont sept sont spécifiques aux souches de *S. aureus* portées par les êtres humains. Ils sont nommés SaPI1, SaPI2 (type I ou II), SaPI3 (type I ou II), SaPI4 et SaPI5. En ce qui concerne les îlots génomiques, trois ont été localisés : vSaα, vSaβ et vSaγ (Gill et al., 2005) (Figure 6).

Les îlots de pathogénie et génomiques portent des gènes codant pour de nombreuses toxines (entérotoxines, toxine de syndrome de choc toxique…), leucocidines ou encore des protéases.

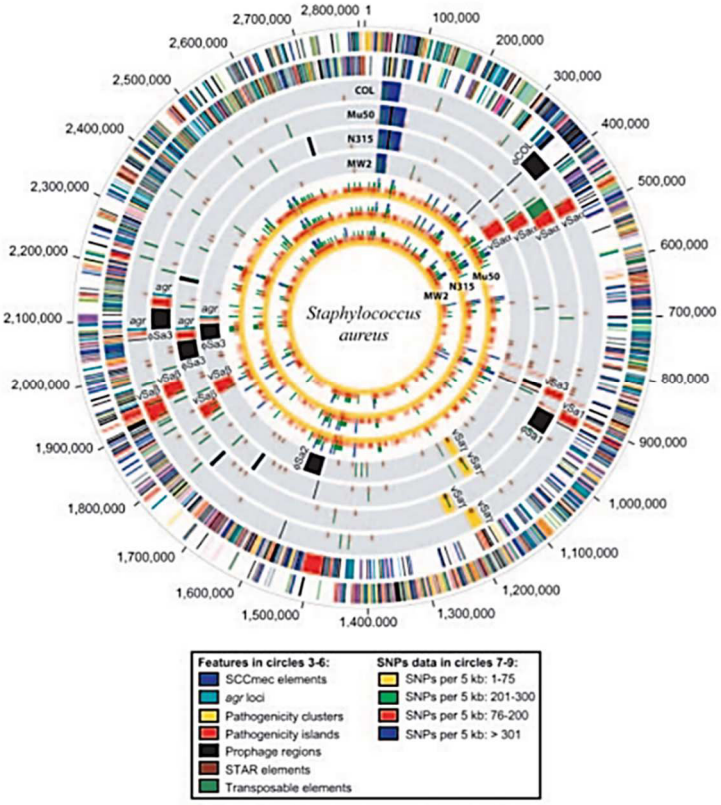


Figure 6 : Représentation circulaire des quatre génomes de *S. aureus* (COL, Mu50, N315 et MW2). (Gill et al., 2005). Les éléments du génome accessoire sont représentés selon un code couleur décrit dans le tableau.

Au cours de ma thèse, j'ai dû m'intéresser particulièrement à l'îlot génomique vSaβ. La structure de cet îlot diffère selon la souche utilisée pour l'étude. Ainsi, trois différents types ont été décrits (vSaβ de type I, II ou III) (Figure 7). Le vSaβ type II diffère du type I par la présence des gènes *bsaABCDPFE* et l'absence en fin d'îlot des gènes codant pour des entérotoxines (*seg, sen, sei,* etc.). Le type III, proche du type I, présente de nombreux gènes supplémentaires en début d'îlot et aussi entre les gènes codant les sérines protéases (*splFEC*) et les entérotoxines (*seg, sen, seu, sei, sem, seo*). La présence de ces derniers implique l'absence des gènes *lukED* (leucocidine).

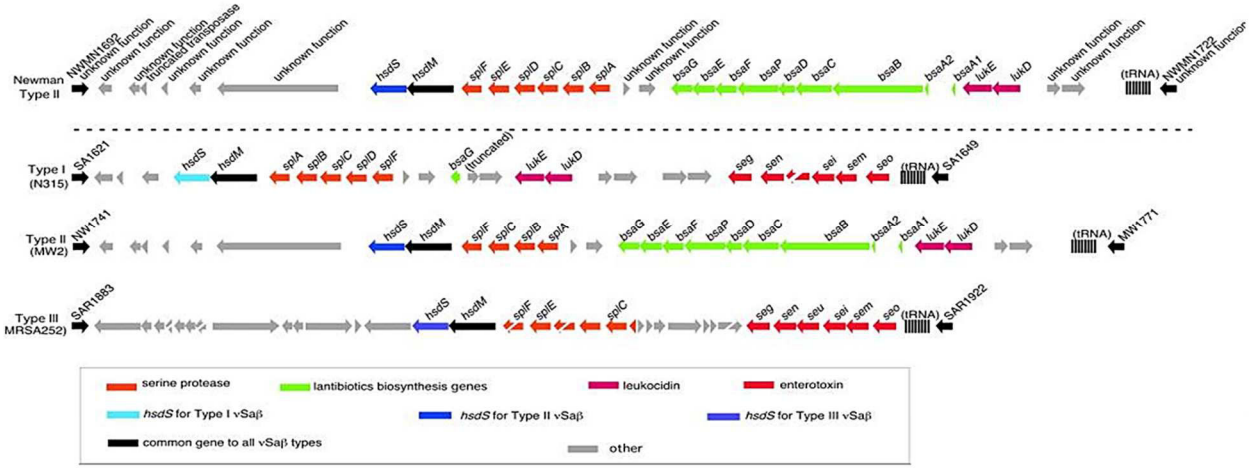


Figure 7 : Les différents types de l'îlot génomique vSaβ. D'après (Baba et al., 2008)

2. La souche de *Staphylococcus aureus* étudiée : HG003

La souche utilisée pour les études présentées dans cette thèse est la souche HG003. Depuis plusieurs années, elle est considérée, dans le monde scientifique, comme la souche idéale pour l'étude de la régulation de l'expression de gènes chez *Staphylococcus aureus*. Dérivée d'une souche clinique de sepsis isolée en 1960 (NCTC8325 ou RN1), la souche HG003 a été obtenue en réparant les gènes *tcaR* et *rsbU* codant respectivement pour le régulateur transcriptionnel TcaR et la protéine RsbU (Herbert et al., 2010). Le génome de la souche HG003 a été séquencé en 2014 (Sassi et al., 2014). Le génome est composé de 2,8 millions de paires de bases et possède un %GC de 32,74%. Le génome partage 99,97% d'identité avec la souche NCTC8325, il est suggéré qu'en dehors de la réintroduction de *tcaR* et *rsbU*, il existe d'autres mutations présentes dans HG003 (Sassi et al., 2014). La souche contient plusieurs éléments du génome accessoire : 3 prophages (Φ11, Φ12 et Φ13), 4 îlots de pathogénies/génomiques (vSaα, vSaß, vSaγ et SaPI2). Enfin, ne faisant pas partie des souches de type SARM, la souche HG003 ne possède pas la cassette SCCmec (Baba et al., 2008).

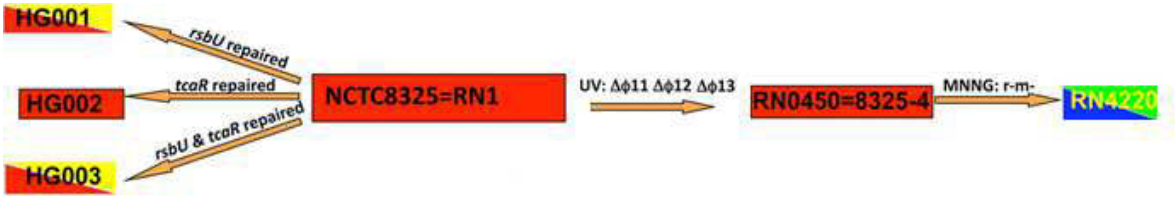


Figure 8 : Relations génétiques entre les différentes souches apparentées à NCTC8325 D'après (Herbert et al., 2010)

3. ARN régulateurs (ARNrég)

Les ARN régulateurs (ARNrég) ont été définis comme des ARN stables, de petites tailles (< 500 nt) et fortement structurés. De par leur séquence nucléotidique et/ou leur structure, les ARNrég ont une activité régulatrice sur d'autres molécules de la bactérie (protéines, ARN).

a. SRD : Staphylococcal Regulatory RNAs Database (Sassi et al., 2015) (ANNEXE 1)

L'intérêt pour les ARN régulateurs a grandi depuis quelques décennies. Afin de les identifier et de comprendre leur rôle, de nombreuses expériences de séquençage à haut débit des ARN (RNA-seq) ont été effectuées et ont engendré la découverte exponentielle de petits ARN. Par contre, elle a entraîné la multiplication des nomenclatures et une redondance de certaines identifications. Au vu de ce constat, le laboratoire a créé une base de données des ARNrég de *Staphylococcus aureus* : SRD. Elle permet d'uniformiser les nomenclatures en regroupant sous un même nom les ARN identifiés dans plusieurs laboratoires et nommés différemment. Ainsi, les ARN sont nommés Srn_XXXX, XXXX étant un nombre correspondant à sa localisation génomique dans la souche référence N315 (de 0000 à 5130). Si l'ARN n'est pas présent dans la souche N315, ce numéro sera supérieur à 9000. Au moment d'écrire ce manuscrit, 215 ARN régulateurs présents chez l'ensemble des souches de *S. aureus* avaient été identifiés.

Depuis maintenant plus de 20 ans, l'étude des ARN régulateurs est prise en compte dans la compréhension générale de la virulence de *S. aureus*. Je vous présenterai les principaux modes d'action des ARNrég, accompagnés de quelques exemples retrouvés chez *S. aureus*.

b. Mécanismes d'action des ARN régulateurs

De manière simplifiée, les ARN régulateurs agissent principalement par appariement de bases avec des ARN cibles. Cette liaison peut avoir plusieurs conséquences : une modification de la stabilité de l'ARN (activation ou inhibition de l'action des ribonucléases (RNase)) ou de la traduction (libération ou séquestration du site de fixation du ribosome (RBS)) (Figure 9).

Les ARN régulateurs ont la possibilité d'agir sur des transcrits codés par un gène localisé à un autre locus génétique, ils agissent alors en *trans*. Au contraire, ils ont une action dite en *cis* si le locus du gène codant l'ARN régulateur est chevauchant à celui de l'ARN cible.

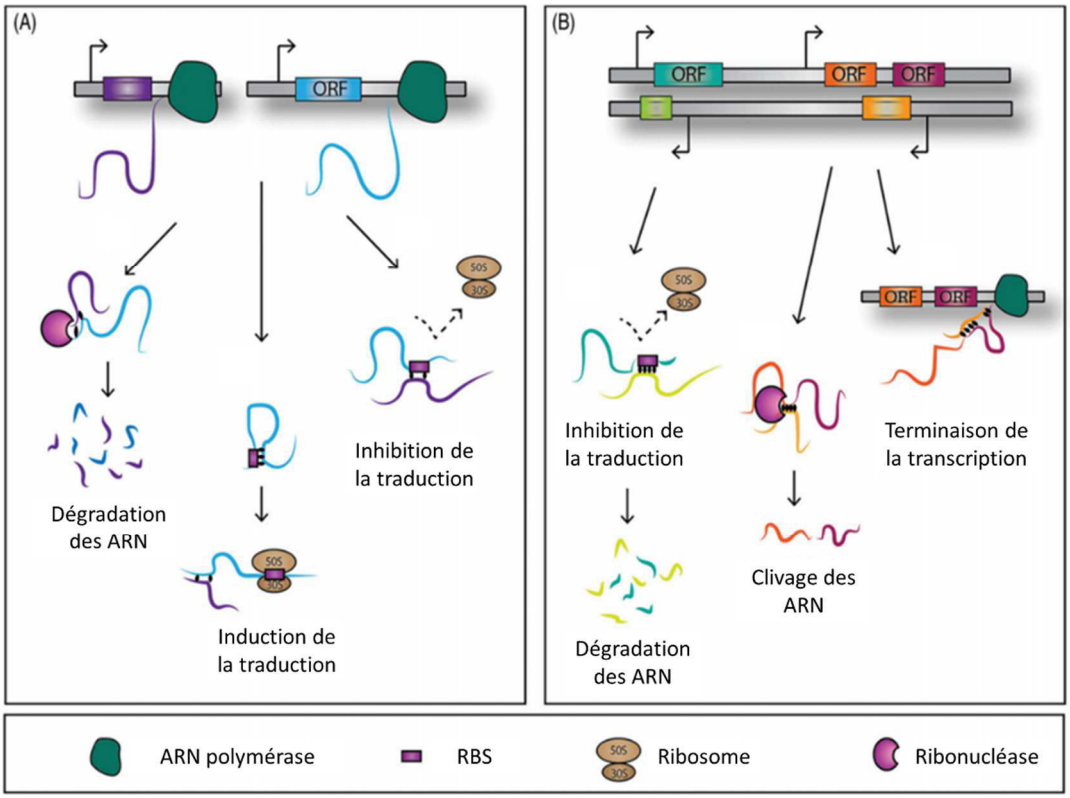


Figure 9 : Exemple de mécanismes d'action des ARN régulateurs codés en *trans* (A) ou en *cis* (B). D'après (Oliva et al., 2015)

- ARN régulateurs de *S. aureus* codés en *trans*
  - L'ARN III (Srn_3910) : effecteur du « quorum sensing »

Révélé en 1993, l'ARNIII est composé de 514 bases et est structuré avec 14 tiges-boucles et 2 longues hélices (Novick et al., 1993). Il possède une partie codante interne pour l'hémolysine δ (codé par le gène *hld*). L'ARNIII réprime, entre autres l'expression de *sbi* (Chabelskaya et al., 2014), *rot* (Boisset et al., 2007; Geisinger et al., 2006), *lytM* (Chunhua et al., 2012), *spa* (Benito et al., 2000; Novick et al., 1993), *ltaS* (Yan et al., 2015) et *coa* (Chevalier et al., 2010). A l'inverse, il active l'expression de *hla* (Morfeldt et al., 1995; Novick et al., 1993), *map (Liu et al., 2011)* et *mgrA* (Gupta et al., 2015). De manière générale, l'ARNIII réprime les facteurs d'adhésion et active la production de protéines nécessaires à l'infection.

  - SprC (Srn_3610)

Le gène *sprC*, présent dans l'îlot de pathogénie vSaβ, est transcrit en un ARN impliqué dans l'atténuation de la virulence et la phagocytose de *S. aureus* par des cellules immunitaires humaines (Le Pabic et al., 2015). La délétion du gène *sprC* entraîne une augmentation de la virulence de la souche de *S. aureus* Newman et une augmentation de la capacité de *S. aureus*

à être phagocytée. La phagocytose de bactéries par des macrophages est initiée par le contact cellule hôte-pathogène. Le Pabic *et al.* ont montré que SprC impacte la phagocytose en agissant sur la protéine autolysine ATL (impliquée dans la division cellulaire, phagocytose et la formation de biofilm). Il a été montré que SprC réprime la traduction de l'ARN *atl*. Néanmoins, des expériences de retards sur gels avec les ARN SprC et *atl* ont montré une très faible affinité entre les deux transcrits *in vitro*. Il est supposé alors que d'autres partenaires (ARN ou protéines) agissent conjointement avec SprC pour effectuer la répression de l'expression de l'autolysine (Le Pabic et al., 2015).

- SprD (Srn_3800)

SprD est impliqué dans la virulence de *S. aureus.* Positionné dans l'îlot de pathogénie Φ13, le gène *sprD* est transcrit en début de phase exponentielle de croissance. Chabelskaya *et al.* ont montré que SprD réprime la traduction de la protéine Sbi, impliquée dans le système d'évasion du système immunitaire (Chabelskaya et al., 2010). Dans un modèle d'infection murin, la délétion du gène *sprD* inhibe la mort contrairement à une souche produisant cet ARNrég. Comme précisé précédemment, l'ARNm *sbi* est également une cible de l'ARNIII. Comme SprD est exprimé en début de phase exponentielle et l'ARNIII en fin de phase exponentielle, la fenêtre d'expression de *sbi* est réduite au milieu de cette phase. Cette co-régulation d'un gène par deux ARN régulateurs fut le premier cas publié dans la littérature (Chabelskaya et al., 2014).

- SprX (Srn_3820)

Contrairement aux autres Spr, SprX a été identifié en 2010 (Bohn et al., 2010). Le nombre de copies de SprX est variable selon les souches de *S. aureus* étudiées. Pour la souche HG001 (souche proche de HG003), deux gènes sont présents dans les îlots de pathogénie Φ13 et Φ12. SprX réprime l'expression du gène *spoVG* (stage V sporulation protein G) dont la protéine est impliquée dans la résistance aux antibiotiques de la famille des glycopeptides (vancomycine et teicoplanine) (Eyraud et al., 2014).

- ARN régulateurs de *S. aureus* codés en *cis*

Parmi ces ARN régulateurs, nous retrouvons les ARN antisens correspondant à la transcription de gènes dont la séquence est chevauchante à celle du gène de l'ARN cible.

- RsaOX (Srn_0230)

Le gène codant RsaOX est positionné en antisens par rapport à la séquence codante d'une transposase hypothétique (SA0062) (Bohn et al., 2010). La ressemblance de cette

organisation avec ce qui est retrouvé chez *E. coli* (l'ARNm de la transposase IS10 est déstabilisé par un ARNrég antisens (Case et al., 1990). Ces similarités supposent qu'un même mécanisme à lieu chez *S. aureus* pour réprimer la production de la transposase SA0062 (Bohn et al., 2010). D'autres ARNrég sont également des ARN antisens de transposases comme Teg17 et Teg24as avec la transposase IS1181 (Beaume et al., 2010).

Parmi les ARNrég codés en *cis,* nous avons les systèmes toxine/antitoxine de type I. Le système toxine/antitoxine de type I correspond à la production d'une toxine dont l'expression du gène est inhibée par un ARN (Wen and Fozo, 2014). Les deux gènes (l'un codant la toxine, l'autre l'antitoxine) sont localisés au même locus génétique.

- SprF1/SprG1 (Srn_3830 / Srn_3840)

Système toxine/antitoxine de type I identifié parmi les Spr, l'ARN SprG1 code pour deux toxines dont l'expression est réprimée par l'antitoxine SprF1. Localisés dans l'îlot de pathogénie Φ13, leurs deux gènes sont positionnés au même locus génétique mais sur des brins opposés (localisation en *cis*). Il a été montré que l'ARN SprG1 est plus stable que SprF1 (120 min vs 10 min). La répression de SprF1 sur SprG1 est présente tout au long de la croissance bactérienne. Néanmoins, dans certaines conditions, il est supposé que l'expression de *sprF1* est éteinte permettant la traduction des peptides. Ces peptides produits par SprG1 sont positionnés à la membrane et entraînent la mort de la bactérie par la formation probable de pores au niveau de celle-ci. Les peptides libérés dans le milieu extracellulaire engendrent également la lyse des autres bactéries et des cellules de l'hôte environnantes (Pinel-Marie et al., 2014).

- ARN régulateurs de *S. aureus* codés en *cis* et agissant en *trans*
  - SprA1/SprA1$_{AS}$ (Srn_3580 / Srn_3590)

Les deux ARN appartiennent à un système toxine/antitoxine de type I. Les deux gènes sont localisés au même locus génétique, c'est-à-dire, que le gène de l'un est l'antisens de l'autre. Néanmoins, l'ARNrég SprA1$_{AS}$ a la particularité d'agir en *trans* sur SprA1 alors que les deux gènes sont localisés en *cis*. Le peptide PepA1, issu de la traduction de SprA1, a une double action : il détruit les érythrocytes humains (action hémolytique) et possède une activité antibactérienne (Sayed et al., 2012). L'étude de ce peptide a permis de créer une série d'analogues de PepA1 qui aurait conservé une action bactéricide mais aurait perdu la capacité hémolytique (Solecki et al., 2015).

L'étude fonctionnelle des ARNrég est un domaine en expansion. Néanmoins, il reste encore difficile aujourd'hui de déterminer les cibles directes et la fonction de ces transcrits. Nous supposons que l'action de ces ARNrég doit être, elle-même, finement régulée et donc

reliée avec l'arrivée de stimuli. Il nous est donc apparu nécessaire d'étudier la régulation de la transcription de ces ARNrég pour mieux comprendre l'expression de leurs gènes en réponse à des signaux extérieurs.

II. La transcription chez les procaryotes

Au milieu du XIX^{ème} siècle, Friedrich Miescher fait part de la découverte de la nucléine, substance issue de noyaux de cellules et dont la composition semblait différente des protéines. Il fallut attendre près d'un siècle pour qu'Oswald Avery, en s'inspirant d'une expérience de Frederick Griffith (1928), définisse cette substance comme un acide désoxyribonucléique (ADN) (1944). Les études qu'il a menées sur la bactérie *Streptococcus pneumoniae* ont initié l'idée que l'ADN soit le support de l'hérédité. Cette hypothèse va à l'encontre du paradigme de l'époque qui mettait les protéines comme le lien avec l'hérédité. Dans ce contexte de méfiance envers le pouvoir héréditaire de l'ADN, Rosalind Franklin a réalisé les premiers clichés aux rayons X de l'ADN (1951). James Dewey Watson et Francis Crick se sont basés sur cette étude pour caractériser la structure en double hélice de l'ADN (1953). Dès lors l'intérêt pour cette molécule n'a cessé d'augmenter et l'ADN fut considéré comme contenant les données nécessaires à la vie et à la transmission de l'information génétique aux générations suivantes. Néanmoins, la question s'est posée de savoir comment l'ADN pouvait produire les protéines considérées alors comme les seules molécules actives des cellules. En 1960, Jacques Monod et François Jacob propose qu'un second acide nucléique, l'acide ribonucléique (ARN) soit le messager entre la molécule d'ADN et la production de protéines. En effet, au cours des années 1950, de nombreux chercheurs se sont intéressés à la synthèse des ARN. Ils ont pu identifier l'ARN polymérase, moteur de cette synthèse, qui nécessite d'avoir un modèle, l'ADN (Hurwitz, 2005). Cette étape-clé de production de « copies » de fragments d'ADN est appelé transcription. Les ARN produits sont soit traduits en protéines soit utilisés tels quels dans différents mécanismes bactériens.

Je vous présenterai ici les généralités de l'étape de transcription chez *S. aureus*, néanmoins lorsque les données bibliographiques seront insuffisantes chez *S. aureus*, j'utiliserai celles des procaryotes en général.

A. Généralités

La transcription est l'un des mécanismes essentiels à la vie. Elle nécessite d'avoir une matrice de type ADN (chromosome ou plasmide), un complexe protéique (l'ARN polymérase) ainsi que les éléments essentiels à la structure des ARN : les nucléotides. Les nucléotides sont composés d'une base azotée (adénine, uracile, cytosine et guanine), d'un sucre (ribose) et d'un groupement phosphate. Quatre nucléotides formant 2 groupes, les purines (adénosine

et guanosine) et les pyrimidines (cytidine et uridine), entrent dans la composition d'une molécule ARN. L'uridine est spécifique à la molécule d'ARN et correspond à un homologue de la thymidine, spécifique de l'ADN.

1. Composants de l'ARN polymérase

Chez les Procaryotes, l'ARN polymérase (holoenzyme) est composée de deux parties : le core-enzyme et la sous-unité σ. Le core-enzyme est lui-même une structure complexe issue de l'association de plusieurs sous-unités dont quatre majeures (2 α, 1 β, 1 β') et plusieurs petites sous-unités annexes (δ, ω et ε) (Keller et al., 2014 ; Minakhin et al., 2001 ; (Murakami and Darst, 2003) ;Weiss et al., 2014 . Mis à part le facteur ε, toutes les autres sous-unités de l'ARN polymérase ont été décrites chez *S. aureus*. Néanmoins, la recherche du facteur ε dans le génome de *S. aureus* tend à montrer que ce facteur pourrait être présent (données personnelles).

- La sous-unité α

Codée par le gène *rpoA*, la sous-unité α est constituée de deux modules (N-Terminal Domain (NTD) et C-Terminal Domain (CTD)) reliés par une région flexible (Blatter et al., 1994; Negishi et al., 1995). Présent en double exemplaire dans le complexe protéique de l'ARN polymérase, les domaines NTD se lient aux sous-unités ß et ß' du core-enzyme. Les modules CTD, quant à eux, permettent la dimérisation des deux sous-unités α et la liaison à l'ADN (Blatter et al., 1994).

- Les sous-unités ß et ß'

De très grandes tailles (>100 kDa), les sous-unités ß et ß' de l'ARN polymérase sont considérées comme le point d'ancrage des autres sous-unités de l'enzyme. Les gènes codant pour ces deux facteurs sont *rpoB* et *rpoC* et sont situés au sein d'un opéron en présence du gène *rplL* (codant pour une protéine ribosomale) (Aboshkiwa et al., 1992).

- La sous-unité δ

Initialement identifié en 1975 chez *B. subtilis*, la sous-unité δ de l'ARN polymérase a été caractérisée chez *S. aureus* en 1998 (Watson et al., 1998). Codée par le gène *rpoE*, la taille de la protéine produite est de 21,5kDa. Des analyses de co-immunoprécipitation avec la sous-unité δ ont montré que la protéine se fixe aux sous-unité ß et ß' de l'ARN polymérase. Elle confirme la présence de la sous-unité δ dans le core-enzyme (Weiss et al., 2014). Chez *S. aureus*, il a été démontré que la délétion du gène *rpoE* entraîne un différentiel d'expression pour 191 gènes (Weiss et al., 2014). L'expression d'une majorité de facteurs de virulence ou de régulateurs transcriptionnels est activée en présence de la sous-unité δ alors que

l'expression de nombreuses protéines phagiques est réprimée. Les auteurs ont montré que la délétion de *rpoE* entraîne une diminution de la virulence dans des modèles d'infection murins ou humains. De plus, ils ont déterminé que les gènes fortement transcrits dans une souche sauvage de *S. aureus* étaient plus négativement impactés et qu'inversement les gènes peu exprimés l'étaient positivement. Les auteurs en ont déduits qu'il y a une normalisation de l'expression génique dans une souche mutée pour *rpoE.* Cette protéine semble donc nécessaire à la spécificité de la transcription de certains gènes. Cette hypothèse est également confortée par des études chez *B. subtilis* où il a été montré que le facteur δ inhibe la formation du complexe ouvert de lecture (dernière étape de l'initiation de la transcription) et que l'effet est d'autant plus fort que le gène est sous contrôle d'un promoteur dit faible, c'est-à-dire peu affin pour la fixation de l'ARN polymérase (Juang and Helmann, 1994). Cet effet favorise donc la transcription de gènes dont l'expression dépend de promoteur fort (forte affinité de l'ARN polymérase).

- La sous-unité ω

Le facteur ω est une sous-unité de l'ARN polymérase aussi bien chez les Procaryotes que les Eucaryotes (Minakhin et al., 2001). L'étude, chez *S. aureus,* du facteur ω et de son gène *rpoZ* a permis de confirmer certains rôles de cette sous-unité retrouvée chez d'autres membres procaryotes (Weiss et al., 2014). La sous-unité ω aurait un rôle dans la stabilité du complexe protéique ARN polymérase et dans l'association du core-enzyme avec le facteur σ. La délétion du gène *rpoZ* provoque une diminution de la capacité de formation de biofilms et donc une diminution de la résistance de la bactérie à des stress environnementaux (Weiss et al., 2014). Des études sur le facteur ω d'*E. coli* ont montré un rôle dans la réponse stringente (en cas de carence nutritive). Cependant, ce rôle n'a pas été retrouvé chez les Firmicutes comme *B. subtilis* ou *S. aureus* (Ross et al., 2013).

- La sous-unité ε

Préalablement nommé ω_1, le facteur ε est codé par *rpoY*, gène issu d'un opéron bicistronique codant également pour une ribonucléase RNase J1 (Keller et al., 2014). Le facteur ε et le core-enzyme de l'ARN polymérase sont co-localisés dans le nucléoïde (correspondant à la localisation du chromosome compacté). Ces données rejoignent l'hypothèse que le facteur ε est réellement une sous-unité du core-enzyme et aurait donc un rôle dans la transcription. En revanche, les expériences pour déterminer le rôle de la sous-unité ε ne permettent pas de déterminer son action sur l'ARN polymérase. En effet, la délétion du gène *rpoY* n'entraîne aucune variation phénotypique (croissance bactérienne, viabilité) et aucune différence d'activité transcriptionnelle pour l'ARN polymérase. Le site de fixation de la sous-unité ε est positionné près de la sous-unité ß' au niveau du canal de positionnement de

l'ADN. Cette localisation du facteur ε est similaire à celle de la protéine Gp2 sur l'ARN polymérase. Gp2 est une protéine phagique qui inhibe l'initiation de la transcription. Les auteurs de l'étude émettent l'idée que le facteur ε protège l'ARN polymérase de la fixation de la protéine Gp2 (Keller et al., 2014).

- La sous-unité σ

Il existe, dans le règne Procaryote, différents types de sous-unité σ et le nombre de facteur σ est variable selon la bactérie étudiée. L'analyse structurelle de ces facteurs σ montre quatre régions structurées en hélices conservées : $\sigma_{1\text{-}1}$, σ_2, σ_3 et σ_4. La présence ou l'absence d'une de ses parties permet de classer les facteurs σ en quatre groupes.

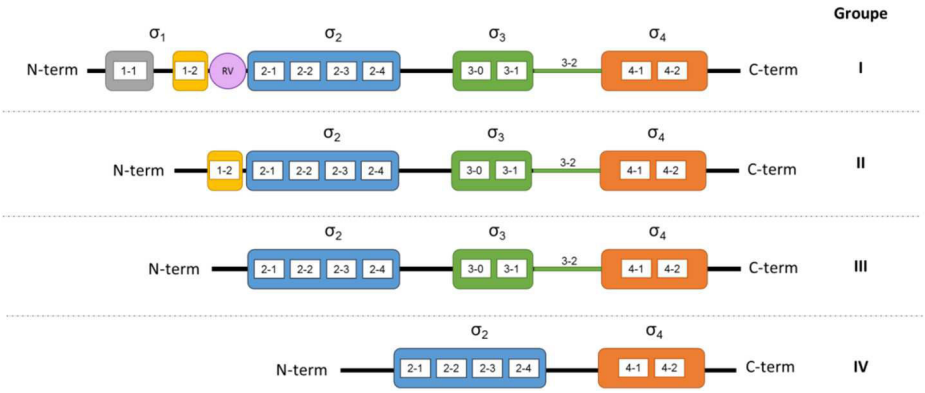


Figure 10 : Les différentes classes de facteurs σ

Le Groupe I contient l'intégralité des régions précédemment citées. Les membres de cette famille regroupent les facteurs σ majoritaires chez les bactéries. Les facteurs σ du Groupe II sont structurellement proches à ceux du Groupe I. Ils ne leur manquent que la région $\sigma_{1\text{-}1}$ ainsi qu'une région dont la séquence n'est pas conservée entre les différents facteurs (RV). Les éléments du Groupe III possèdent les régions σ_2, σ_3 et σ_4 et enfin le Groupe IV seulement les parties σ_2 et σ_4. La conservation de ces deux dernières indique leur rôle important au sein de ce facteur. Elles sont impliquées dans la fixation de la sous-unité σ à l'ARN polymérase et à la reconnaissance de l'ADN.

Quatre facteurs σ ont été identifiés chez *S. aureus* (σ^A, σ^B, σ^S et σ^H). Dans cette partie, je vous présenterai le facteur σ^A, les autres seront décrits dans une prochaine section de l'introduction. Caractérisé chez *S. aureus* en 1996, le facteur σ^A a un rôle essentiel dans l'expression des gènes dits de ménage, c'est-à-dire, les gènes exprimés dans des conditions de culture classique (Deora and Misra, 1996a). Associé au core-enzyme de l'ARN polymérase, le facteur σ^A reconnaît différentes séquences génétiques précises au cœur de régions promotrices des gènes.

2. Éléments essentiels d'un gène pour la transcription

Le génome de *S. aureus* est composé d'environ 2700 gènes. Pour pouvoir réaliser la transcription, l'ARN polymérase a besoin d'identifier une zone particulière en amont de gènes appelée promoteur. La reconnaissance des promoteurs se fait par la fixation d'une des sous-unités σ sur des séquences conservées.

Le facteur σ^A est l'homologue du facteur σ^{70} d'*E. coli*, facteur nécessaire à l'expression des gènes de ménage de la bactérie. Du fait d'une conservation importante des séquences promotrices bactériennes, nous pouvons faire l'analogie entre les promoteurs d'*E. coli* et de *S. aureus*.

La transcription du gène commence au niveau d'un nucléotide particulier (+1 de transcription) et il sert de base pour localiser les zones importantes du promoteur. Le premier nucléotide transcrit est préférentiellement une purine (Adénine ou Guanine) (Liu and Turnbough, 1994).

Les promoteurs dépendants du facteur σ^{70} sont composés de plusieurs boîtes, chacune étant fixée par une région spécifique du facteur σ^{70}. Deux boîtes ont été identifiées comme nécessaires à la fixation de l'holoenzyme, il s'agit des boîtes -35 et -10. L'alignement des séquences promotrices d'*E. coli* a permis d'établir une séquence consensus pour ces deux boîtes. Le consensus trouvé est composé de deux hexamères (boîte -35 : TTGACA et boîte -10 : TATAAT) séparés par un espaceur de 17nt (+/- 1) dont la séquence n'est pas conservée (deHaseth and Helmann, 1995; Lisser and Margalit, 1993). Une étude récente a permis d'identifier, entre autres, chez *S. aureus*, le nucléotide +1 des gènes transcrits et ainsi d'identifier les séquences des boîtes -35 et -10 de ces gènes. Les auteurs ont pu montrer que la séquence -10 de *S. aureus* (TAaAaT) est proche de la séquence consensus d'*E. coli* (TATAAT). Par contre, aucune séquence consensus pour la boîte -35 n'est ressortie de cette analyse chez *S. aureus*. L'explication donnée est que l'hexamère -35 a une séquence plus variable que la boîte -10 et que la distance avec le +1 de transcription n'est pas fixe (Prados et al., 2016).

L'ARN polymérase positionnée sur le promoteur recouvre les nucléotides allant du -50 à +5 (Li and McClure, 1998). Parmi ces 55 nt, d'autres éléments ont été caractérisés. Nous trouvons ainsi la boîte -10 prolongée (en amont de la boîte -10), le discriminateur localisé entre la boîte -10 et le +1 de transcription. Au regard des données de Prados *et al.*, nous supposons que la séquence -10 prolongée de *S. aureus* est TGT et que le discriminateur est riche en adénine et thymine (Prados et al., 2016).

Enfin, la région en amont de la boîte -35 est également importante pour une transcription efficace des gènes ciblés. Cette séquence, très riche en adénie et thymine (-59 nnAAA(A/T)(A/T)T(A/T)TTTTnnAAAAnnn -38) est fixée par les deux sous-unités α, contrairement aux autres boîtes identifiées qui le sont par le facteur σ.



Figure 11 : Composition d'un promoteur bactérien. Les séquences consensus de chaque boîtes sont placées en dessous et selon les couleurs apliquées.

B. Les étapes de la transcription

1. Reconnaissance du promoteur par l'ARN polymérase

La transcription est un processus moléculaire complexe qui nécessite la fixation de l'ARN polymérase sur l'ADN. Cependant, la fixation initiale de l'holoenzyme (core-enzyme + facteur σ) n'est pas spécifique. Il a été proposé que le complexe protéique de l'ARN polymérase diffuse sur l'ADN, cela signifie que l'enzyme glisse le long de l'ADN (Park et al., 1982; Sakata-Sogawa and Shimamoto, 2004). Dans un phénomène de diffusion sur l'ADN, l'holoenzyme s'approche d'un promoteur, la sous-unité σ fixe les différentes boîtes présentées ci-dessus. Ainsi les boîtes -35, -10 prolongée, -10 et Discriminateur sont respectivement fixées par les parties $\sigma_{4\text{-}2}$, σ_3, $\sigma_{2\text{-}4}$ et $\sigma_{1\text{-}2}$. La fixation de la sous-unité σ entraînerait alors un arrêt de l'holoenzyme sur le promoteur permettant ainsi d'initier la transcription (Kabata et al., 1993).

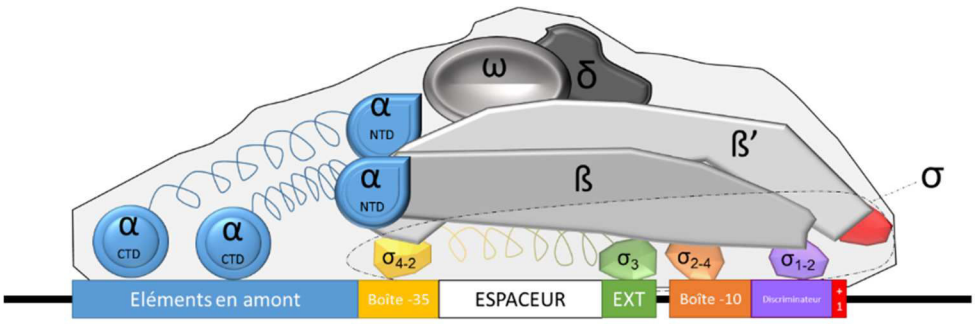


Figure 12 : Positionnement de l'ARN polymérase sur la région promotrice.

2. Initiation et élongation de la transcription

L'arrêt de l'ARN polymérase au niveau de promoteur permet de définir un complexe fermé entre l'holoenzyme et l'ADN. L'ADN se courbe au niveau des éléments en amont de la boîte -35 grâce à la flexibilité des sous-unités α qui le fixent. La courbure de l'ADN autour de l'holoenzyme entraîne une fragilisation des liaisons entre les thymines et les adénines présentes dans la boîte -10. L'ADN s'ouvre à ce niveau et le brin anti-sens peut s'insérer dans le site catalytique de l'enzyme. Enfin, l'holoenzyme se positionne dans sa conformation finale correspondant un complexe ouvert stable. (Ruff et al., 2015).

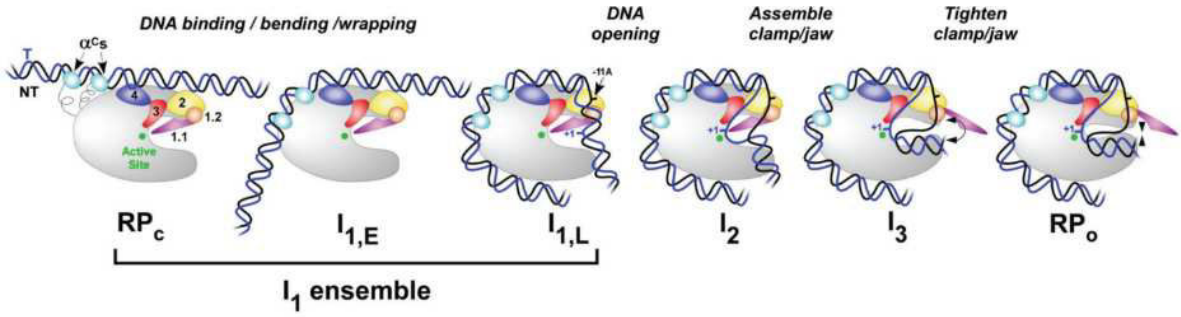


Figure 13 Les différents étapes de l'initiation de la transcription. Le passage d'un complexe fermé (RP_C) à un complexe ouvert (RP_O) est réalisé par la fixation de l'ARN polymérase sur l'ADN puis par l'émergence d'une courbure du chromosome ($I_{1,E}$). Puis, les deux brins d'ADN au niveau de la boîte -10 se séparent ($I_{1,L}$), le brin complémentaire se positionne dans le site catalytique (I_2 et I_3). D'après Ruff et al., 2015.

La transcription commence par l'appariement d'un nucléotide au nucléotide complémentaire au +1 de transcription. La complémentarité des nucléotides est la suivante : à une adénine est associée un uracile, à une thymine est associée une adénine et à une cytosine est appariée à une guanine (et inversement). Il est important de noter que le nucléotide uracile est spécifique à l'ARN et la thymine à l'ADN. La différence entre ces deux nucléotides est l'absence d'un groupement méthyle (-CH3) sur la base de l'uracile. Néanmoins, cette particularité n'impacte pas la liaison entre l'adénine et l'uracile.

Une fois, la base complémentaire du +1 transcription placée, la transcription continue sur les sept premières bases. Cependant, la fixation stable sur le promoteur de l'holoenzyme empêche l'élongation de la transcription et le petit ARN est relâché du complexe (Murakami et al., 2002). Cet évènement se répète mais cette fois-ci, l'énergie accumulée lors de la première transcription déstabilise le complexe ouvert. La sous-unité σ se dissocie du core-enzyme et l'élongation de la transcription continue. Au fur et à mesure de l'avancée de l'ARN polymérase sur l'ADN, les nucléotides complémentaires s'ajoutent et ce jusqu'à la dernière étape de la transcription (Imashimizu et al., 2014).

3. Terminaison de la transcription

La terminaison de la transcription permet la libération de l'ARN dans le cytoplasme. Il existe deux types de terminaison : une terminaison intrinsèque et une terminaison Rho-dépendante (Figure 14).

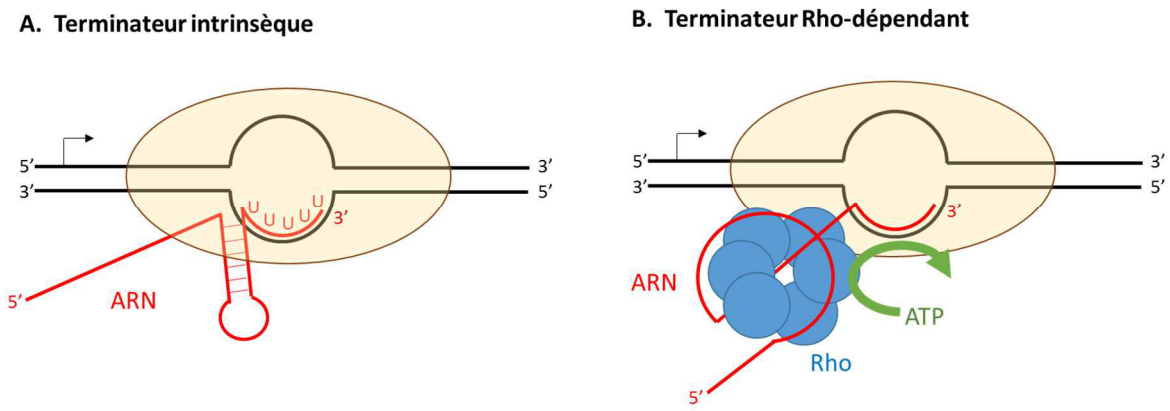


Figure 14 : Différents types de signaux de terminaison de la transcription (A : Terminaison intrinsèque ; B : Terminaison Rho-dépendante)

a. Terminaison intrinsèque

La terminaison intrinsèque de la transcription est indépendante de tous facteurs. L'élongation de l'ARN permet l'apparition d'une structure secondaire tige-boucle, riche en guanines et cytosines, suivie par une succession de plusieurs uraciles. Cette structure, positionnée dans le core-enzyme permet une déstabilisation du complexe entre l'ADN et l'ARN nouvellement formé, au niveau des uraciles. La transcription prend fin et l'ARN est libéré du complexe (Gusarov and Nudler, 1999).

b. Terminaison Rho-dépendante

Lorsque la séquence nucléotidique ne permet pas la présence d'une structure secondaire pour la terminaison de la transcription, le facteur de terminaison Rho entre en jeu. Rho s'associe en homo-hexamère qui se positionne sur le transcrit au niveau d'une région riche en cytosine et pauvre en guanine appelée Rut (Rho utilization site). L'interaction de l'ARN dans le cœur catalytique de Rho entraîne l'activation de sa fonction ATPase permettant la dissociation de transcrit du core-enzyme de l'ARN polymérase (Porrua et al., 2016; Proshkin et al., 2014; Santangelo and Artsimovitch, 2011).

III. La régulation de la transcription

Etape essentielle à la vie, la transcription est soumise à de nombreuses régulations qui peuvent être constitutives ou ponctuelles. Les régulations de l'ARN polymérase peuvent faire intervenir des facteurs σ alternatifs ou des régulateurs transcriptionnels agissant en *trans*.

A. Régulations intrinsèques de la transcription :

1. La région promotrice

Nous avons vu que le promoteur est essentiel pour la reconnaissance d'un gène par l'ARN polymérase. L'étude des promoteurs a permis de mettre en place une séquence

consensus de fixation du facteur σ^A. Or, les régions promotrices ne sont pas toutes identiques et cela pourrait expliquer les différences dans les niveaux d'expression de gènes. Des variations dans les séquences des boîtes -10/-35 ainsi que leur positionnement par rapport au +1 de transcription entraînent des variations de l'efficacité de la transcription (Rhodius and Mutalik, 2010). Le +1 de transcription est également important. Ce nucléotide indispensable à l'initiation de la transcription doit être préférentiellement une adénine ou une guanine. La présence d'une thymine ou d'une cytosine à cette position entraîne une diminution importante de l'activité transcriptionnelle (Liu and Turnbough, 1994).

2. Régulations par interférence transcriptionnelle

Le positionnement des gènes et de leurs promoteurs peut constituer un niveau de régulation. L'interférence transcriptionnelle correspond à l'action négative de la transcription d'un promoteur sur la transcription d'un autre. Si on considère que les promoteurs ont des affinités différentes avec l'ARN polymérase et qu'un est considéré comme fort et l'autre comme faible, alors ces organisations chromosomiques entreront en compétition pour la fixation de l'ARN polymérase. Trois cas de figure sont possibles : des promoteurs convergents, en tandem et chevauchants (Figure 15). Le mécanisme commun aux trois est la compétition : la fixation de l'ARN polymérase sur le promoteur fort empêche la fixation d'un second complexe transcriptionnel sur le promoteur faible. En ce qui concerne les promoteurs convergents, quatre autres mécanismes ont été caractérisés. Le délogement de l'ARN polymérase du promoteur faible peut être effectué lors de l'étape d'élongation de la transcription du promoteur fort. La transcription du promoteur faible n'est donc pas initiée. Sur le même principe, lors de l'élongation de la transcription du promoteur fort, l'ARN polymérase se positionne au niveau du promoteur faible empêchant ainsi la fixation d'une seconde ARN polymérase (c'est le mécanisme d'occlusion). A l'inverse, si la transcription du promoteur faible a débuté, alors l'ARN polymérase peut empêcher l'élongation du promoteur fort. Néanmoins, il n'y a pas de délogement de l'ARN polymérase du fait du caractère « fort » du promoteur. Enfin, si la transcription des deux promoteurs a lieu, alors les deux ARN polymérases vont entrer en collision rendant impossible la transcription complète des deux gènes (Shearwin et al., 2005).

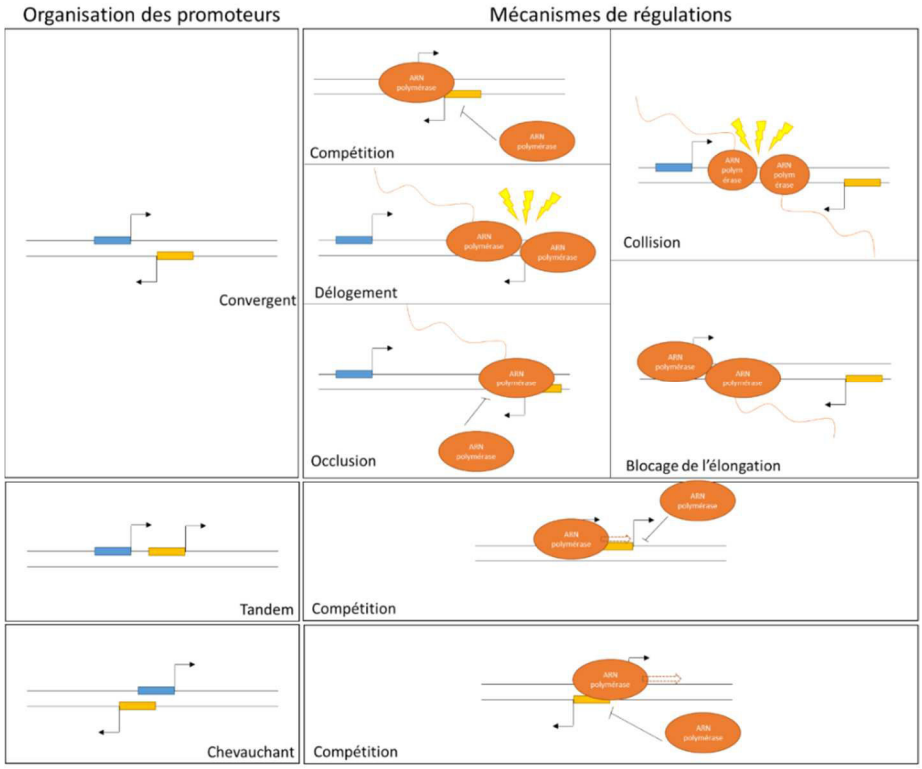


Figure 15 : Interférence transcriptionnelle par l'organisation chromosomique des promoteurs. Les promoteurs forts et faibles sont respectivement représentés par des cadres bleus et oranges.

3. Les facteurs σ alternatifs

Nécessaire à la spécificité de l'ARN polymérase pour les promoteurs, les quatre facteurs σ de *S. aureus* (σ^A, σ^B, σ^S et σ^H) permettent d'ajouter une spécificité supplémentaire dans la reconnaissance de promoteurs spécifiques entraînant une expression de gène dépendante de l'expression du facteur σ. Nous avons vu dans la section précédente que le facteur σ^A est majoritairement utilisé pour la transcription de gènes chez *S. aureus*.

a. Le facteur σ^B

Le facteur σ^B a été identifié par homologie de séquence avec le facteur alternatif σ^B de *Bacillus subtilis* (Wu et al., 1996). Issu d'un opéron de quatre gènes, les gènes *rsbU*, *rsbV*, *rsbW* et *sigB* sont co-exprimés de manière constitutive par l'ARN polymérase associée à la sous-unité σ^A. Dans des conditions de croissance, la protéine RsbV est phosphorylée et le facteur σ^B est séquestré par l'anti-facteur-σ RsbW. En condition de stress, l'anti-RsbV phosphatase RsbU est activée ce qui permet la déphosphorylation de RsbV. RsbV déphosphorylé est ensuite fixé par RsbW, libérant ainsi la sous-unité σ^B. Le facteur σ^B peut ensuite s'associer au core-enzyme de l'ARN polymérase permettant la transcription des gènes cibles de σ^B (Senn et al., 2005).

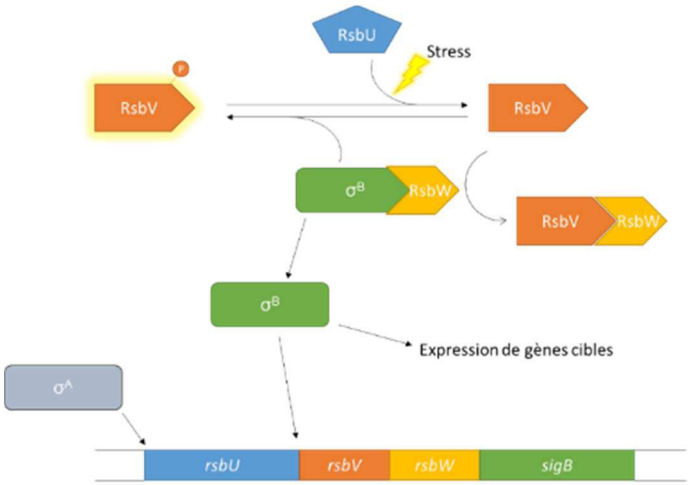


Figure 16 : Régulation protéique du facteur σ^B par RsbUVW.

La délétion du gène *sigB* entraîne plusieurs modifications phénotypiques. Il a été montré que les souches Δ*sigB* ont, entre autres, une capacité de sédimentation et d'agrégation augmentée et ne présentent pas une pigmentation orangée caractéristique de *S. aureus* (Kullik et al., 1998). σ^B serait également impliqué dans la survie dans les phagolysosomes (mécanisme de dégradation de pathogène par les cellules immunitaires) du fait de son implication dans l'adaptation aux pH acides et aux stress oxydants. Enfin, les hautes températures semblent être également un stress permettant la libération de σ^B par RsbW (Chan et al., 1998). L'ensemble de ces observations montre l'importance de σ^B dans la régulation générale de *S. aureus*.

Quelques cibles de ce facteur alternatif ont déjà été identifiées. Le facteur σ^B active ainsi l'un des trois promoteurs de SarA, facteur de transcription majeur chez *S. aureus* (Bischoff et al., 2001; Cheung et al., 1999). σ^B active la transcription de Asp23, l'une des protéines associées à la membrane les plus abondantes de *S. aureus* (Gertz et al., 1999). Suite à des expériences de protéomique (2D-DIGE), 27 protéines, nommées Csg (pour controlled by sigma B), ont été identifiées comme appartenant au régulon du facteur σ^B. En 2004, une séquence consensus de la fixation de σ^B a été proposée en alignant plusieurs gènes candidats pour une régulation par ce facteur alternatif. Très proche de la séquence consensus pour la fixation de σ^B de *B. subtilis*, σ^B se fixe préférentiellement sur cette séquence : GttTaa-$N_{12\text{-}15}$-gGGTAt. La recherche de cette séquence dans le génome de la souche N315 de *S. aureus* a ensuite permis de révéler 18 gènes régulés par σ^B (Homerova et al., 2004).

b. Le facteur σ^H

Peu de choses sont encore connues pour le facteur alternatif σ^H. Identifié par homologie de séquences d'acides aminés avec le facteur σ^H de *B. subtilis* (Morikawa et al., 2003), il est impliqué dans l'évolution de la bactérie. En effet, le facteur σ^H permet l'expression

du gène *int,* qui produit les intégrases nécessaires à l'intégration ou l'excision chromosomique d'éléments accessoires du génome (prophages ou îlots de pathogénie) (Tao et al., 2010).

c. Le facteur σ^S

Dernier facteur σ alternatif découvert, σ^S appartient à la famille des facteurs σ ECF (Groupe 4 : fonction extracytoplasmique) (Shaw et al., 2008). Comme la plupart des ECF, σ^S contribue à l'adaptation aux stress environnementaux et à la survie de la bactérie. La présence de *sigS* permet une meilleure résistance à la lyse au Triton X-100 (détergent) et est un avantage lors d'expériences de compétition avec la souche mutant Δ*sigS* (Shaw et al., 2008). Ce facteur σ alternatif est aussi essentiel à la virulence de *S. aureus* (Shaw et al., 2008). Le gène *sigS* est exprimé à environ 34 et 48 heures de croissance, son expression est donc principalement réprimée dans une croissance classique.

B. Régulations extrinsèques de la transcription :

1. Les facteurs de transcription

Alors que nous avons vu que la régulation de l'initiation de la transcription sur des gènes cibles peut se faire par l'appariement du core enzyme avec des facteurs σ différents, d'autres protéines peuvent entrer en jeu pour réguler l'efficacité de la transcription de gènes. Les facteurs de transcription sont des protéines capables de se lier à des séquences ADN spécifiques. Nous distinguons les facteurs qui vont favoriser la transcription (les activateurs) de ceux qui vont la réduire (les répresseurs).

a. Activation de la transcription

Pour activer la transcription, les facteurs de transcription peuvent se localiser à trois endroits du promoteur, chaque position entraînant un mode d'activation différent.

L'activation de classe I correspond à une fixation du facteur de transcription en amont de la boîte -35 et permet le recrutement de la partie CTD (C-terminal domain) des sous-unités α de l'ARN polymérase.

Les activateurs de classe II vont, quant à eux, se lier à l'ADN juste en amont de la zone de fixation du domaine 4 de la sous-unité σ de l'ARN polymérase. Cette position permet un recrutement amélioré de l'holoenzyme sur l'ADN et entraîne donc une augmentation de l'activité transcriptionnelle.

Le dernier mécanisme est mis en place lorsque la conformation de l'ADN n'est pas optimale pour la transcription du gène. En effet, lorsque le nombre de nucléotides entre les boîtes -35 et -10 est supérieur à celui l'optimum, la fixation de la sous-unité σ est affectée, le recrutement de l'holoenzyme est moins efficace, la transcription ne se fait pas. Alors, des

facteurs de transcription peuvent se positionner entre ces 2 boîtes et permettre le rapprochement spatial de celles-ci. Le facteur σ pourra alors fixer plus efficacement ces sites de fixation, la transcription est activée.

b. Répression de la transcription

La répression de la transcription est tout aussi importante que l'activation pour réguler au mieux l'expression des gènes. Trois types de mécanisme de répression sont recensés.

Lorsqu'un facteur de transcription se positionne au niveau du site de fixation de l'ARN polymérase, il y a alors un encombrement stérique entre les deux pour la zone d'intérêt. Le facteur de transcription empêche la fixation de l'ARN polymérase sur le promoteur, la transcription ne peut avoir lieu.

Le deuxième mécanisme entraîne un changement conformationnel important du chromosome. La fixation des facteurs de transcription dans la région promotrice ou génique entraînent une courbure de l'ADN sur lui-même. Cette conformation empêche l'ARN polymérase de se fixer sur le promoteur et d'initier donc la transcription.

Enfin, le dernier mécanisme n'agit pas directement sur l'ARN polymérase mais sur les activateurs qui lui sont liés. Le facteur de transcription répresseur se fixe à proximité du facteur de transcription activateur. Cette proximité inhibe la fixation de l'activateur transcriptionnel ce qui impacterait le recrutement de l'ARN polymérase ou l'initiation de la transcription.

c. Dynamique des facteurs de transcription

Il paraît évident que ces simples mécanismes de répression ou d'activation ne peuvent expliquer l'ensemble des nuances dans la régulation de l'expression de gènes. Il faut considérer l'ensemble (promoteur, ARN polymérase et facteurs de transcription) comme un mécanisme dynamique où se jouent plusieurs niveaux de régulation. Ainsi des activateurs et des répresseurs peuvent se lier à un même promoteur pour réguler finement son activité. La fixation de plusieurs types d'activateurs permet d'amplifier le signal pour le recrutement de l'ARN polymérase, que leur fixation soit indépendante ou conjointe.

2. Modifications post-traductionnelles des facteurs de transcription

Nous avons vu que les facteurs de transcription sont des régulateurs majeurs de l'initiation de la transcription. Il existe, entre autres, deux grandes familles de facteurs : les systèmes à deux composants et les facteurs agissant seuls.

Les systèmes à deux composants (TCS) sont des éléments essentiels à la communication entre la bactérie et l'environnement extérieur. En effet, les TCS sont composés d'un senseur (protéine de type histidine kinase liée à la membrane) et d'un régulateur

cytoplasmique. De manière simplifiée, l'arrivée d'un signal à la membrane entraîne une activation du senseur par phosphorylation. Le senseur activé peut alors à son tour phosphoryler le régulateur. Le changement conformationnel du régulateur, par modification post-traductionnelle, permet ensuite la reconnaissance de l'ADN par le régulateur phosphorylé. Ce dernier peut donc agir sur la transcription.

Cette phosphorylation entraîne une modification de l'affinité du régulateur pour l'ADN. Ce phénomène est également retrouvé pour les facteurs agissant seuls. En effet, il existe des protéines capables d'ajouter (les kinases) ou de supprimer (les phosphatases) des groupements phosphates à une protéine. L'ajout de ce groupement se fait sur des acides aminés spécifiques : sérine, thréonine, cystéine et tyrosine. Chez *S. aureus*, deux kinases ont été identifiées (Stk1 et Stk2) et la phosphatase majeure est Stp1 (Debarbouille et al., 2009). Ces protéines sont impliquées dans la modification post-traductionnelles de plusieurs facteurs de transcription tels que VraR, GraR, SarZ, MgrA et SarA (Belcheva et al., 2009; Fridman et al., 2013; Sun et al., 2012). Ces trois derniers sont aussi sujets à une autre modification : l'oxydation. En présence d'agents oxydants, les cystéines présentes dans les protéines sont modifiées entraînant une modification de la structure tridimensionnelle de la protéine (Reddie and Carroll, 2008). La modification de ces régulateurs entraîne, selon le type de modification et le résidu modifié, une diminution ou une augmentation de l'affinité de la protéine pour l'ADN. Ainsi, l'effet régulateur de ces facteurs de transcription en sera modulé (Ballal and Manna, 2009b, 2010; Chen et al., 2006; Chen et al., 2009).

La fixation non covalente de petits ligands sur les facteurs de transcription modifie leurs propriétés de liaison à l'ADN. Notons l'exemple des alarmones émises lors d'un stress nutritif. En cas de carence nutritive, une régulation globale de la transcription et de la traduction sont nécessaires. Pour ce faire, la bactérie émet des alarmones ppGpp et pppGpp (appelées (p)ppGpp) synthétisées par l'enzyme RSH à partir de GTP ou d'ATP (guanosine ou adénosine tri-phosphate). (p)ppGpp ainsi produit va ainsi diminuer la quantité de GTP capable de se lier au facteur de transcription CodY, initialement localisé sur les promoteurs de gènes spécifiques dont il réprime la transcription. En cas de carence nutritive, la liaison CodY-GTP est dissocié et entraîne le décrochage de CodY de l'ADN permettant ainsi à l'ARN polymérase d'initier la transcription de gènes spécifiques habituellement réprimés par CodY(Geiger and Wolz, 2014).

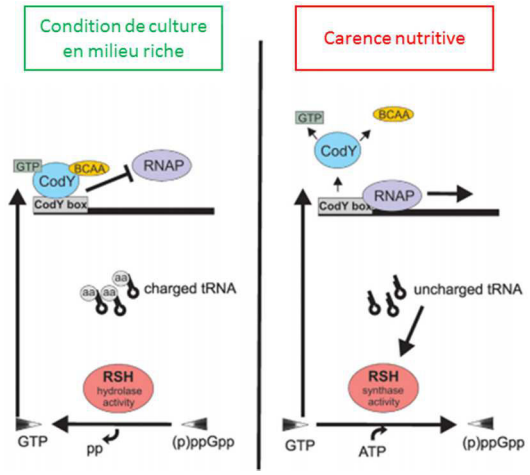


Figure 17 : Régulation transcriptionnelle en cas de carence nutritif. Dans un milieu non carencé, le facteur de transcription CodY empêche la transcription de ces gènes cibles. En cas de carence, la production de l'alarmone (p)ppGpp entraîne la dissociation de CodY du promoteur permettant la transcription. D'après (Geiger and Wolz, 2014)

Les différents types de régulations de la transcription présentés précédemment ne reflètent qu'une partie des régulations pouvant être recensées. Notons à titre d'exemple les changements de la topologie de l'ADN. En effet, l'état de surenroulement négatif ou positif du chromosome permet la reconnaissance plus ou moins facile de l'ARN polymérase pour le promoteur bactérien. Ce surenroulement peut être modifié par des enzymes spécifiques, les topoisomérases ou des protéines associées aux chromosomes comme la protéine HU, retrouvée entre autres chez *E. coli* et *S. aureus* et qui aurait un rôle architectural de la chromatine comparable aux histones des Eucaryotes.

IV. Les facteurs de transcription de *Staphylococcus aureus*

S. aureus est un pathogène produisant de nombreux facteurs de virulences tels que des toxines, adhésines, etc. L'expression de ces facteurs de virulence doit être finement régulée pour permettre la réponse appropriée aux conditions environnementales. La bactérie exprime de nombreux facteurs agissant sur la transcription de gènes de virulence. Ces facteurs de transcription ont la particularité de lier l'ADN pour activer ou réprimer la transcription. En 2013, une étude sur la conservation des gènes codant pour des facteurs de transcription dans les différentes souches de *S. aureus* a permis de déterminer la présence d'environ 130 facteurs de transcription. Parmi ceux-ci, deux groupes sont identifiés : les facteurs agissant en binôme (TCS : système à deux composants) et ceux ayant une activité propre (Ibarra et al., 2013).

A. Les systèmes à deux composants

Seize TCS ont été recensés chez *S. aureus* (Cheung et al., 2004). Le système WalKR est le seul TCS à être essentiel à la bactérie, c'est-à-dire que la délétion de ces gènes est létale pour *S. aureus* (Dubrac et al., 2007). Les TCS sont impliqués dans de nombreuses fonctions bactériennes comme l'antibiorésistance, la formation de biofilm, la virulence…

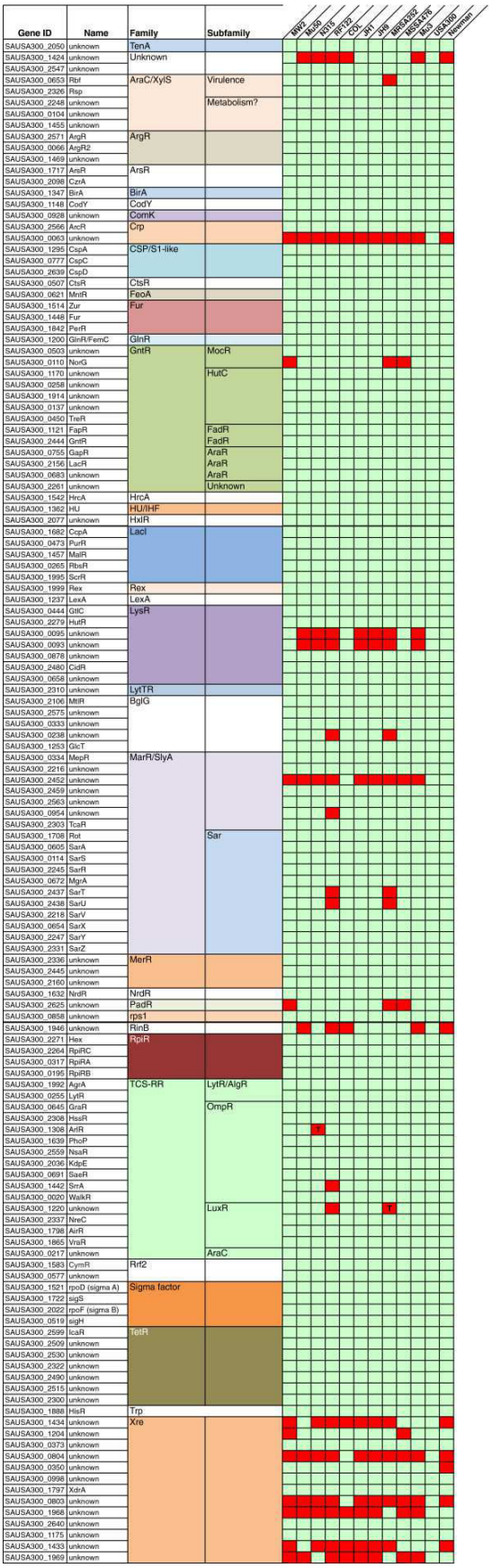


Figure 18 : Comparaison des facteurs de transcription de la souche USA300-FPR avec d'autres souches de *S. aureus* (présence du gène en vert, absence en rouge). D'après (Ibarra et al., 2013).

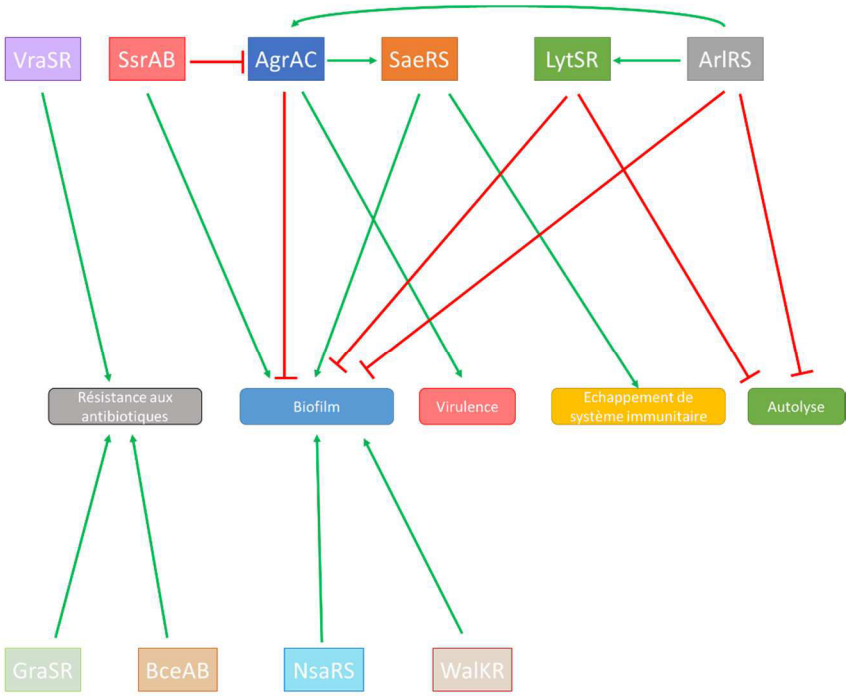


Figure 19 : Implication des TCS dans différentes fonctions cellulaires. Les effets positifs des TCS sur les fonctions sont représentés par une flèche verte et les effets négatifs par une flèche interrompue rouge.

Un des TCS le mieux caractérisé de *S. aureus* est AgrAC. Senseur de la densité bactérienne, il est le cœur de la régulation de nombreuses protéines permettant une expression génique différentielle selon le voisinage bactérien de la bactérie.

Au cours de la croissance bactérienne, *S. aureus* a besoin de mesurer en permanence la concentration bactérienne qui l'entoure. Ce système de mesure, appelé « quorum sensing », est géré par le système *agr*. Ce système possède deux promoteurs P2 et P3 produisant deux transcrits distincts : l'ARNII et l'ARNIII. L'ARNII est le transcrit d'un opéron de 4 gènes codant pour 4 protéines nécessaires au quorum sensing. : AgrA, AgrB, AgrC et AgrD. *AgrD* code pour un peptide qui est modifié en AIP (Auto-inducing peptide). AIP est sécrété dans le milieu extracellulaire par le transporteur membranaire AgrB. Plus la concentration bactérienne est forte, plus la concentration en AIP extracellulaire sera forte. La quantité d'AIP présente sera détectée par AgrC qui s'activera par autophosphorylation. Ensuite, AgrC active par phosphorylation le facteur de transcription AgrA qui pourra réguler la transcription de ses cibles. AgrA active la transcription des ARNII et ARNIII du système *agr.* Ainsi, nous avons une augmentation de la quantité de ces deux ARN au cours de la croissance. Ce système de « quorum sensing » permet le passage des étapes de colonisation de l'hôte aux mécanismes d'infection (Queck et al., 2008).

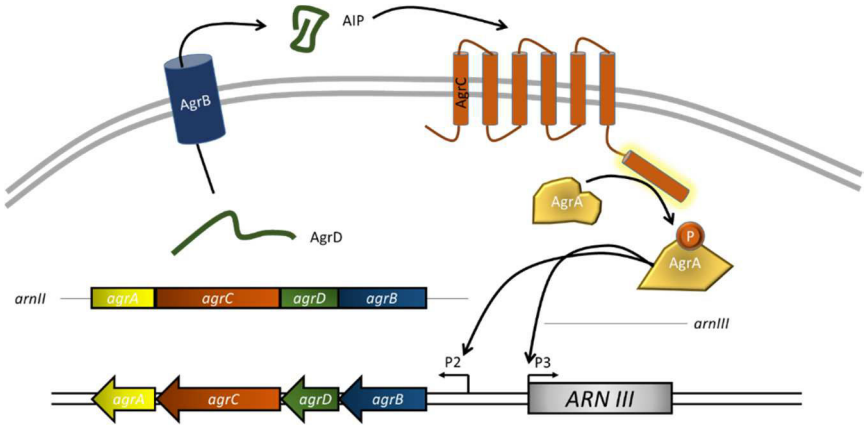


Figure 20 : Mécanisme du système de « quorum-sensing » par *agr*

Cette bascule de ces deux étapes par la répression ou l'activation de l'expression de gènes se fait par AgrA (régulateur du système) et par l'ARN régulateur ARNIII. Le système *agr* régule de nombreux mécanismes cellulaires : il réprime la formation de biofilm et active la virulence de *S. aureus.* Il est également le régulateur de nombreux autres facteurs de transcription tel que SarT, Rot et SarS (Cheung et al., 2001; Geisinger et al., 2006; Schmidt et al., 2001). Ces régulations sont les indices d'une interconnexion entre régulateurs.

B. Les facteurs de transcription

Parmi les gènes codant pour des facteurs de transcription, de nombreux ont été caractérisés comme pouvant se lier à l'ADN sans nécessité d'être « activés » par un senseur associé à la membrane. Ces facteurs de transcription sont largement présents dans la bactérie et permettent une régulation de l'expression de gènes. Environ 110 protéines sont susceptibles de réguler l'expression de gènes mais une faible partie d'entre eux a d'ores et déjà été étudiée. Nous nous intéresserons principalement aux facteurs de transcription, largement étudiés, de la famille SarA, regroupant les protéines régulatrices majeurs de *S. aureus*.

1. La famille SarA

Les facteurs de transcription de la famille SarA se caractérisent par une structure tridimensionnelle en « winged-helix ». Cette conformation, composée de 3 hélices α et de 3 feuillets ß, permet la liaison de la protéine sur des acides nucléiques. Nous pouvons distinguer dans cette structure une partie « hélice-coude-hélice », représentant le cœur de la structure, et la partie « ailée » (winged) qui est en dehors de la première partie. Les protéines « winged-helix » sont les facteurs de transcription les plus retrouvés chez les procaryotes (Aravind et al., 2005).

a. Membres de la famille SarA

11 protéines ont été ainsi caractérisées comme faisant partie de cette famille (SarA, SarR, SarS, SarT, SarU, SarV, SarX, SarY, SarZ, MgrA et Rot). Les facteurs de transcription peuvent être séparés en 2 groupes. Nous avons les protéines, de petite taille, agissant en dimère (SarA, SarR, SarT, SarV, SarX, SarZ, MgrA et Rot) et les protéines de plus grosse taille (SarS, SarU et SarY) agissant en monomères. Il est intéressant de noter que la structure de ces dernières mime un dimère de petites protéines.

SarA	Dimère	124 Acides aminés
SarR		115 Acides aminés
SarT		118 Acides aminés
SarV		116 Acides aminés
SarX		141 Acides aminés
SarZ		148 Acides aminés
Rot		166 Acides aminés
MgrA		147 Acides aminés
SarS	Monomère	250 Acides aminés
SarU		247 Acides aminés
SarY		247 Acides aminés

Figure 21 : Protéines membres de la famille SarA

Malgré une structure tridimensionnelle en « winged-helix » commune, la séquence en acides aminés de ces protéines est très différente. Néanmoins, il est important de noter la présence, chez 8 membres de la famille SarA (SarA, SarR, SarS, SarU, SarX, SarZ, MgrA et Rot), d'une cystéine ainsi qu'un enchaînement d'acides aminés particuliers : acide aspartique, acide glutamique et arginine. Quelques autres résidus sont également conservés afin de garder la structure et la fonction de la protéine (Cheung et al., 2008b).

En 2009, l'expression des gènes codant pour certaines protéines de la famille SarA a été étudiée dans différentes souches (Ballal and Manna, 2009a). Des analyses, à différents temps de croissance, ciblant les transcrits ou les protéines cibles ont été effectuées par respectivement Northern Blot ou Western Blot. L'analyse des résultats obtenus pour la souche Newman montre que trois facteurs de transcription sont produits de manière constitutive : SarA, SarZ et Rot. Les facteurs de transcription SarR et SarS seraient plus présents en début de phase exponentielle de croissance. En phase exponentielle, nous retrouvons les facteurs de transcription MgrA, SarR et SarX.

Début de phase exponentielle	Phase exponentielle	Fin de phase exponentielle
SarA	SarA	SarA
SarZ	SarZ	SarZ
Rot	Rot	Rot
	SarX	
	MgrA	
SarS		
SarR	SarR	

Figure 22 : Présence des protéines dans la souche Newman selon la phase de croissance de la bactérie.

b. Mécanismes de régulations

L'étude de ces facteurs de transcription a révélé différents mécanismes de régulation. Le premier correspond à la fixation du facteur (monomère ou dimère) sur l'ADN. La localisation de la liaison entraîne l'activation ou la répression de la transcription (cf pages 42 et 43). La structure active, après fixation, peut entraîner une courbure de l'ADN provoquant la régulation de l'expression du gène cible (Cheung et al., 2004). Liu *et al.* ont proposé l'existence d'une répression par fixation de plusieurs dimères sur un promoteur (Liu et al., 2006). La structure en dimère peut se faire avec deux protéines non-identiques (formation d'hétérodimère) pouvant entraîner un changement de fonction (Cheung et al., 2004). Nous avons vu qu'une cystéine est présente dans la plupart des membres de la famille SarA. Cet acide aminé peut être sujet à une oxydation. Ainsi, SarA, SarZ et MgrA peuvent être oxydées provoquant leur décrochage de l'ADN (Ballal and Manna, 2009b, 2010; Chen et al., 2006; Chen et al., 2009). Parmi les autres modifications post-traductionnelles qui ont été identifiées nous pouvons citer la phosphorylation. La kinase Stk1 peut ajouter un groupement phosphate aux protéines SarA ou MgrA et, à l'inverse la phosphatase Stp1 peut l'enlever (Didier et al., 2010). Enfin, la régulation des facteurs de transcription de la famille SarA ne peut être complète sans parler de la compétition de certaines protéines entre-elles pour la fixation d'un promoteur (par exemple, SarA et SarR pour la régulation d'*agr*) (Manna and Cheung, 2006b).

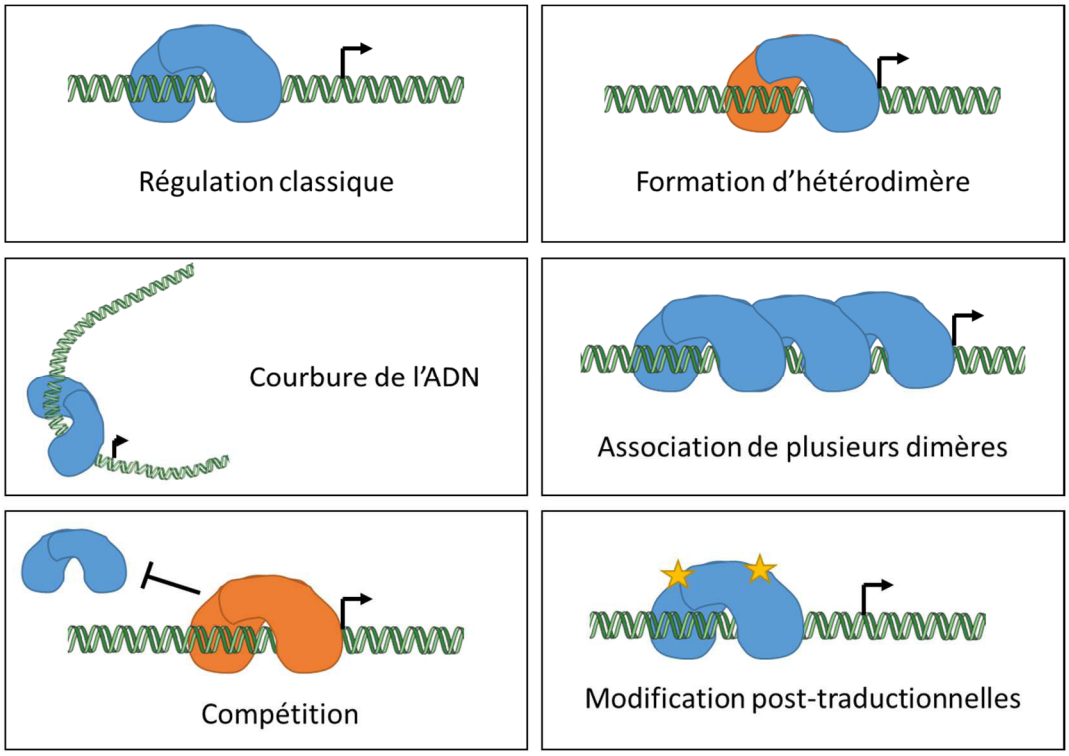


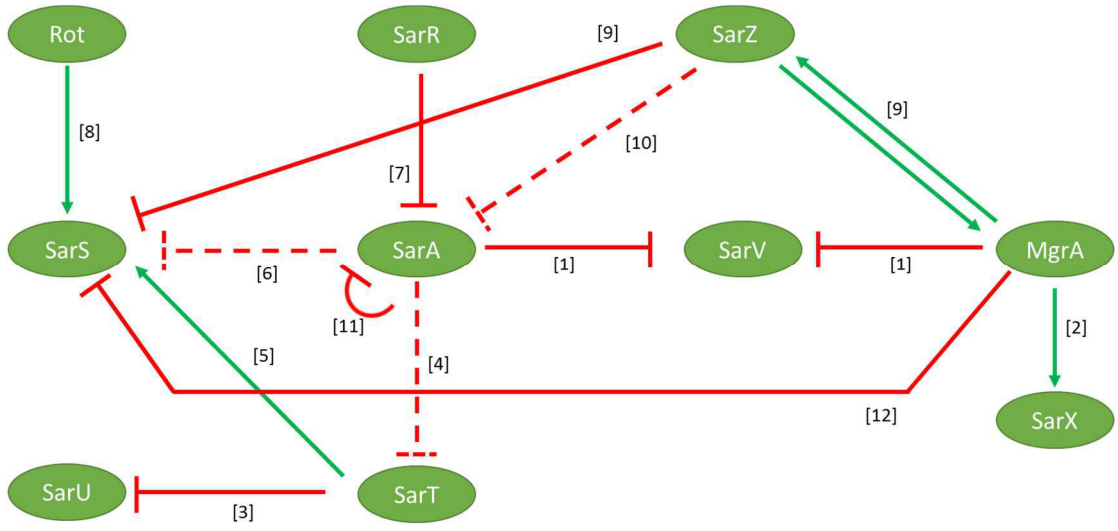
Figure 23 : Différents mécanismes d'action des facteurs de transcription de la famille SarA

Un autre niveau de régulation est possible. En effet, les protéines de la famille SarA régulent l'expression des gènes de ces propres membres.

c. Inter-régulation des membres de la famille SarA

Les membres de la famille SarA sont des régulateurs transcriptionnels de l'expression de gènes, qu'ils soient capables de se réguler entre eux mais également de participer à la régulation transcriptionnelle de leur propre gène.

Comme présenté dans la Figure 24, nous constatons que la protéine SarA réprime l'expression de *sarT, sarS, sarV* (Cheung et al., 2001; Manna et al., 2004; Schmidt et al., 2001). Le gène *sarA* est quant à lui réprimé par sa propre protéine et également par SarR et SarZ (Manna and Cheung, 2001a; Tamber and Cheung, 2009). SarZ réprime *sarS* et active MgrA (Ballal et al., 2009). MgrA active l'expression de *sarZ* et *sarX* et réprime, comme SarA, *sarV* et *sarS* (Ballal et al., 2009; Ingavale et al., 2005; Manna and Cheung, 2006a; Manna et al., 2004). Enfin, SarT a été défini comme répresseur de l'expression de *sarU* mais il peut également activer, comme Rot, l'expression de *sarS* (Manna and Cheung, 2003; Oscarsson et al., 2005; Schmidt et al., 2003). L'ensemble de ces données montre la complexité de la régulation transcriptionnelle de *S. aureus* alors que nous avons abordé ici qu'une portion infime de l'ensemble des facteurs de transcription de *S. aureus*.



[1] Manna *et al.* (2004)
[2] Manna & Cheung (2006)
[3] Manna & Cheung (2003)
[4] Schmidt *et al.* (2001)
[5] Schmidt *et al.* (2003)
[6] Cheung *et al.* (2001)
[7] Manna & Cheung (2001)
[8] Oscarsson et al. (2005)
[9] Ballal et al. (2009)
[10] Tamber et al. (2009)
[11] Cheung et al. (2008)
[12] Ingavale et al. (2005)

Figure 24 : Inter-régulations des membres de la famille SarA. Les flèches vertes représentent une activation de l'expression du gène et les flèches interrompues rouges correspondent à une répression. Les traits pointillés signifient une régulation indirecte.

2. SarA, régulateur majeur de *S. aureus*

a. Description du facteur de transcription SarA

SarA a été mis en évidence en 1992 par A.L. Cheung grâce à l'insertion du transposon Tn*971*LTV1 au niveau d'un gène dont le produit impacte la production de protéines de surface ou d'exotoxines (Cheung et al., 1992). Le gène *sarA* est sous le contrôle de trois promoteurs (P1, P3 et P2) mais produit une seule protéine : SarA (Figure 25). La région promotrice de *sarA* recouvre donc une grande région d'environ 800 pb. Les promoteurs de *sarA* sont régulés par différents facteurs. Ainsi, le promoteur P3 est réprimé par les protéines SarR et SarA, ce dernier réprimant également le promoteur P1 (Cheung et al., 2008a; Manna and Cheung, 2001a). Alors que les promoteurs P2 et P1 sont reconnus par l'ARN polymérase contenant le facteur σ^A, le promoteur P3 est, quant à lui, dépendant de la fixation de la sous-unité σ^B (Cheung et al., 1999). Ainsi, la production de SarA est constitutive quelles que soient les conditions de cultures de *S. aureus* (Ballal and Manna, 2009a).

Les promoteurs de *sarA* présentent deux gènes exprimant deux ARN régulateurs (*srn_1540_teg48* et *srn_1550_teg49*) (Kim et al., 2014). *Srn_1550_teg49* est présent au niveau du promoteur P3 alors que *srn_1540_teg48* est dans le promoteur P1. Tout comme le promoteur P3, il semble que l'expression de *srn_1550_teg49* soit dépendante du facteur σ^B. Néanmoins, les auteurs supposent que Srn_1550_Teg49 serait un dérivé du transcrit *sarA* P3. Srn_1550_Teg49 régule l'expression de *sarA* en agissant en *trans* et agirait comme un stabilisateur du transcrit P3 de *sarA* (Kim et al., 2014).

SarA est une protéine d'une masse moléculaire d'environ 14,5kDa (124 acides aminés). L'analyse de la structure par cristallographie de la protéine SarA a permis de confirmer la structure en « winged-helix » (Liu et al., 2006; Schumacher et al., 2001). De plus, la formation en dimère de la protéine SarA est essentielle pour la liaison à l'ADN (Rechtin et al., 1999; Schumacher et al., 2001).

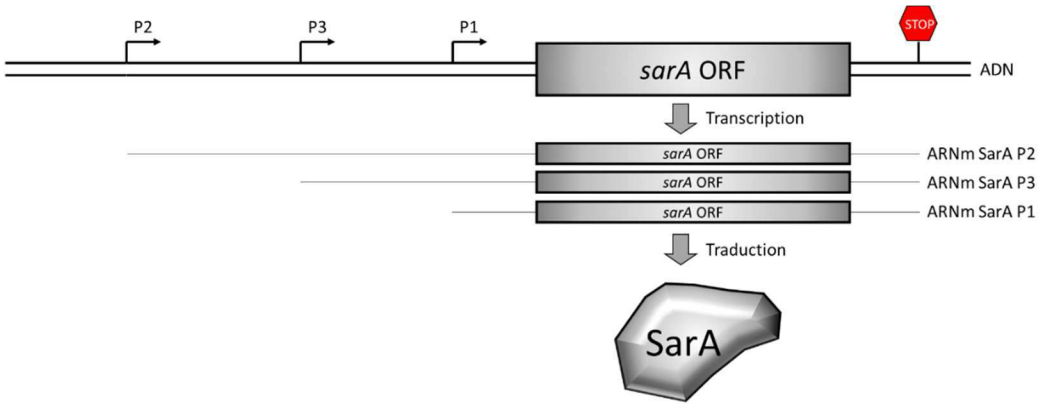


Figure 25 : Organisation génétique du gène *sarA* sous contrôle de ses trois promoteurs.

b. SarA, architecte de la chromatine

La détermination de la structure tridimensionnelle du complexe SarA-ADN a permis de montrer que la région « hélice-coude-hélice » fixe le grand sillon de l'ADN alors que la partie « ailée » se lie au petit sillon (Liu et al., 2006). Plusieurs études ont émis l'hypothèse de la fixation de plusieurs dimères de SarA sur une même région promotrice (Chien et al., 1999; Morfeldt et al., 1995; Rechtin et al., 1999) entraînant des changements conformationnels de la chromatine. Il a été proposé que SarA agirait comme une protéine architecturale du chromosome tout comme l'est la protéine HU d'*E. coli* (Schumacher et al., 2001).

c. Les cibles de SarA

Une étude a essayé de déterminer l'ensemble des cibles dont l'expression est impactée positivement ou négativement par SarA (Dunman et al., 2001). En utilisant la technique de GeneChip sur les ARN d'une souche sauvage et d'une souche mutante pour le gène *sarA*, les auteurs ont identifiés 76 cibles dont l'expression est activée par SarA et 44 gènes réprimés par ce facteur de transcription. Cette étude a permis de retrouver des cibles déjà connues pour être régulées par SarA (*agr, fnbB, hla, spa,* etc.) (Dunman et al., 2001). Il faut noter que l'ensemble de ces résultats présentent les cibles directes et indirectes de SarA, elle reflète donc une régulation globale de SarA pouvant impliquer d'autres régulateurs transcriptionnels.

Comme nous l'avons vu, la protéine SarA pourrait être une protéine architecturale de la chromatine. Chez *E. coli*, des protéines apparentées aux histones ont aussi un rôle sur la stabilité d'ARNm. Deux études ont étudié ce rôle pour la protéine SarA (Morrison et al., 2012; Roberts et al., 2006). En comparant la demi-vie d'ARNm, ils ont montré l'impact de la protéine SarA sur la stabilité d'ARNm. De plus, la fixation de SarA sur les ARNm *hla*, *spa* et *ssr42* a été validée par des expériences de retard sur gel ARN-Protéine (Morrison et al., 2012). L'hypothèse avancée est donc que SarA modifierait la stabilité de certains ARNm par fixation de la protéine sur ceux-ci.

d. SarA et l'antibiorésistance

L'antibiorésistance d'une bactérie est définie par sa capacité à développer des stratégies pour lutter contre l'action de certains antibiotiques.

SarA serait un acteur favorisant l'antibiorésistance de *S. aureus*. *De facto*, une souche délétée pour le gène *sarA* présente une plus forte sensibilité à une certaine classe d'antibiotique, les ß-lactamines. SarA apporterait un avantage pour la survie bactérienne chez l'hôte en cas de traitement à l'ajout d'oxacilline (Li et al., 2016; Trotonda et al., 2009). Le développement de cette résistance nécessite l'activation, par SarA, de *mecA* produisant ainsi

la protéine PBP2a (penicillin binding protein 2A) (Li et al., 2016). Cette protéine est une molécule capable de lutter contre l'action, entre autres, de la méticilline appartenant au groupe des ß-lactamines. FmtA, une autre protéine ciblant les antibiotiques, est activée par SarA (Zhao et al., 2012). La fixation de SarA sur le promoteur de *fmtA* a été confirmée, reliant directement l'action de SarA à l'antibiorésistance. SarA agit également sur l'expression d'*abcA* dont la protéine participe à l'autolyse bactérienne et à l'antibiorésistance (Villet et al., 2014). SarA est également requise pour permettre la résistance aux antibiotiques (vancomycine et ciprofloxacine) (Lamichhane-Khadka et al., 2009).

La résistance à la méticilline peut être réduite par l'utilisation d'un détergent non-ionique comme le Triton X-100 (Komatsuzawa et al., 1995). Il a été montré que l'action de Triton X-100 entraîne une modification de la capacité de fixation de SarA par modification de la structure tridimensionnelle de la protéine (Mandal et al., 2016). Ainsi, l'effet négatif du Triton X-100 sur la résistance à la méticilline pourrait passer par des changements d'affinité de SarA pour l'ADN.

e. SarA et la résistance aux stress oxydants

Les métabolismes cellulaires engendrent la production de nombreuses molécules oxydantes telles que le peroxyde d'oxygène (H_2O_2), le radical hydroxyle (HO•) ou l'anion superoxide (O_2^-). Ces dérivés oxydants sont toxiques pour les cellules et sont utilisés par les neutrophiles et les cellules phagocytaires pour lutter contre les infections bactériennes et virales. *S. aureus* a développé différentes stratégies pour se défendre contre l'attaque de ces stress oxydants. La bactérie exprime des gènes (*trxAB, sodAM, katA,* etc.) capables de réduire ces agents oxydants et inactiver leur toxicité.

SarA réprime l'expression de trois de ces gènes (*trxB, sodA* et *sodM*) par fixation sur leur promoteur (Ballal and Manna, 2009b, 2010). Nous pourrions en déduire que la fixation de SarA sur leur promoteur est délétère lorsque *S. aureus* subit un stress oxydant. Il est important de noter que le choix de l'agent oxydant utilisé pour oxyder SarA est primordial. Ainsi, l'ajout de diamide à des cultures cellulaires permet la levée de répression de SarA sur les trois gènes étudiés. Le peroxyde d'hydrogène permet, en présence de SarA, la forte expression de *trxB* (Ballal and Manna, 2010) et le *tert*-butyl hydroperoxide (t-BOOH), quant à lui, est idéal pour permettre une expression importante de *sodA* (Ballal and Manna, 2009b).

f. SarA et le biofilm bactérien

Le biofilm bactérien désigne une communauté de microorganismes, présents dans une matrice riche en sucres et protéines extracellulaires, et qui se positionne sur une surface (Costerton et al., 1978). Cette structure tridimensionnelle est un moyen efficace de lutter contre

les variations de pH, d'osmolarité et l'action du système immunitaire et d'antibiotiques. Par exemple, les biofilms sont capables de survivre à des concentrations d'antibiotiques jusqu'à 1000 fois supérieures à celle utilisées pour éradiquer des bactéries issues de cultures planctoniques (Ceri et al., 1999). Le biofilm est un moyen de résistance aux antibiotiques.

Les protéines impliquées dans la formation du biofilm chez *S. aureus* sont les polysaccharides PNAG (poly-*N*-acteyl ß-1-6-glucosamine) codés par les gènes *icaADBC* et la protéine de surface Bap.(Cramton et al., 1999; Trotonda et al., 2005). A l'inverse, diverses protéases telles que l'auréolysine et les sérines ou cystéines protéases peuvent être exprimées pour disperser le biofilm, permettant la libération massive de bactéries dans l'organisme.

Valle *et al.* ont montré que le facteur de transcription SarA est indispensable à la formation de biofilm (Valle et al., 2003). SarA est un activateur transcriptionnel de l'opéron *icaADBC*. Ce facteur de transcription active également l'expression du gène *bap* qui produit une seconde protéine impliquée dans la formation de biofilm (Trotonda et al., 2005). Ainsi, SarA est un acteur majeur pour la production de biofilm bactérien en activant les gènes codant pour les protéines nécessaires à sa formation.

D'un autre côté, SarA réprime l'expression de gènes produisant les protéases qui permettent la dispersion du biofilm. Il a été ainsi montré que SarA réprime la production d'auréolysine et de sérine protéases (Karlsson and Arvidson, 2002; Oscarsson et al., 2006; Shaw et al., 2004; Tegmark et al., 2000). SarA empêche donc la dispersion du biofilm.

g. SarA, modulateur de la virulence de *S. aureus*

Nous avons vu que le système *agr* de « quorum-sensing » permet de participer à l'infection conjointe à une augmentation de la densité bactérienne. SarA régule l'expression des deux promoteurs d'*agr* permettant d'activer les transcrits P2 et P3 (ARNIII). SarA se fixe au niveau du promoteur commun de ces deux gènes permettant d'obtenir une conformation de la chromatine favorable à leur expression (Rechtin et al., 1999; Reyes et al., 2011).

De plus, SarA active la transcription de *hla* qui produit l'hémolysine α (Chien et al., 1999; Oscarsson et al., 2006). L'hémolysine α est une toxine qui induit la formation de pores sur les cellules environnantes (Menzies and Kourteva, 2000). SarA active également l'expression de deux autres toxines, Seb et Tsst-1 (Chan et al., 1998). Néanmoins, le rôle activateur de SarA sur *hla* et *tst* n'est pas retrouvé dans toutes les souches de *S. aureus* ce qui indiquerait que d'autres régulateurs sont impliqués dans l'expression de *hla* et *tst* (Andrey et al., 2015; Chan et al., 1998).

La protéine IsaB est fortement produite dans certaines conditions (culture anaérobie, internalisation dans des cellules épithéliales, présence de neutrophiles). Il a été montré que SarA réprimait l'expression d'*isaB* (Liu et al., 2015). Nous pouvons supposer que, comme dans le cas *sodM* ou *trxB*, c'est une modification post-traductionnelle de SarA qui est responsable de la levée de répression d'*isaB*.

h. Modifications post-traductionnelles de SarA

Les modifications post-traductionnelles des protéines sont des moyens, pour la bactérie, de s'adapter et produire des réponses rapides à l'arrivée de signaux extérieurs. Ces modifications correspondent soit à la modification (oxydation/réduction) du radical de certains résidus soit à l'ajout de différents groupes fonctionnels (acétate, phosphate, etc.) aux niveaux d'acides aminés spécifiques. SarA, comme d'autres membres de sa famille (MgrA, SarZ), est capable de subir différentes modifications post-traductionnelles. La cystéine présente dans la séquence protéique de ces protéines semble être le point d'ancrage de nombreuses modifications (Chen et al., 2006; Chen et al., 2009; Fujimoto et al., 2009). Le groupe fonctionnel thiol (-SH) de la cystéine est susceptible d'être oxydé en présence d'agents oxydants (peroxyde d'oxygène, radical hydroxyle, etc.). L'oxydation de cette cystéine provoque une dissociation du facteur de transcription de l'ADN.

SarA peut également être phosphorylée (ajout d'un groupement phosphate) grâce aux kinases Stk1 ou Stk2 ciblant respectivement des résidus thréonines et sérines (Didier et al., 2010). La phosphorylation de SarA entraîne une diminution de l'affinité de la protéine pour ces promoteurs cibles (Didier et al., 2010). Stk1 a la possibilité d'ajouter le groupement phosphate sur la cystéine (Sun et al., 2012). Cette phosphorylation semble diminuer la capacité de SarA à fixer le promoteur *hla* (Sun et al., 2012). Cette modification post-traductionnelle est réversible par l'action de la phosphatase Stp1 (Zheng et al., 2015).

Les modifications post-traductionnelles de SarA permettent d'amplifier ses possibilités de régulation en ajoutant, au simple principe d'activation/répression, les nuances nécessaires aux réponses environnementales.

3. SarA et SarR, deux régulateurs antagonistes

Les études des différents membres de la famille SarA ont permis d'identifier SarR comme une protéine antagoniste du facteur de transcription SarA.

SarR est capable de fixer le promoteur de *sarA* afin de réprimer son expression (Manna and Cheung, 2001a). De plus, ces deux facteurs de transcription ont souvent des rôles inverses quant à l'expression de gènes. Ainsi, alors que SarA active l'expression du promoteur P2 du système *agr* en courbant la chromatine, SarR, en se substituant à SarA, réprime

l'expression en phase post-exponentielle de croissance (Reyes et al., 2011). SarA est décrite comme un répresseur des protéases de *S. aureus.* A l'inverse, SarR active la production des protéases comme la sérine protéase SspA et l'auréolysine Aur (Gustafsson and Oscarsson, 2008).

Malgré des séquences en acides aminés proches, les structures tridimensionnelles de SarA et SarR diffèrent. Il a été proposé que ces deux protéines reconnaissent une même séquence nucléotidique (Cheung and Zhang, 2001). La raison d'un rôle antagoniste, dans l'expression de gènes entre SarA et SarR, serait que le dimère de SarR aurait une structure plus rigide que le dimère de SarA (Cheung and Zhang, 2001).

V. Régulation transcriptionnelle d'ARN régulateurs

Nous avons vu qu'il existe une relation fonctionnelle entre ARN régulateurs et facteurs de transcription. Nous avons cité les exemples de Srn_1550_Teg49 qui régule la stabilité de l'ARN *sarA* ou encore l'ARN III qui agit sur les ARN *rot* et *mgrA*. De plus, nous savons que l'expression des gènes codant des ARN régulateurs doit être aussi régulée. Cependant, peu d'exemples sont décrits chez *S. aureus*. AgrA (TCS du « quorum-sensing ») réprime l'expression de l'ARNrég *srn_4050_artR* et active celle de l'*arnIII* (Reyes et al., 2011; Xue et al., 2014). Ces deux ARNrég, quant à eux, activent l'expression de *hla,* codant l'hémolysine α.

Ces exemples sont rares chez *S. aureus* et ils montrent pourtant l'intérêt d'étudier la régulation transcriptionnelle des ARN régulateurs. En effet, dans d'autres bactéries (*E. coli*, *Salmonella*,…), la recherche des facteurs de transcription impliqués dans la régulation de l'expression d'ARN régulateurs a permis d'intégrer ces derniers dans des fonctions bactériennes telles que la motilité et la formation de biofilm (Mandin and Guillier, 2013). De manière intéressante, des boucles de régulation ont été identifiées entre ARN régulateurs et système à deux composants. Par exemple, chez *E. coli*, le TCS EnvZ-OmpR régule l'expression de 4 gènes transcrits en ARNrég. Deux d'entre eux, OmrA/OmrB réprime l'expression du régulateur OmpR (Brosse et al., 2016).

Ces exemples, caractérisés dans d'autres bactéries, peuvent aider à la compréhension des régulations d'ARNrég et, comme vous le verrez, mon travail de thèse apporte des connaissances nouvelles sur la régulation transcriptionnelle d'ARN régulateurs chez *S. aureus* par un facteur de transcription.

CONTEXTE DE MA THESE

L'étude des ARN régulateurs chez *S. aureus* a commencé depuis une trentaine d'années. Elle a permis de mieux appréhender et de comprendre certains mécanismes de régulations de ce pathogène majeur de l'Homme. C'est dans ce contexte que le laboratoire de Biochimie Pharmaceutique (INSERM U1230) a découvert, en 2005, plusieurs ARN régulateurs localisés dans d'anciennes régions mobiles du génome, les îlots de pathogénie (Pichon and Felden, 2005). Depuis, l'étude de leurs fonctions a révélé plusieurs types de mécanismes impliquant la résistance aux antibiotiques, la production de peptides antimicrobiens ou de manière générale la virulence de ce procaryote.

Parmi les ARN régulateurs influençant la virulence, Le Pabic *et al.* a montré que l'ARN Srn_3610_SprC a un rôle atténuateur de la virulence de *S. aureus* (Le Pabic et al., 2015). Or, *srn_3610_sprC* est peu exprimé dans la bactérie dans des conditions de culture en laboratoire. Cette observation suggère que *S. aureus* a besoin de garder un potentiel virulent au cours de sa croissance. Mais, pourquoi *S. aureus* maintient la présence du gène *srn_3610_sprC* à travers l'évolution ? L'hypothèse émise est que la bactérie nécessite la production de Srn_3610_SprC à des moments précis lors d'arrivée de signaux. Il a été démontré que *srn_3610_sprC* est fortement exprimé dans les premières minutes de la phagocytose pour, ensuite, être rapidement réprimé (Le Pabic et al., 2015).

Mon sujet de thèse se base sur ces observations pour comprendre les mécanismes de régulation transcriptionnelle de *srn_3610_sprC.*

La première partie de mon travail a été d'étudier l'expression de *srn_3610_sprC*, dans différentes souches mutantes de *S. aureus*, afin d'identifier les régulateurs de son expression. Parmi les facteurs de transcription étudiés, nous avons mis en avant SarA, facteur de transcription majeur de *S. aureus*. SarA est un répresseur fort de la transcription de *srn_3610_sprC*. La répression de ce gène par SarA a été étudiée par des expériences de gènes rapporteurs, de retard sur gel et d'empreinte à la DNase I et nous avons ainsi pu localiser le site de fixation de SarA sur *srn_3610_sprC*. La recherche de séquences similaires, à ce site de fixation, dans le génome de *S. aureus* HG003 a révélé une séquence présente en amont d'un second gène transcrit en ARN régulateur, *srn_9340*. Nous avons donc, comme nous l'avions fait pour *srn_3610_sprC*, étudié l'action de SarA sur ce petit ARN.

La seconde partie s'appuie sur l'observation d'une surexpression de *srn_3610_sprC* aux premières minutes de la phagocytose. Nous avons recherché des signaux permettant la dérépression effectuée par SarA sur l'expression de *srn_3610_sprC* et *srn_9340*. Pour cela, nous avons utilisé des plasmides porteurs du gène codant pour une protéine fluorescente,

dont l'expression est sous le contrôle du promoteur de *srn_3610_sprC* (P_{sprC}) ou *srn_9340* (P_{srn_9340}). Nous avons ainsi pu suivre l'activité des promoteurs lorsque différents types de stress étaient appliqués à la bactérie. L'identification du ou des stress entrainant une forte augmentation de l'activité des promoteurs permettrait de nous aiguiller sur les conditions où SarA ne réprime plus l'expression des deux gènes d'intérêt et donc sur les moments ou ces deux ARNrég sont utiles à la bactérie.

Enfin, grâce à la littérature et aux données obtenues lors de cette thèse, nous avons, dans une troisième partie, défini une séquence consensus de la fixation de SarA sur les promoteurs de gènes. L'identification de cette séquence nous a permis de révéler de nouvelles cibles de SarA, que SarA agisse en tant que répresseur ou activateur de l'expression

RESULTATS ET DISCUSSIONS

PARTIE I :

Régulation de l'expression de deux ARN régulateurs par le facteur de transcription SarA

RESULTATS ET DISCUSSIONS

En 2005, plusieurs ARN régulateurs ont été localisés dans des éléments mobiles du génome (Pichon and Felden, 2005). L'un d'entre eux, Srn_3610_SprC, joue un rôle atténuateur de la virulence de *S. aureus* et influence la phagocytose de la bactérie par les macrophages humains (Le Pabic et al., 2015). Par ailleurs, son expression semble variable lors de la phagocytose. L'étude de la régulation de *srn_3610_sprC* est d'autant plus importante que l'ARN intervient dans le pouvoir infectieux de *S. aureus* et sur les mécanismes de défenses de l'hôte.

I. Caractérisation de l'ARN Srn_3610_SprC

A. Localisation de *srn_3610_sprC*

Srn_3610_sprC est positionné entre les gènes codant pour la leucocidine *lukE* et pour une protéine hypothétique *saouhsc_01956*. Le transcrit identifié au préalable par RACE (Rapid amplification cDNA ends) dans la souche Newman (Le Pabic et al., 2015) a d'abord permis d'identifier, par homologie de séquence, ses coordonnées chromosomiques dans la souche *S. aureus* NCTC8325 (taxid : 93061). Dans cette dernière, le gène *srn_3610_sprC* est localisé sur le brin - aux coordonnées 1861438-1861584 (Figure 26). La souche HG003, utilisée dans cette thèse, est issue de la souche NCTC8325 (99,97% d'identité) et la séquence du gène *srn_3610_sprC* est conservé à 100% entre ces deux souches. Ce gène, composé de 147pb, est présent en une copie au niveau de l'îlot génomique vSaß. Nous avons étudié cet îlot vSaß dans les différentes souches de *S. aureus*. Nous avons conclu que la présence de s*rn_3610_sprC* est dépendante de la présence de l'îlot génomique de types I ou II. Par exemple, la souche MRSA 252 (type III) ne contient pas ce gène.

Enfin, la région promotrice de *srn_3610_sprC* (P_{sprC}) a été définie comme la région intergénique entre *srn_3610_sprC* et *saouhsc_01956* (Figure 26), elle est d'une taille de 144 nucléotides.

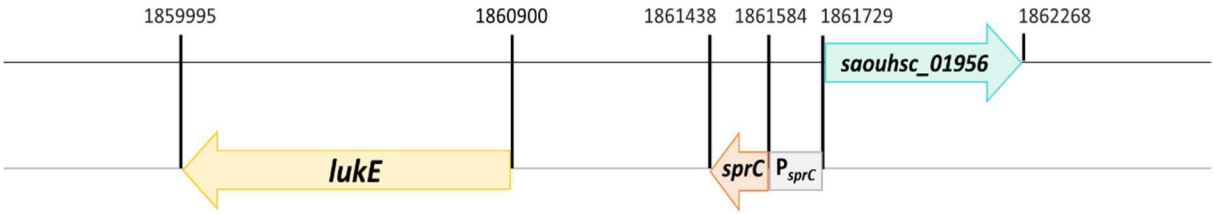


Figure 26 : Localisation génomique du gène *srn_3610_sprC* dans la souche *S. aureus* NCTC8325.

B. Séquence de *srn_3610_sprC* et de son promoteur

Les séquences géniques et promotrices de *srn_3610_sprC* de l'ensemble des souches de *S. aureus* disponibles dans la base de données NCBI ont été extraites et analysées (96 souches de *S. aureus*). *Srn_3610_sprC* est présent dans 82% des souches soit dans 79 souches sur les 96 répertoriées. Cette analyse a également montré que la séquence du gène *srn_3610_sprC* est hautement conservée dans les 79 souches. Ainsi, sur les 147 nucléotides qui composent le gène, seulement 10 points de mutations sont recensés. De plus, certains nucléotides sont systématiquement et conjointement mutés et pourraient se compenser (mutations compensatoires), c'est-à-dire que ces mutations ne modifieraient pas la structure secondaire de l'ARN (Figure 27). Au final, cette analyse montre peu de mutations pour le gène *srn_3610_sprC*.

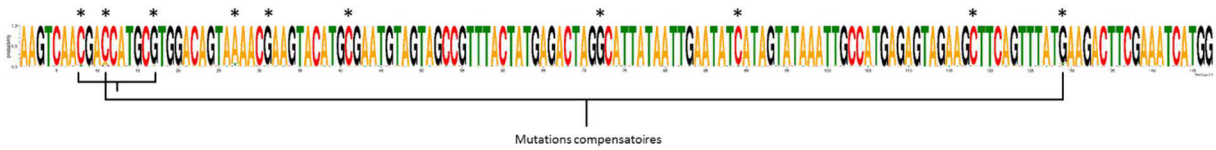


Figure 27 : Représentation des séquences de *srn_3610_sprC* des 79 souches de *S. aureus*. Les points de mutations sont marqués par un astérisque (*). La représentation a été réalisée par le logiciel WebLogo.

Après analyse de la séquence génique de *srn_3610_sprC* par ORFFinder, *srn_3610_sprC* ne présenterait pas de cadre ouvert de lecture putatifs permettant la production de peptides.

Les 144 nucléotides du promoteur de *srn_3610_sprC* présentent une plus grande variabilité : 24 points de mutations sont identifiés dans les 79 souches étudiées. Ils semblent répartis de manière aléatoire sur le promoteur (répartition régulière). Il est intéressant de noter qu'aucun point de mutation n'est localisé dans les 30 derniers nucléotides, correspondant à la partie essentielle du promoteur pour l'expression du gène.

Par homologie avec les séquences consensus des boîtes -10 et -35 d'*E. coli*, nous avons identifié ces dernières dans le promoteur de *srn_3610_sprC*. De plus, l'utilisation de la méthode d'identification globale des +1 de transcription des gènes de *S. aureus* a confirmé la localisation des boîtes -10 et -35 sur le promoteur P_{sprC} (Prados et al., 2016). Seize nucléotides séparent ces deux boîtes ce qui est cohérent avec les 17±1 bases d'espacement compris dans la séquence consensus d'*E. coli*. L'ensemble des caractéristiques des boîtes -10 et -35 de *srn_3610_sprC* permet de penser que la sous unité σ de l'ARN polymérase impliquée dans la transcription de *srn_3610_sprC* est la sous-unité σ^A, homologue de σ^{70} d'*E. coli* (Figure 28) (Deora and Misra, 1996b).

srn_3610_sprC (promoteur + gène)

1861730-*AAGTGTATAATATCAATTTACTACCTCGCAATATAATTTTATACAATGTATAATAAGATTTAAAG*
*ATGTAGCTCATCAACATTTTCATAGGGGATTTTTAAAATTAAGTA**TTGAAA**AATAAAATATTTATAT**TATAAT***
*ATAAAT**A**AGTCAACGACCATGCGTGGACAGTAAAACGAAGTACATGCGAATGTAGTAGCCGTTTACTATGAGA*
CTAGGCATTATAATTGAATATCATAGTATAAATTGCCATGAGAGTAGAAGCTTCAGTTTATGAAGACTTCGAA
ATCATGG-1861438

Figure 28: Séquences nucléotidiques du promoteur (P_{sprC}) et du gène *srn_3610_sprC.* Le promoteur est indiqué en italique. Les boîtes -10 et -35 ont été respectivement identifiées en rouge et vert. Le nucléotide +1 de transcription est marqué en gras et souligné.

C. Structure secondaire de l'ARN

La structure secondaire de l'ARN Srn_3610_SprC a été prédite par le logiciel MFold-RNA Folding Form (Figure 29). Comme la plupart des ARN régulateurs, Srn_3610_SprC est très fortement structuré avec 75% de bases impliquées dans la formation d'ARN double brin (structure en tige). Même si cette structure secondaire reste prédite, elle rejoint des précédents résultats de structure de l'ARN Srn_3610_SprC obtenues au laboratoire (Le Pabic et al., 2015). La structure de Srn_3610_SprC se terminerait par une structure tige-boucle. Cette conformation terminale semble compatible avec une terminaison de transcription Rho-indépendante.

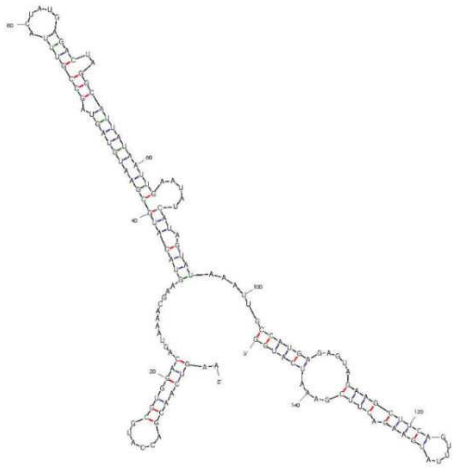


Figure 29 : Structure de l'ARN Srn_3610_SprC. La structure est prédite par le logiciel Mfold. (ΔG estimé = -34,80 kcal/mol)

II. Recherche de régulateurs de la transcription de *srn_3610_sprC*

Dans le but d'identifier les acteurs impliqués dans la transcription de *srn_3610_sprC*, nous avons fait le choix d'étudier l'expression de cet ARN dans différents contextes génétiques mutants. La plupart des souches mutantes étudiées proviennent de la souche HG003. Soit les mutations ont été réalisées directement dans la souche HG003, soit elles étaient présentes dans une autre souche de *S. aureus* puis ont été transduites dans le contexte HG003. Lorsqu'il n'a pas été possible de travailler avec un contexte génétique HG003, l'expression de

srn_3610_sprC a été comparée entre la souche parentale et la souche mutée (par exemple, les souches 8325-4 et 8325-4Δ*rot*).

Chez *S. aureus,* les facteurs de transcription de la famille SarA représentent la plus grande famille de facteurs de transcription et sont les plus décrits dans la littérature (Cheung et al., 2008b). Nous nous sommes donc intéressés à l'expression de *srn_3610_sprC* en absence de l'un de ces facteurs de transcription.

A. Expression de *srn_3610_sprC* dans différents contextes mutants

Nous avons étudié, par Northern Blot, la quantité d'ARN Srn_3610_SprC présente dans des souches mutées pour les gènes *sarA*, *sarR*, *sarS*, *sarT*, *sarV*, *sarX* ou *rot*. Les ARN totaux des différentes souches ont été extraits après 5 heures de croissance en milieu BHI à 37°C. Après migration sur gel d'acrylamide, les ARN Srn_3610_SprC et ARNtm ont été révélés par des sondes spécifiques (Figure 30).

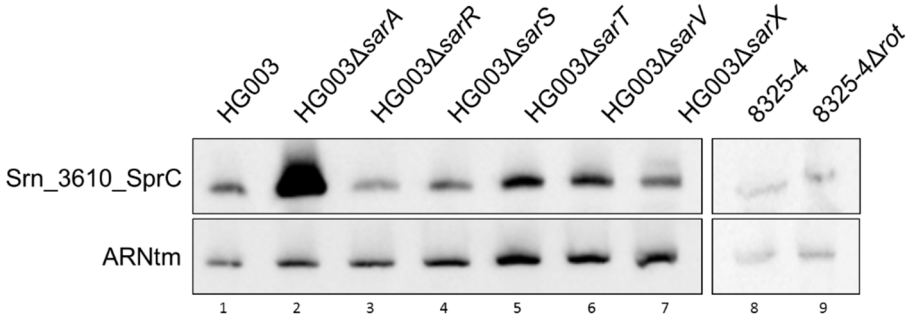


Figure 30 : Expression du gène *srn_3610_sprC* dans différentes souches mutantes pour l'un des facteurs de transcription de la famille SarA. Les ARN totaux ont été extraits après 5 heures de croissance et les ARN Srn_3610_SprC et ARNtm ont été révélés par des sondes spécifiques (Tableau 11).

Pour identifier les acteurs de la régulation transcriptionnelle de *srn_3610_sprC*, nous avons quantifié le signal pour Srn_3610_SprC et l'ARNtm dans chaque contexte génétique. L'ARNtm est l'ARN contrôle et nous sert à normaliser la quantité d'ARN Srn_3610_SprC.

Le niveau basal de Srn_3610_SprC correspond à la quantité d'ARN mesurée dans la souche parentale HG003 (Figure 30 ; puits 1). Ce niveau est comparé avec les quantités observées dans les différentes souches. Ainsi, la comparaison des quantités de Srn_3610_SprC observées à la Figure 30 dans les puits 4 à 7 (souches mutées pour *sarS, sarT, sarV* ou *sarX)* avec ceux du puits 1 ne révèle aucune différence significative. Les facteurs de transcription SarS, SarT, SarV et SarX ne semblent pas impliqués dans la production de l'ARN d'intérêt.

De plus, en comparant les niveaux d'expression de *srn_3610_sprC* des puits 8 (8325-4) et 9 (8325-4Δ*rot*), nous ne notons pas de différence de la quantité de Srn_3610_SprC. Le facteur de transcription Rot n'aurait pas d'effet sur l'expression de *srn_3610_sprC.*

La comparaison des puits 1 et 3 (respectivement HG003 et HG003Δ*sarR*) montre que SarR a un effet activateur sur l'expression du gène d'intérêt. L'absence de SarR entraîne une diminution du niveau à 0,7 fois la quantité d'ARN de référence.

A l'opposé, au puits 2 (HG003Δ*sarA*), nous constatons que la quantité de Srn_3610_SprC est très fortement augmentée en absence de la protéine SarA. Cette augmentation d'un facteur 10 (comparée à la quantité d'ARN Srn_3610_SprC dans le puits 1) indique un rôle répresseur très fort du facteur de transcription SarA sur l'expression de *srn_3610_sprC*.

A ce stade, il est difficile de savoir si ces deux régulateurs (SarR et SarA) agissent directement ou indirectement. De plus, ces régulateurs sont connus pour avoir des interconnections (Cf. Introduction page 51). Il a été montré que SarR réprime l'expression de *sarA* (Manna and Cheung, 2001b). De ce fait, en absence de SarR, une plus grande quantité de SarA est produite dans la cellule ce qui pourrait conduire à une diminution de la transcription de *srn_3610_sprC* comme observée (Figure 30 ; puits 3).

En plus des facteurs de transcription de la famille SarA, nous disposions au laboratoire deux souches HG003 mutantes, la première mutée pour l'ARN régulateur ARNIII et la seconde pour le facteur alternatif σ^B. Il est intéressant d'étudier la quantité d'ARN Srn_3610_SprC dans un contexte où l'ARNIII est absent. En effet, l'ARNIII est un régulateur majeur chez *S. aureus*, capable d'agir sur des facteurs de transcription (Rot ou MgrA) ou sur des facteurs de virulence (Spa, Hla, …) (Cf. Introduction page 27). Dans un second temps, nous avons étudié la souche HG003Δ*sigB*. L'étude du promoteur nous a permis de supposer que la sous-unité σ^A est le facteur σ le plus probable recruté par l'ARN polymérase pour la transcription de *srn_3610_sprC*. Nous savons qu'en cas de stress, le facteur σ^B, libéré de l'anti-facteur SigW, est recruté sur certains promoteurs et entraîne leur transcription. Nous avons, grâce à la souche mutante HG003Δ*sigB*, voulu vérifier que dans les conditions de culture standard, σ^B n'intervient pas dans la transcription de *srn_3610_sprC*.

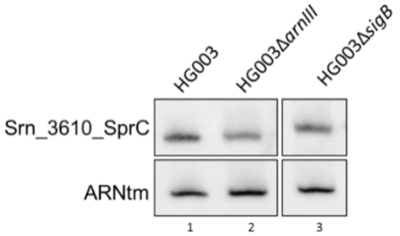


Figure 31 : Expression du gène *srn_3610_sprC* dans les souches mutantes HG003Δ*arnIII* ou HG003Δ*sigB*. Les ARN Srn_3610_SprC et ARNtm (contrôle de dépôt) ont été révélés avec des sondes spécifiques (Tableau 11).

A la Figure 31, la comparaison des quantités de Srn_3610_SprC (après normalisation avec l'ARNtm) révélées

dans les souches HG003 (puits 1), HG003Δ*arnIII* (puits 2) ou HG003Δ*sigB* (puits 3) n'a pas montré de différence significative. Ni l'ARNIII, ni la sous-unité σ^B ne semblent impliqués dans la régulation de la transcription de *srn_3610_sprC*.

Dans les conditions conventionnelles de croissance, la transcription de *srn_3610_sprC* est faible, il paraît alors difficile d'identifier un activateur par son absence. En effet, la délétion d'un activateur conduirait à un niveau de l'ARN Srn_3610_SprC encore plus faible. C'est pourquoi, nous avons souhaité étudier l'effet de l'absence des facteurs de transcription de la famille SarA dans une souche HG003Δ*sarA*. Pour cela, des double-mutants ont été construits. En absence de SarA, la quantité de Srn_3610_SprC est augmentée de 10 fois. Si nous délétons, dans ce contexte, le gène de l'activateur, nous nous attendons à voir le niveau de l'ARN Srn_3610_SprC diminuer.

Nous avons regardé la quantité d'ARN Srn_3610_SprC dans des souches doublement mutantes, mutante pour *sarA* et mutante pour un deuxième facteur de transcription de la famille SarA ou pour l'ARN régulateur ARNIII (Figure 32 et Figure 33).

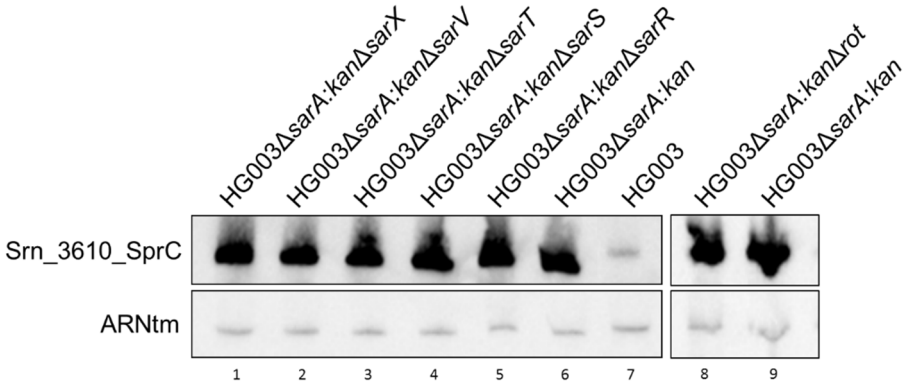


Figure 32 : Quantité d'ARN Srn_3610_SprC dans les souches doublement mutées pour *sarA* et un autre facteur de transcription. Les ARN ont été extraits à 5 heures de croissance. L'ARNtm sert de contrôle de dépôt.

La quantité d'ARN Srn_3610_SprC dans les souches doublement mutantes est comparée avec celle présente dans les souches HG003Δ*sarA* (Figure 32 ; puits 6) et HG003 (puits 7). Pour les souches doublement mutées HG003Δ*sarA*Δ*sarR,* HG003Δ*sarA*Δ*sarS,* HG003Δ*sarA*Δ*sarT,* HG003Δ*sarA*Δ*sarV*, HG003Δ*sarA*Δ*sarX* (puits 1 à 5) et HG003Δ*sarA*Δ*rot* (puits 8), la quantité d'ARN Srn_3610_SprC révélée est identique à celle de la souche HG003Δ*sarA* (puits 6). Nous pouvons dire que, parmi les facteurs de transcription de la famille SarA étudiés, seule SarA semble impliquée dans l'expression de *srn_3610_sprC*.

De plus si nous analysons le résultat obtenu pour la souche HG003Δ*sarA*Δ*sarR,* cette expérience montre que la diminution de l'expression de *srn_3610_sprC* observée dans la souche HG003Δ*sarR* (Figure 30 puits 3) nécessite la présence de SarA : cela renforce la

proposition d'un effet indirect de SarR. SarR réprime *sarA*, la souche HG003Δ*sarR* présente donc une quantité de protéine SarA plus importante, *srn_3610_sprC* est alors encore plus réprimé.

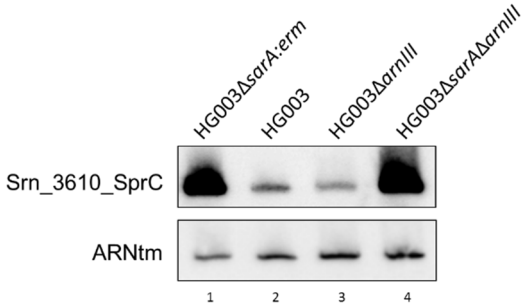


Figure 33 : Expression de *srn_3610_sprC* en absence de SarA et/ou l'ARNIII. Les ARN Srn_3610_SprC et ARNtm (contrôle de dépôt) ont été révélés avec des sondes spécifiques sur des ARN totaux extraits après 5 heures de croissance.

En comparant la quantité d'ARN Srn_3610_SprC entre la souche HG003Δ*sarA* (puits 1) et la souche HG003Δ*sarA*Δ*arnIII* (puits 4) de la Figure 33, nous observons une légère augmentation (x1,2) dans la souche HG003Δ*sarA*Δ*arnIII.* L'absence de l'ARNIII et de SarA entraîne une production de Srn_3610_SprC augmentée. Cet effet positif de l'absence de l'ARNIII n'est visible que dans la souche HG003Δ*sarA*, ce qui indique un effet indirect de l'ARNIII. Cet ARN régulateur pourrait activer un répresseur dont l'effet serait masqué par le répresseur SarA. Une autre hypothèse serait que l'ARNIII réprimerait un activateur lequel pourrait agir seulement en absence de SarA.

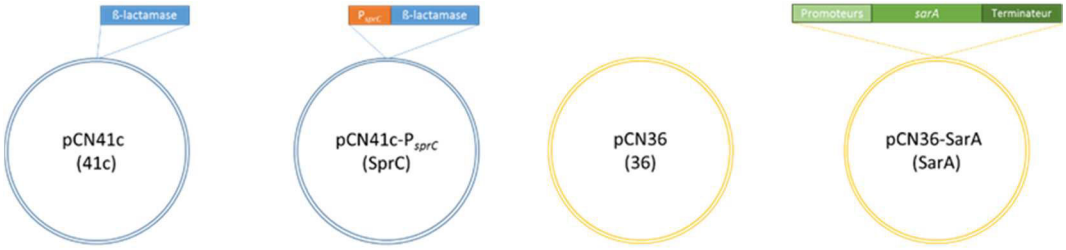
L'ensemble des expériences de Northern Blot a identifié le facteur de transcription SarA comme répresseur fort de la transcription de *srn_3610_sprC*. De ce fait, nous nous sommes focalisés sur l'étude du mécanisme de régulation de *srn_3610_sprC* par SarA.

B. Action de SarA sur le promoteur de Srn_3610_SprC

Afin de vérifier que, dans la souche mutante HG003Δ*sarA*, l'absence de SarA est responsable de la surexpression de *srn_3610_sprC*, des expériences de complémentation fonctionnelle ont été réalisées. Cette technique consiste à réintroduire le gène *sarA* grâce à un plasmide ou par réintégration au locus chromosomique. Nous avons opté pour une complémentation fonctionnelle plasmidique en *trans*. La souche HG003Δ*sarA* a été transformée par électroporation avec le plasmide pCN36-SarA contenant la région promotrice, le gène et le terminateur endogène de SarA. En parallèle, le plasmide vide (pCN36) a été introduit dans la souche HG003Δ*sarA*. Ces transformations ont également été effectuées dans la souche sauvage HG003.

Pour étudier l'activité du promoteur de *srn_3610_sprC* par l'utilisation de la technique du gène rapporteur, l'introduction d'un second plasmide a été nécessaire. Ainsi les 4 souches préparées ci-dessus (HG003 pCN36 ; HG003 pCN36-SarA ; HG003Δ*sarA* pCN36 ;

HG003Δ*sarA* pCN36-SarA) ont reçu soit le plasmide pCN41c, soit le plasmide pCN41c-P_{sprC}. Dans ce dernier, la production de la ß-lactamase (codée par le gène *blaZ*) est sous le contrôle du promoteur P_{sprC}.



Souches	HG003		HG003		HG003Δ*sarA*		HG003Δ*sarA*	
	1	2	3	4	5	6	7	8
Plasmide	36	36	SarA	SarA	36	36	SarA	SarA
Contexte SarA	Sauvage		Surexpression de SarA		Délétion de SarA		Complémentation de SarA	
Plasmide	41c	SprC	41c	SprC	41c	SprC	41c	SprC
Activité ß-lactamase	Non	Oui	Non	Oui	Non	Oui	Non	Oui

Figure 34 : Récapitulatif des combinaisons plasmidiques transformées dans les souches HG003 et HG003Δ*sarA* pour l'expérience de complémentation fonctionnelle.

Au final, huit combinaisons plasmidiques ont été obtenues, elles sont résumées à la Figure 34. Dans les souches porteuses des combinaisons 1 ou 2, le gène *sarA* n'est présent qu'une seule fois à son locus chromosomique. Ces souches expriment donc un niveau sauvage de SarA. Les souches ayant les combinaisons 3 ou 4 possèdent la copie génomique de *sarA* et le plasmide porteur du gène *sarA*. Nous nous attendons à observer une surexpression de *sarA* dans ces souches. Les souches délétées du gène *sarA* (combinaison 5 ou 6) ne produisent pas la protéine SarA. Enfin, les souches avec les combinaisons 7 ou 8 ont perdu la copie chromosomique de *sarA* mais portent ce gène sur un plasmide (complémentation fonctionnelle plasmidique en *trans*).

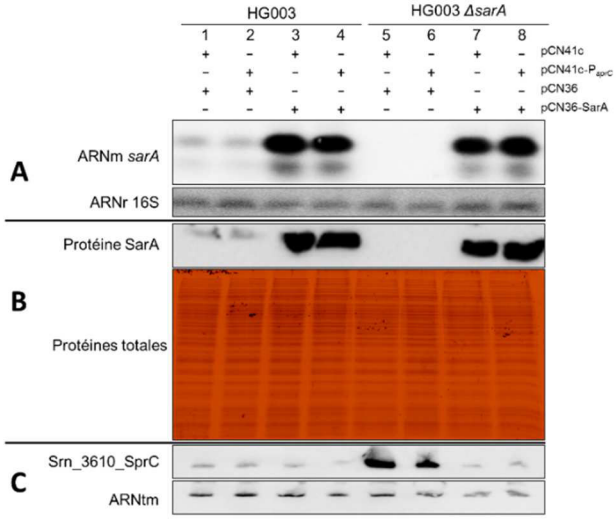


Figure 35 : Expression de *srn_3610_sprC* et *sarA* dans les différentes souches utilisées pour les expériences de complémentation. Les ARN Srn_3610_SprC, SarA, ARNtm et l'ARN ribosomal 16S ont été détectés par Northern Blot et la protéine SarA est révélée par Western Blot en utilisant un anticorps polyclonal anti-SarA. Les protéines totales, révélées au SyproRuby, permettent de contrôler l'uniformité de la quantité de protéines déposées.

Dans un premier temps, nous avons comparé les niveaux d'ARN *sarA* ou de protéines SarA dans les différentes souches. Les niveaux de référence sont ceux observés dans les combinaisons 1 et 2 (Figure 35 A et B). Ensuite, nous avons comparé le niveau d'ARN Srn_3610_SprC entre les souches. Les souches 1 et 2 nous donnent, là aussi, les niveaux de référence pour Srn_3610_SprC (Figure 35 C).

La présence du gène *sarA* plasmidique dans la souche HG003 (puits 3 et 4) entraîne une augmentation de la quantité de transcrits *sarA* laquelle est suivie d'une augmentation de la production de la protéine SarA. L'absence de transcrits et de protéine SarA au niveau des puits 5 et 6 valide la délétion chromosomique de *sarA* dans la souche HG003. La réintroduction d'une copie plasmidique de *sarA* dans une souche HG003Δ*sarA* (puits 7 et 8) entraîne la présence de l'ARN *sarA* et de la protéine. Les niveaux observés sont comparables à ceux observés pour les contextes 3 et 4 c'est-à-dire que les niveaux transcriptionnels de *sarA* sont forts et la quantité de protéine SarA est également augmentée. L'introduction du plasmide pCN36-SarA dans la souche HG003Δ*sarA* a restauré la production de la protéine SarA. Nous pouvons même considérer que nous observons une surexpression de SarA lorsque le gène de celle-ci est porté par un plasmide.

Quant à l'ARN Srn_3610_SprC, son niveau est dépendant de la quantité de SarA. En effet, alors que l'absence de SarA (Figure 35C ; puits 5 et 6) conduit à une augmentation de la quantité d'ARN Srn_3610_SprC, la réintroduction de SarA (puits 7 et 8) corrige ce phénomène et il n'existe plus de différence entre la quantité de Srn_3610_SprC observée dans les souches sauvages (puits 1 et 2) et celle observée dans les souches complémentées (puits 7 et 8). C'est donc bien l'absence de SarA qui est responsable de la surexpression de *srn_3610_sprC* observée dans la souche HG003Δ*sarA.* Par ailleurs, la surexpression de *sarA* (puits 3 et 4) entraîne une légère sur-répression de *srn_3610_sprC*.

Nous observons ici que l'expérience de complémentation fonctionnelle a fonctionné puisque la réintroduction de *sarA* a révélé un niveau d'expression du gène *srn_3610_sprC* comparable à celui observé dans la souche sauvage HG003. Donc, dans la souche HG003Δ*sarA*, c'est bien l'absence de *sarA* qui est responsable de la dérépression transcriptionnelle de *srn_3610_sprC.*

Les facteurs de transcription sont connus pour agir préférentiellement au niveau du promoteur d'un gène. Pour vérifier que l'action de SarA se situe sur le promoteur de *srn_3610_sprC*, nous avons réalisé des expériences de gène rapporteur en utilisant le gène *blaZ*, lequel code pour la ß-lactamase. Le gène rapporteur est présent dépourvu de son promoteur dans le plasmide pCN41c. Nous avons amplifié les 144 nucléotides du promoteur de *srn_3610_sprC* (P_{sprC}) pour les placer en amont du gène rapporteur. Ainsi, la production de

ß-lactamase est en corrélation avec l'activité du promoteur P_{sprC} (plus le promoteur est actif, plus l'enzyme ß-lactamase sera produite).

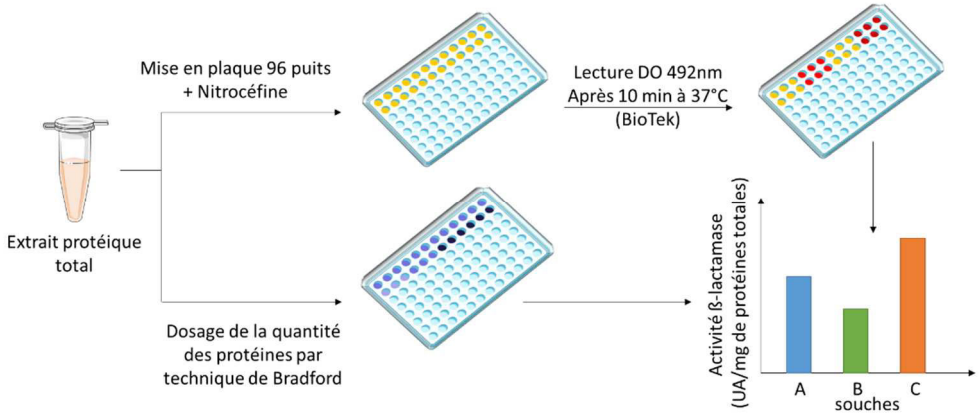


Figure 36 : Protocole de l'expérience de gène rapporteur (ß-lactamase)

Comme expliqué précédemment, les plasmides pCN41c-P_{sprC} et pCN41c (plasmide vide) ont été introduits par transformation dans les souches HG003 pCN36, HG003 pCN36-SarA, HG003Δ*sarA* pCN36 et HG003Δ*sarA* pCN36-SarA. Les protéines totales des différentes souches ont été extraites. L'activité de l'enzyme est mesurée à 492 nm en présence de nitrocéfine et normalisée avec la quantité de protéines totales extraites (Figure 36).

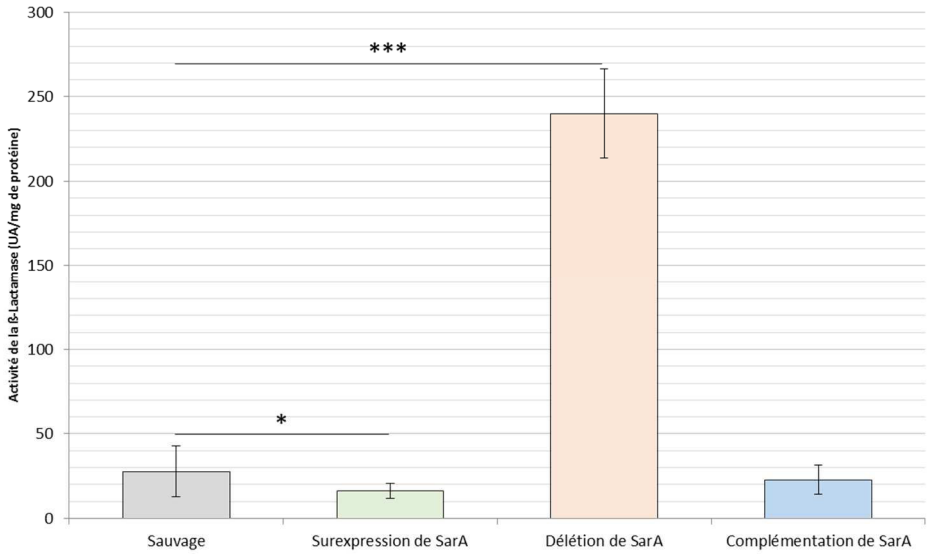


Figure 37 : Mesure de l'activité du promoteur P_{sprc} en fonction de la quantité de protéine SarA présente. L'activité de la ß-lactamase a été mesurée dans les souches HG003 (sauvage), HG003 + pCN36-SarA (surexpression de SarA), HG003Δ*sarA* (délétion de SarA) et HG003Δ*sarA* + pCN36-SarA (complémentation de SarA). La moyenne de l'activité de la ß-lactamase de trois expériences indépendantes a été calculée (Test de Mann-Whitney : * P<0,05 ; *** P<0,001)

En comparant les niveaux d'expression de la ß-lactamase dans les souches HG003 pCN36 (niveau endogène de SarA) et la souche HG003 pCN36-SarA (surexpression de SarA), nous observons une diminution significative de 2,5 fois de l'activité de la ß-lactamase ($p<0,017$) (Figure 37). Dans la souche HG003Δ*sarA* pCN36 (délétion de SarA), le promoteur est environ neuf fois plus actif par rapport à la souche sauvage ($p<0,0006$). La dérépression de l'expression de *srn_3610_sprC* dans le mutant HG003Δ*sarA* est également observée dans cette expérience. La réintroduction, dans la souche HG003Δ*sarA,* du gène *sarA* porté par le plasmide permet de diminuer l'activité du promoteur P_{sprC} si nous la comparons à celle observée dans la souche HG003Δ*sarA*. De plus, les activités de P_{sprC} dans la souche sauvage HG003 et complémentée (HG003Δ*sarA* + pCN36-SarA) ne sont pas significativement différentes. Ces résultats sont en accord avec les expériences de Northern Blot sur l'ARN Srn_3610_SprC (Figure 35C).

Dans cette expérience, la réintroduction de *sarA* dans la souche mutante HG003Δ*sarA* a corrigé le phénotype observé en absence de *sarA.* Dans la souche HG003Δ*sarA*, l'absence de SarA est responsable de la dérépression de *srn_3610_sprC*. Ces expériences de gène rapporteur (dont l'expression est sous contrôle du promoteur de *srn_3610_sprC*) montrent que cette dérépression transcriptionnelle implique le promoteur P_{sprC}. La complémentation fonctionnelle est également validée et confirme l'action de SarA sur l'expression de *srn_3610_sprC*.

C. Lien entre SarA et le promoteur de *srn_3610_sprC*

Pour savoir si SarA agit de manière directe sur le promoteur de *srn_3610_sprC* (P_{sprC}), des expériences de retard sur gel ont été réalisées. Pour cela, le promoteur de *srn_3610_sprC* a été amplifié par PCR à partir de l'ADN génomique de la souche HG003 et radiomarqué. En ajoutant la protéine SarA purifiée (de 6,5 à 65 pmol ; Figure 38 puits 2 à 7), nous constatons un retard de migration de la sonde P_{sprC}. SarA est donc capable de fixer le promoteur de *srn_3610_sprC.* Nous avons déterminé la constante de dissociation (K_D) pour la fixation de SarA sur P_{sprC}. Le K_D correspond à la concentration de SarA qui fixe la moitié de la quantité de sonde radiomarquée. Ici, la valeur du K_D a été mesurée à 2,8µM.

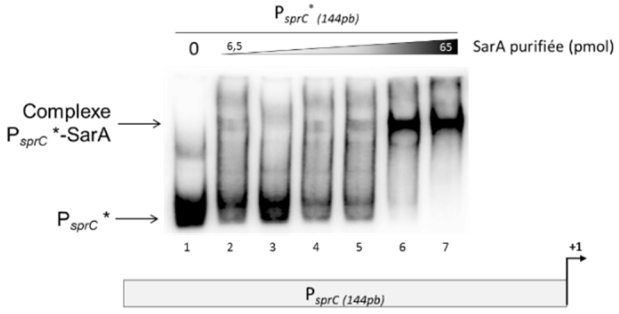


Figure 38 : SarA est capable de fixer *in vitro* le promoteur de *srn_3610_sprC* (P_{sprC}). 134 fmol de sonde ADN P_{sprC} radiomarquée sont mis en présence croissante de la protéine SarA purifiée (de 6,5 à 65 pmol). Dans le panel du bas, la sonde utilisée pour cette expérience est représentée.

Pour localiser le site de fixation de SarA sur le promoteur P_{sprC}, des expériences d'empreinte à la DNase I ont été effectuées. Cette technique consiste à faire agir une endonucléase sur un fragment d'ADN. Si une protéine se fixe sur l'ADN, l'endonucléase ne pourra pas agir et nous observerons une région protégée de la dégradation correspondant au site de fixation de la protéine.

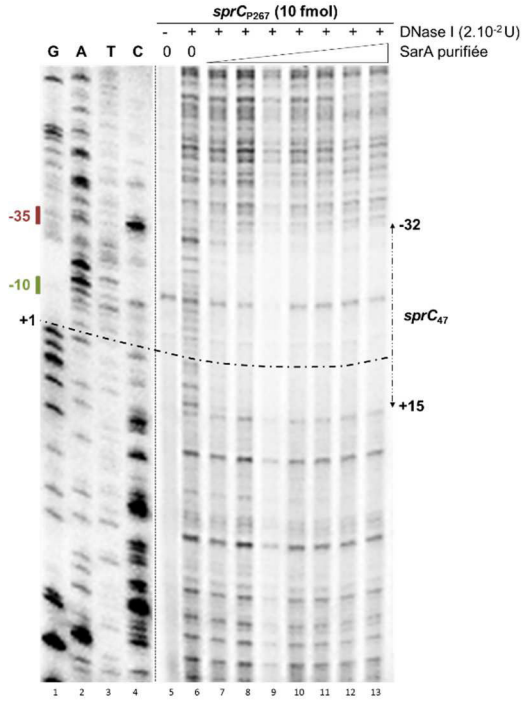


Figure 39 : SarA chevauche la promoteur et le gène de *srn_3610_sprC*. Les expériences d'empreinte à la DNase ont été réalisées avec 10 fmol de $sprC_{P267}$radiomarquée, une quantité croissante de SarA (0,2-5 pmol) et une quantité de DNase I (2.10^{-2} U). Le séquençage de la sonde est effectué dans les puits G, A, T et C. Les boîtes -10 et -35 ainsi que le +1 de transcription ont été placées en marge du séquençage. La zone protégée $sprC_{47}$ est représentée par une double-flèche.

Pour des raisons techniques, nous devions utiliser une sonde plus grande que le promoteur de *srn_3610_sprC* : une sonde ADN ($sprC_{P267}$) correspondant au promoteur de *srn_3610_sprC* (144pb) et une partie du gène (123pb) a donc été amplifiée et radiomarquée à son extrémité 5' ($sprC_{P267}$). Dans les puits 1 à 4 de la Figure 39, le séquençage de la sonde est présenté : ceci a permis, entre autres, de placer dans la marge le +1 de transcription ainsi que les boîtes -10 et -35. Dans le puits 6, nous observons la dégradation de la sonde $sprC_{P267}$ par la DNase I en absence de protéine. La quantité de DNase I a été choisie pour qu'il y ait statistiquement une coupure par fragment d'ADN. Ici, elle est de 2.10^{-2} U. La protéine SarA purifiée est ajoutée en quantité croissante (0,2 à 5 pmol) dans les puits 7 à 13 (Figure 39). La présence de SarA protège un site de la dégradation par la DNase I. Cette zone de 47 paires de bases recouvre les nucléotides allant de -32 au +15 par rapport au +1 de transcription de *srn_3610_sprC*. Parmi ces 47 pb, nous retrouvons la boîte -10 et une partie de la boîte -35, sites de fixation de l'ARN polymérase.

La première expérience de retard sur gel ayant été réalisée en utilisant un fragment d'ADN ne contenant que la région promotrice du gène *srn_3610_sprC*, il nous est apparu nécessaire de tester l'effet de la présence d'une partie du gène *srn_3610_sprC* sur la fixation

de la protéine SarA. Nous avons alors utilisé, dans cette deuxième série de retard sur gel, la sonde *sprC*$_{P267}$ utilisée dans les expériences d'empreinte à la DNase I.

Les expériences de retard sur gel ont été réalisées avec des quantités croissantes de SarA (0,1-8 pmol). Un retard de migration de la sonde *sprC*$_{P267}$ commence à apparaître avec 0,5 pmol de SarA (Figure 40 ; puits 4). Plus la quantité de SarA est forte, plus la quantité de sonde *sprC*$_{P267}$ retardée est importante. A 4 pmol (puits 9), l'intégralité de *sprC*$_{P267}$ est retardée. Cette expérience nous a permis de déterminer la constante de dissociation de la fixation de SarA sur *sprC*$_{P267}$. Ici, le K_D est de 86±3 nM.

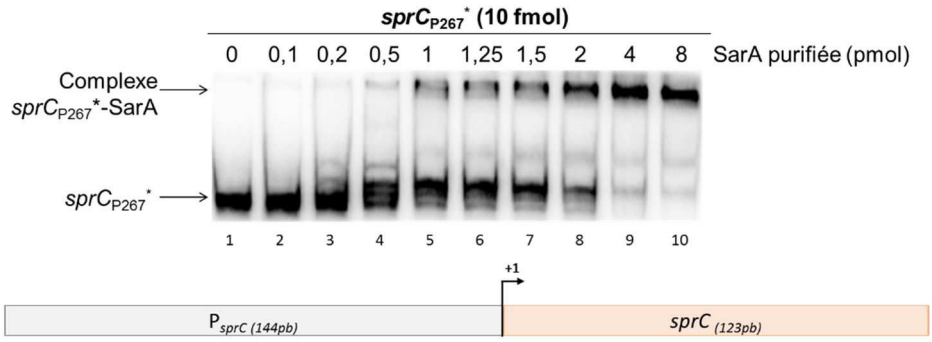


Figure 40 : SarA fixe *in vitro* la sonde *sprC*$_{P267}$. La sonde *sprC*$_{P267}$ (10 fmol) est mise en présence d'une quantité croissante de SarA (de 0,1 à 8 pmol : puits 2 à 10).

Précédemment, nous avions estimé le K_D de la fixation de SarA sur P_{sprC} (144 nt) à 2,8 µM. Nous observons ici que SarA a une meilleure affinité pour la sonde de 267 pb que pour le promoteur seul P_{sprC}. Ces résultats sont en accord avec les expériences d'empreinte à la DNase I qui montrent que le site de fixation de SarA chevauche la partie transcrite de *srn_3610_sprC*.

SarA fixe *in vitro* la sonde *sprC*$_{P267}$, nous avons voulu vérifier la spécificité de cette fixation.

Pour cela, nous avons réalisé une expérience de compétition en ajoutant à la réaction de fixation (SarA-*sprC*$_{P267}$*) une sonde *sprC*$_{P267}$ froide (compétiteur spécifique, Figure 41 ; puits 3 à 6) ou une sonde de 267pb correspondant au gène de l'ARNr 16S (compétiteur non spécifique, puits 7 à 10). L'ajout de 50X de compétiteur spécifique (puits 5) entraîne un relargage de SarA de la sonde radiomarquée (*sprC*$_{P267}$*) vers la sonde froide (*sprC*$_{P267}$) alors que 100X de compétiteur non spécifique (puits 10) ne provoque pas ce phénomène. Nous pouvons donc conclure que SarA fixe *in vitro* de manière spécifique la sonde *sprC*$_{P267}$.

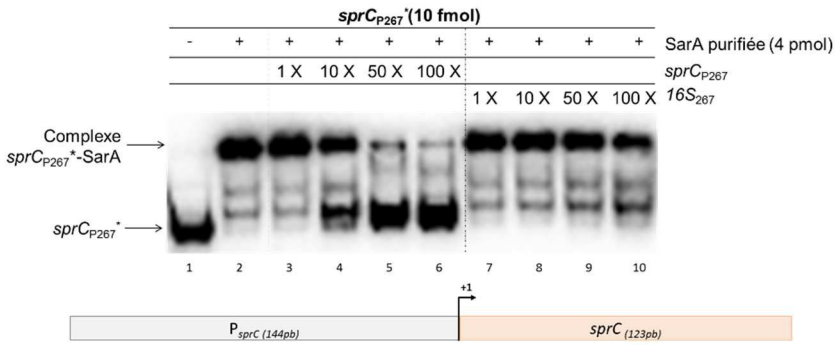


Figure 41 : SarA fixe *in vitro* la sonde *sprC*$_{P267}$ de manière spécifique. Des quantités croissantes de compétiteurs spécifiques (*sprC*$_{P267}$ non marquée) et non spécifiques (16S$_{267}$) sont ajoutés à la réaction de fixation (10 fmol de *sprC*$_{P267}$* / 4 pmol de SarA).

Les résultats des expériences d'empreinte à la DNase I ont permis de définir que la région de 47pb située entre -32 et +15 est impliquée dans la fixation de SarA. Afin de confirmer ces résultats, cette zone a été retirée de la sonde de 267pb pour générer une sonde ADN de 220pb, appelée *sprC*$_{P267\Delta47}$ (Figure 42). Dans les puits 2 à 5, une quantité croissante de protéine SarA est ajoutée à la sonde de 267 pb (*sprC*$_{P267}$). Un retard de migration de la sonde est observé à partir 0,5 pmol de SarA et la quasi-totalité de la sonde est complexée à la protéine à 2 pmol de SarA. Quelle que soit la quantité de SarA ajoutée à la sonde *sprC*$_{P267\Delta47}$ (puits 6 à 10), aucun retard de migration de la sonde n'est observé. Ainsi les 47 nucléotides, protégés de la dégradation de la DNase I par SarA, sont nécessaires à la fixation de SarA.

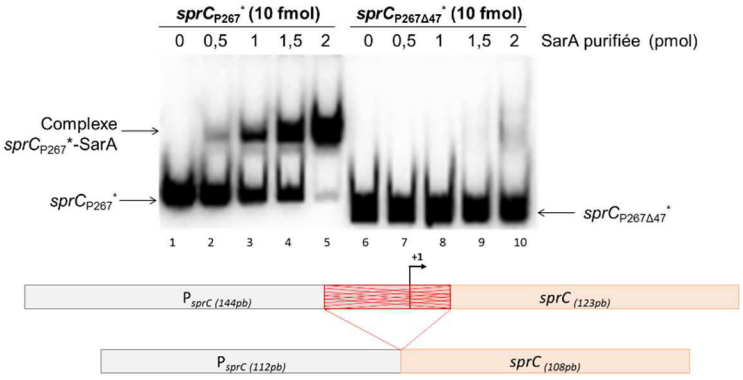


Figure 42 : la séquence *sprC*$_{47}$ est nécessaire à la fixation de SarA. 10 fmol de *sprC*$_{P267}$ ou *sprC*$_{P267\Delta47}$ (sonde délétée de *sprC*$_{47}$) sont mis en contact avec la protéine SarA (0,5-2 pmol).

Nous avons ensuite voulu savoir si la séquence de 47pb (*sprC*$_{47}$) est suffisante pour permettre la fixation de SarA. Cette séquence a donc été mise en contact avec la protéine SarA (Figure 43). La sonde *sprC*$_{47}$ a été retardée par la présence de SarA (puits 2). SarA est donc capable *in vitro* de fixer la séquence *sprC*$_{47}$. De plus, cette fixation est spécifique puisque l'ajout de 50X de compétiteur spécifique (*sprC*$_{47}$ froid, puits 4) déplace SarA de la sonde radiomarquée vers la sonde froide alors qu'une même quantité de sonde non spécifique ne le fait pas (Random$_{47}$, puits 5). Pour avoir un compétiteur non spécifique, nous avons créé une

sonde avec une séquence nucléotidique aléatoire mais dont le pourcentage en guanine et cytosine (%GC) est identique à *sprC*$_{47}$.

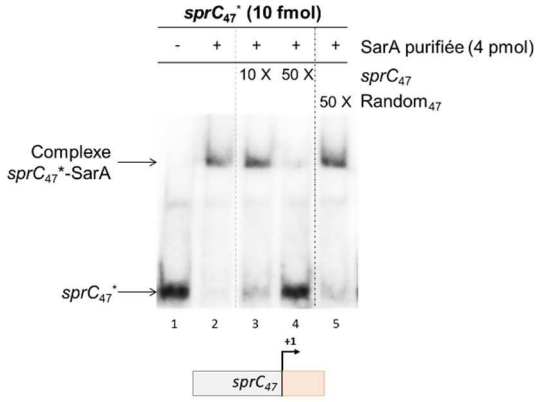


Figure 43 : la sonde *sprC*$_{47}$ est suffisante à la fixation de SarA. 10 fmol de *sprC*$_{47}$ radiomarquées sont mis en contact avec 4 pmol de la protéine purifiée SarA (puits 2 à 5). Les sondes froides *sprC*$_{47}$ (à 10X : puits 3 ou à 50X : puits 4) ou Random$_{47}$ (à 50X : puits 5) sont ajoutées pour tester la spécificité de la fixation.

Nous pouvons conclure que, *in vitro,* cette zone de 47pb, située dans le promoteur et le gène de *srn_3610_sprC*, est donc nécessaire et suffisante à la fixation spécifique de SarA.

Pour réduire la zone de fixation de SarA, trois sondes de 30pb chevauchantes entre elles et recouvrant les 47pb, ont été synthétisées et utilisées dans des expériences de retard sur gel. Les résultats présentés à la Figure 44 montrent que la protéine SarA est capable de fixer les sondes SprC$_{30Milieu}$ et SprC$_{30Droite}$ (4 pmol, puits 3 des différents gels). L'analyse des séquences SprC$_{30Milieu}$ et SprC$_{30Droite}$ permet de révéler une séquence commune de 22 nucléotides, susceptible d'être nécessaires à SarA pour sa liaison au promoteur de *srn_3610_sprC.*

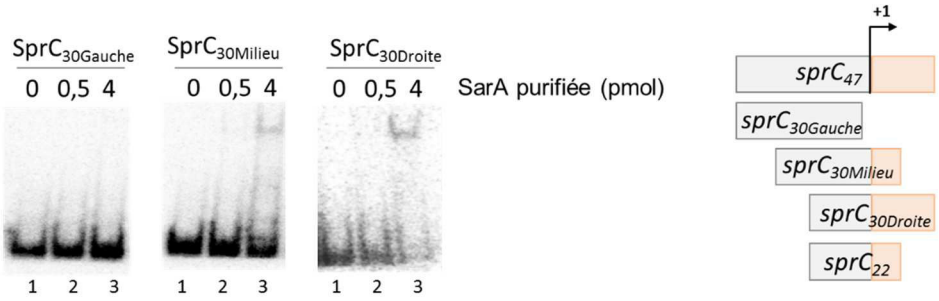


Figure 44 : Identification de 22 nucléotides fixés par SarA. 10 fmol de sondes de 30nt (SprC$_{30Gauche}$, SprC$_{30Milieu}$ et SprC$_{30Droite}$) sont mis en contact avec la protéine SarA purifiée (0,5-4 pmol).

Pour vérifier l'implication des 22 nt dans la fixation de SarA, nous avons mis en compétition la sonde radiomarquée *sprC*$_{P267}$*, avec la sonde *sprC*$_{P267}$ froide ou *sprC*$_{P267\Delta22}$ (sonde de 267pb délétée des 22 nucléotides identifiés précédemment). Nous constatons que l'ajout de sonde *sprC*$_{P267}$ froide (Figure 45 ; puits 3 à 5) provoque une diminution de la quantité de sonde radiomarquée *sprC*$_{P267}$ retardée. A l'inverse, la sonde *sprC*$_{P267\Delta22}$ n'interfère pas dans la fixation de SarA sur *sprC*$_{P267}$* (puits 6 à 8). Cette expérience montre que ces 22pb sont

indispensables à la fixation *in vitro* de SarA sur la région promotrice et génique de *srn_3610_sprC* (Figure 45).

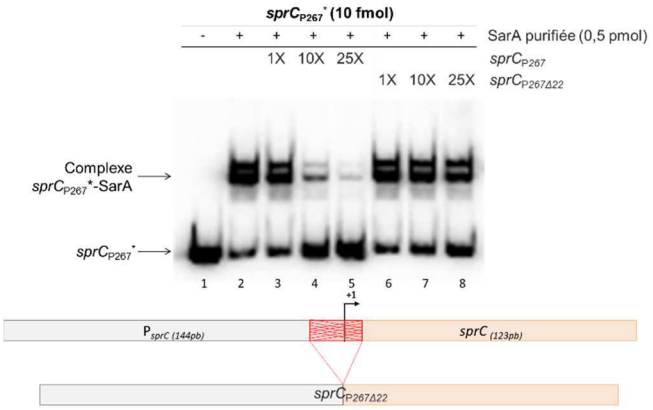


Figure 45 : Les 22 nt sont nécessaires à la fixation de SarA. 0,5 pmol de SarA est fixé à 10 fmol de *sprC*$_{P267}$* (puits 2 à 8). La sonde *sprC*$_{P267}$ non radiomarquée (de 1 à 25X : puits 3 à 5) ou la sonde *sprC*$_{P267Δ22}$ (de 1X à 25X : puits 6 à 8) est ajoutée à la réaction de fixation *sprC*$_{P267}$*-SarA.

En conclusion, si les expériences d'analyse transcriptionnelle (Northern Blot, gène rapporteur) ont montré que SarA est un répresseur fort de la transcription de *srn_3610_sprC*, les expériences d'empreinte à la DNase I et de retard sur gel montrent que SarA agit de manière directe par fixation sur une région du promoteur et du gène de *srn_3610_sprC*.

SarA étant un facteur de transcription majeur chez *S. aureus*, nous nous sommes posé la question de savoir si la séquence de protection de SarA pour *srn_3610_sprC* pouvait se retrouver ailleurs dans le génome notamment au niveau du promoteur d'autres gènes.

III. Identification d'un ARN régulateur : Srn_9340

A. Recherche bio-informatique

A l'aide du logiciel BLASTN (via la plateforme NCBI), une recherche bio-informatique a été menée à partir de la séquence de 47 pb (région de *srn_3610_sprC* protégée par SarA) et le génome de *S. aureus* NCTC8325. La séquence étant inférieure à 50 nucléotides, nous avons décidé d'optimiser notre recherche avec le programme Blastn, recommandé pour les séquences nucléotidiques courtes. Le programme nous a permis d'identifier 91 résultats dans la souche NCTC8325 avec des E-value allant de 10^{-18} à 3,1. L'E-value correspond à la probabilité que les deux séquences se ressemblent par hasard (plus l'E-value est faible, plus la probabilité est grande que la ressemblance entre les deux séquences ne soit pas due au hasard). Arbitrairement, nous avons décidé d'écarter les alignements avec une E-value inférieure à 0,01. L'utilisation de ce seuil d'exclusion a permis de réduire le nombre de candidats, ainsi sur les 91 candidats initialement trouvés seulement 6 entraient dans cette catégorie. Parmi les résultats écartés, de nombreuses séquences d'une quinzaine de nucléotides correspondaient à des boîtes -10 ou -35 de promoteurs d'autres gènes.

Candidat	Coordonnées (Brin)	Taille (nt)	Identité	E-Value	Séquence	Localisation
n° 1	1861616 - 1861570 (-)	47	47/47 (100%)	10^{-18}	GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG ||||||||||||||||||||||||||||||||||||||||| GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG	Promoteur de *srn_3610_sprC*
n° 2	1863936 - 1863908 (-)	29	27/29 (93%)	$3x10^{-6}$	AAAAATAAAATATTTATATTATAATATAA ||||||||||| | ||||||||||||| AAAAATAAAATAATAATATTATAATATAA	Région intergénique
n° 3	1861584 - 1861605 (+)	22	22/22 (100%)	$4x10^{-5}$	TATTTATATTATAATATAAATA |||||||||||||||||||||| TATTTATATTATAATATAAATA	Promoteur de *srn_3610_sprC*
n° 4	1692309 - 1692281 (-)	29	25/29 (86%)	0,002	AATATTTATATTATAATATAAATAAGTCA || ||| |||||| | ||||||||||||| AAGATTAATATTACACTATAAATAAGTCA	Terminateur de *coaE*
n° 5	749723 - 749745 (+)	23	21/23 (91%)	0,006	TAAAATATTTATATTATAATATA ||||||||||||| |||||| || TAAAATATTTATAATATAATTTA	Région intergénique
n° 6	889131 - 889102 (-)	30	26/30 (87%)	0,006	AAAATAAAATATTTATATTATA-ATATAAA ||||||||||||| ||||| | ||||||| AAAATAAAATATTATTATTAAATATATAAA	Promoteur de *saouhsc_00917*

Tableau 2 : Résultats de la recherche de *sprC*$_{47}$ dans le génome de *S. aureus* NCTC8325. Les six résultats avec les E-value < 0,01 sont répertoriés avec leurs coordonnées génomiques, leur pourcentage d'identité avec *sprC*$_{47}$ et leur localisation dans le génome de NCTC8325.

Logiquement, le premier résultat obtenu par le logiciel BLASTN est une séquence de 47pb avec 100% d'identité. Ces coordonnées génomiques permettent de localiser cette séquence sur le gène et le promoteur de *srn_3610_sprC*. Un autre résultat pour *srn_3610_sprC* est également trouvé (candidat n°3). Cette séquence de 22 nt (avec 100% d'identité), positionnée sur le brin +, correspond à une particularité de la séquence utilisée pour cette étude. En effet, parmi ces 47 nt, se trouve une région palindromique, c'est-à-dire, qu'elle peut se lire à la fois sur le brin + et -. C'est pourquoi les 22 nt qui composent le palindrome sont retrouvés une seconde fois.

Pour les autres candidats, les séquences candidates contenaient de 23 à 30 nt et ces séquences ciblaient toutes la région promotrice de la séquence de *sprC*$_{47}$. Pour les différencier, la localisation génomique a été le deuxième critère de classement. Les séquences cibles identifiées sont des séquences putatives de fixation du facteur de transcription SarA. Il nous est apparu important de noter que ces séquences se situent dans une région promotrice d'un gène identifié, ou comme pour *srn_3610_sprC*, dans la région promotrice d'un petit ARN. Ainsi, nous avons pu exclure de l'analyse le candidat n°4 (séquence de 29 nt avec 86% d'identité) qui est localisé au niveau du terminateur du gène *coaE*, codant pour une kinase de la coenzyme A.

Les trois derniers candidats ont été localisés sur le génome. La séquence de 30 nt du candidat n°6 (87% d'identité) est positionnée au niveau du promoteur de *saouhsc_00917* (E-value = 0,002). Les candidats n°2 et n°5 sont, quant à eux, dans des régions intergéniques. Les ARN régulateurs n'étant pas tous annotés, il est possible que ces candidats se situent en amont de gènes exprimant des ARNrég. Nous avons donc visualisé, grâce au logiciel Artémis, les séquences candidates (n°2 et n°5) parmi les données du séquençage à haut débit des ARN de la souche HG003 (Figure 46). Ainsi, le candidat n°5 (séquence de 23 nucléotides avec

91% d'identité) semble se positionner en amont du transcrit *secA* codant pour la sous-unité A de la préprotéine translocase (E-value = 0,006).

De manière intéressante, le candidat n°2 est retrouvé en amont d'un transcrit annoté pour un petit ARN : Srn_9340. Cette séquence de 29 nucléotides possède 93% d'identité avec la séquence *sprC*$_{47}$ (E-value = 3x10^{-6}).

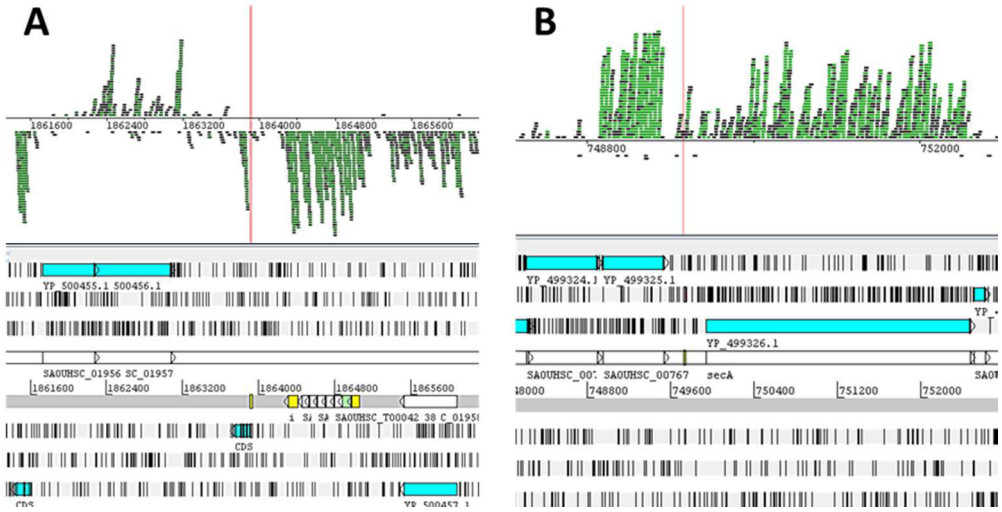


Figure 46 : Localisation du résultat n° 2 (A) et n°5 (B) du BLASTN. La visualisation est effectuée avec le logiciel Artémis sur les données de séquençage à haut débit des ARN de la souche HG003 (P. Bouloc).

La recherche bio-informatique de séquences similaires à *sprC*$_{47}$ a permis d'identifier trois autres régions promotrices potentiellement fixées par SarA. Toutes partagent une homologie de séquence avec la partie promotrice de *sprC*$_{47}$. Néanmoins, pour la suite de l'étude, nous nous sommes focalisés sur le gène *srn_9340* exprimant un ARN régulateur. En effet, au vu de la ressemblance entre la zone protégée de la dégradation de la DNase I par SarA pour *srn_3610_sprC* et la séquence située en amont de *srn_9340,* nous pouvions penser que SarA était également acteur de la régulation d'un second ARNrég, Srn_9340.

B. Caractérisation de Srn_9340

La mise en place par le laboratoire de la base de données SRD a permis de nommer l'ARN Srn_9340 trouvé ci-dessus (Sassi et al., 2015). Cet ARN avait préalablement été identifié par Howden *et al.* dans la souche JKD6008, comme l'ARN sRNA287 parmi 408 autres ARNrég (Howden et al., 2013). En 2016, Mäder et *al.* ont retrouvé Srn_9340 dans la souche HG001 et l'ont renommé S774 (Mader et al., 2016). Pour des questions d'uniformité des nomenclatures des ARN régulateurs, nous avons décidé de garder le nom de Srn_9340.

L'étude de *srn_9340* est effectuée dans la souche HG003 qui est la souche de référence pour l'étude de la régulation transcriptionnelle et que nous avions utilisée pour

l'étude pour *srn_3610_sprC*. Nous avons voulu d'abord confirmer l'expression de *srn_9340* dans la souche HG003.

Avec l'utilisation d'une sonde spécifique pour l'ARN Srn_9340, nous avons montré, par Northern Blot, que le gène *srn_9340* produit, dans la souche HG003, deux transcrits d'une taille d'environ 110 bases (Figure 47). La taille a été déterminée grâce à la migration conjointe d'un marqueur de taille préalablement marqué à la digoxigénine. Nous constatons, après 5 heures de croissance, que la quantité du grand transcrit est plus important que celle de l'ARN court.

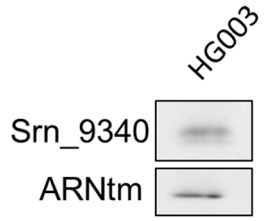


Figure 47 : *srn_9340* produit deux transcrits. 10µg d'ARN totaux sont déposés sur gel de polyacrylamide (8%) et les ARN Srn_9340 et ARNtm (contrôle de dépôt) sont révélés avec des sondes adaptées.

Les extrémités 5' et 3' des deux transcrits de Srn_9340 ont été déterminées par la technique de RACE (Rapid amplification cDNA ends) (Figure 48). Nous avons montré que les transcrits partagent la même extrémité 5' (coordonnée : 1863905) mais diffèrent à leur extrémité 3' (coordonnées : 1863797 et 1863790). Ainsi, le transcrit court possède 108 nucléotides et le transcrit long 116 nucléotides.

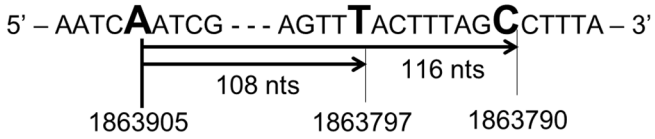


Figure 48 : Détermination des extrémités 5' et 3' des deux transcrits de *srn_9340* par RACE.

Le gène *srn_9340* est localisé à la fin de l'îlot génomique vSaß entre le gène *saouhsc_01957* codant pour une protéine hypothétique (appartenant à l'îlot) et le regroupement de gènes d'ARN de transfert appartenant au core génome. La présence du gène *srn_9340* est dépendante de la présence de l'îlot de pathogénie mais également du type de l'îlot vSaß qui a été intégré. Ainsi, les souches N315, Mu3 ou JH1, par exemple, possèdent l'îlot vSaß de type I différent de la souche HG003 à son extrémité avec la présence de gènes codant pour de nombreuses toxines mais ne possèdent pas *srn_9340* (Baba et al., 2008).

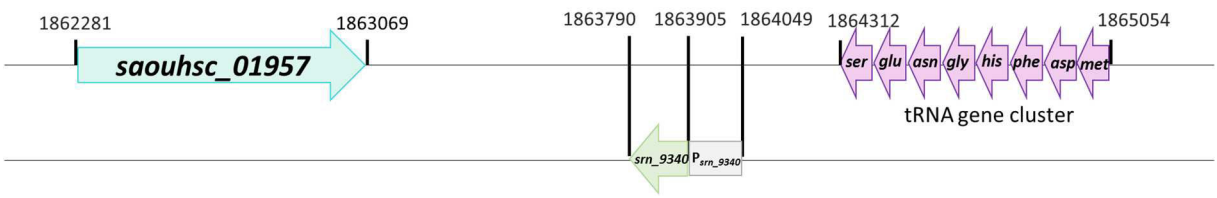


Figure 49 : Localisation génomique de la souche NCTC8325 de *srn_9340*.

De manière intéressante, *srn_9340* est génomiquement proche de *srn_3610_sprC* (2338pb les séparent), néanmoins la présence de l'un n'est pas gage de la présence du second. Dans le Tableau 3 (page 80) sont listées les souches en fonction de la présence de *srn_3610_sprC* et/ou *srn_9340*. Alors que la plupart des souches contiennent les deux petits ARN (52/96 étudiées), seules quelques-unes ne possèdent que l'un d'entre eux. La présence de *srn_3610_sprC* seul (27/96 souches) suppose que l'intégration de l'îlot de pathogénie a été effectuée sans son extrémité dans laquelle se situe *srn_9340*, à moins que, comme dans le cas de N315, il ne s'agisse de l'intégration d'un type particulier de vSaß (Baba et al., 2008). La présence de *srn_9340* seul (6/96) reflète la présence passée de l'îlot de pathogénie qui aurait été partiellement éliminé lors de l'évolution. Enfin, dans les 11 souches où les deux gènes sont absents, il semblent que ce soit tout l'îlot génomique qui soit absent.

C. Séquence de *srn_9340* et de son promoteur

Comme pour l'analyse de *srn_3610_sprC*, les séquences promotrices et géniques de *srn_9340* ont été étudiées dans les 58 génomes où *srn_9340* est présent (Figure 50).

srn_9340(promoteur + gène)
1864049-*ATCAAATGTGAATATGTTTTAAGCGTTGTATGTACATCTGCATGACCTAATCGTTTTGAAATGTA*
*ATGTTTAGCTGTTACTTAAAGAAAATCAAATATTATATATACGCT**TGAAA**AAATAAAATAATAATAT**TATAAT***
*ATAATC***A**ATCGGCATCTACGTGCGGATTATAAAGTAAAGCGTAGCATAACCAGGAGTTAGTCGATTTTTAATA
AATTCGACTAAATTATTAATAAAGGCTAAAGTAATTAGTT**T**ACTTTAGC-1863790

Figure 50 : Séquences nucléotidique du promoteur (P*srn_9340*) et du gène *srn_9340*. Le promoteur est indiqué en italique. Les boîtes -10 et -35 ont été respectivement identifiées en rouge et vert. Le nucléotide +1 de transcription est marqué en gras et souligné. Le dernier nucléotide du transcrit court est distingué en violet.

L'organisation chromosomique de *srn*_9340 étant différente de *srn_3610_sprC,* le promoteur de *srn_9340* a été défini arbitrairement comme étant composé de 144 nucléotides. Parmi cette séquence, une quinzaine de points de mutations ont été recensés et se répartissent sur l'ensemble de la séquence. Les boîtes -10 (rouge) et -35 (vert) ont été localisées et sont identiques à celles de *srn_3610_sprC*. Nous pouvons donc supposer que c'est également la sous-unité σ^A qui est impliquée dans la transcription de *srn_9340.* Cette hypothèse rejoint les observations de Mäder *et al.*(Mader et al., 2016). Le gène *srn_9340* est composé de 116 nucléotides et sa séquence est très conservée dans l'évolution de *S. aureus,* avec seulement cinq points de mutations constatés dont deux qui pourraient se compenser (Figure 51).



Figure 51 : Séquence de *srn_9340* de la souche *S. aureus* NCTC 8325. Les mutations possibles sont notées par un astérisque (*).

***srn_3610_SprC* et *srn_9340* présents (52 souches)**	***srn_3610_SprC* présent uniquement (27 souches)**	***srn_9340* présent uniquement (6 souches)**	***srn_3610_SprC* et *srn_9340* absents (11 souches)**
HG003	ST228 isolates	SA268	MRSA 252
NCTC 8325	ECT-R2	SA40	H-EMRSA-15
Newman	ED98	SA957	HO 5096 0412
COL	Mu3	JKD6008	AUS0325
USA300-FPR3757	JH1	M013	08-02300
USA300-TCH1516	JH9	MS4	XQ
USA300-2014.CO1	Mu50		FDA209P
USA300-2014.CO2	N315		RKI4
TW20	ST772-MRSAV		TMUS2134
T0131	6850		TMUS2126
1189-97	TMUS2134		93b S9
VC40	TMUS2126		
Z172	CN1		
Gv69	JKD6159		
Gv88	04-02981		
Gv51	2148.N		
DSM-20231	NCCP14562		
GR2	NCCP14558		
HC1340	ZJ5499		
HC1335	RIVM6519		
Be62	ST20130942		
5118.N	ST20130943		
1969.N	FDAARGOS 159		
1625.CO1	NZAK3		
1971.CO1	MI		
Tager104	SA564		
71193	HOU1444-VR		
ST398			
MSSA476			
MW2			
11819-97			
Bmb9393			
XN108			
JS395			
UA-S391			
08-02119			
V605			
V521			
ST20130938			
ST20130939			
ST20130940			
ST20130941			
UTSW MRSA55			
HUV05			
CA12			
V2200			
USA300 2014.1			
USA300 2014.2			
CA15			
M121			
DSM 20231			
NCTC13435			

Tableau 3 : Liste de souches selon la présence ou l'absence de *srn_3610_sprC* ou *srn_9340*.

Aucun cadre ouvert de lecture n'a été identifié par le logiciel ORFFinder, nous supposons alors que ce gène ne code pas de peptide.

D. Prédictions de structures des transcrits de Srn_9340

Prédite par le logiciel Mfold, la structure du transcrit long (116 nt) serait composée de trois « tiges-boucles » dont la dernière porterait le signal de terminaison de la transcription (Figure 52). L'ARN est assez structuré avec 57% des nucléotides impliqués dans des liaisons ARN double brin.

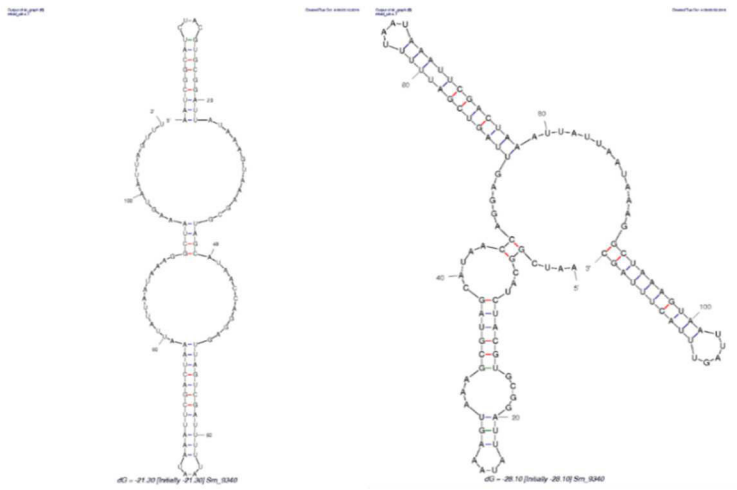


Figure 52 : Structure des deux transcrit de *srn_9340* prédites par Mfold.
(ΔG estimé Srn_9340_{court} = -21,30kcal/mol ; Srn_9340_{long} = -28,10kcal/mol).

L'absence des huit bases à l'extrémité 3' de l'ARN semble entrainer une modification importante de la structure du transcrit avec notamment une perte du signal de terminaison (Figure 52). Dans ce cas, nous pouvons supposer que la terminaison de la transcription du transcrit court est dépendante du facteur Rho à moins qu'il ne soit le résultat d'une modification post-transcriptionnelle du transcrit long (action de désoxyribonucléases).

IV. Recherche de régulateurs de la transcription de *srn_9340*

Nous avons effectué, pour *srn_9340,* les mêmes études que celles réalisées pour l'étude de la régulation pour *srn_3610_sprC*.

A. Expression de *srn_9340* dans différents contextes mutants

L'expression de *srn_9340* a été étudiée par Northern Blot dans différentes souches HG003 mutées pour un des facteurs de transcription de la famille SarA ainsi que dans la souche 8325-4 et son contexte mutant Δ*rot* (Figure 53).

L'ARN Srn_9340 a été quantifié dans les souches et normalisé par rapport à la quantité d'ARNtm (ARN servant de contrôle de dépôt). Les valeurs obtenues pour les souches mutantes sont comparées à celles de la souche sauvage HG003 (puits 1). Ainsi, la délétion de *sarS*, *sarT*, *sarV* ou *sarX* (Figure 53 ; puits 4 à 7) n'entraîne pas de modification de la quantité d'ARN Srn_9340 par rapport à HG003. Le Northern Blot ne présente pas de différence notable entre la souche sauvage 8325-4 et la souche mutante 8325-4Δ*rot* (respectivement puits 8 et 9). Ces facteurs de transcription ne semblent pas impacter l'expression de *srn_9340.*

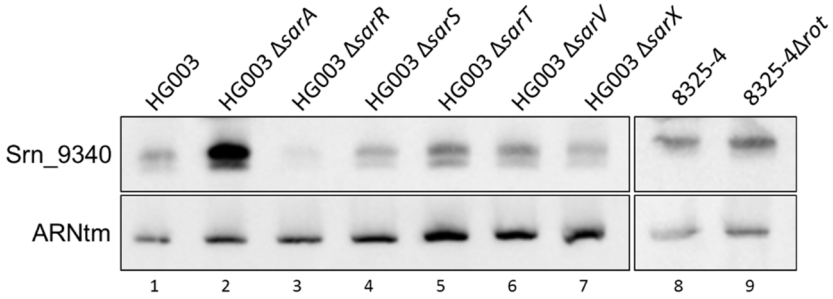


Figure 53 : Expression de *srn_9340* dans différents contextes mutants pour l'un des facteurs de transcription de la famille SarA. Les ARN totaux ont été extraits à 5 heures de croissance et les ARN Srn_9340 et ARNtm ont été révélés par des sondes spécifiques.

La comparaison de la quantité de Srn_9340 entre HG003 et HG003Δ*sarR* (respectivement puits 1 et 3 ; Figure 53) permet de constater une diminution de la quantité de transcrits dans la souche mutante. SarR aurait donc un effet positif sur la transcription de *srn_9340*. Dans le mutant HG003Δ*sarA* (puits 2), une augmentation d'un facteur 3,5 de la quantité de Srn_9340 est observée par rapport à HG003 (puits 1). SarA a donc un effet négatif sur la transcription de *srn_9340* et serait alors un répresseur de l'expression de *srn_9340.*

Au vu des similitudes de séquences entre le promoteur de *srn_3610_sprC* et *srn_9340*, il paraissait logique que SarA soit un répresseur fort de la transcription de *srn_9340* et l'étude de l'expression de *srn_9340* dans la souche mutante HG003Δ*sarA* a confirmé cette idée. En ce qui concerne l'effet observé dans la souche HG003Δ*sarR*, nous savons que SarR est un répresseur de SarA. L'effet vu sur la transcription de *srn_9340* pourrait, comme nous l'avions déjà observé pour *srn_3610_sprC*, correspondre à une action indirecte de SarR nécessitant SarA.

L'étude de l'expression de *srn_9340* dans des contextes mutants pour certains facteurs de transcription de la famille SarA a permis d'identifier SarA et SarR comme des régulateurs de la transcription de *srn_9340.* Grâce à la souche HG003Δ*arnIII* dont nous disposons au laboratoire, nous avons voulu savoir si l'ARN régulateur ARNIII est un acteur de la régulation de *srn_9340.*

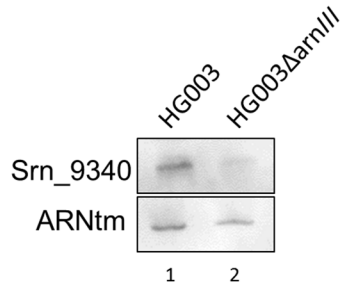


Figure 54 : Expression de *srn_9340* en absence d'expression de l'ARNIII. Les ARN Srn_9340 et ARNtm (contrôle de dépôt) ont été révélé avec des sondes spécifiques.

A la Figure 54, la comparaison de la quantité d'ARN Srn_9340 dans la souche HG003 (puits 1) et HG003Δ*arnIII* (puits 2) montre, après normalisation avec la quantité d'ARNtm, une diminution d'un facteur 2. Ainsi, l'ARNIII est un régulateur positif de l'expression de *srn_9340.* L'ARNIII n'est pas caractérisé comme pouvant agir directement sur les promoteurs comme un facteur de transcription peut le faire. Donc, nous pouvons supposer que l'ARNIII agit soit indirectement sur l'expression de *srn_9340* via la régulation d'un facteur de transcription soit directement sur l'ARN Srn_9340 en stabilisant celui-ci. En effet, l'ARNIII est connu pour s'apparier aux ARN et améliorer ou diminuer la stabilité des ARN cibles (Huntzinger et al., 2005).

Comme observé pour *srn_3610_sprC*, la transcription de *srn_9340* est faible dans des conditions conventionnelles de croissance. De plus, SarA a un effet répresseur fort sur l'expression de *srn_9340* pouvant cacher l'action d'un activateur. La délétion de *sarA* en plus d'un second facteur de transcription permettrait d'identifier ce régulateur (Figure 55). Pour cela, nous avons regardé l'expression de *srn_9340* dans les souches doublement mutées déjà utilisées pour l'étude de la régulation de *srn_3610_sprC*.

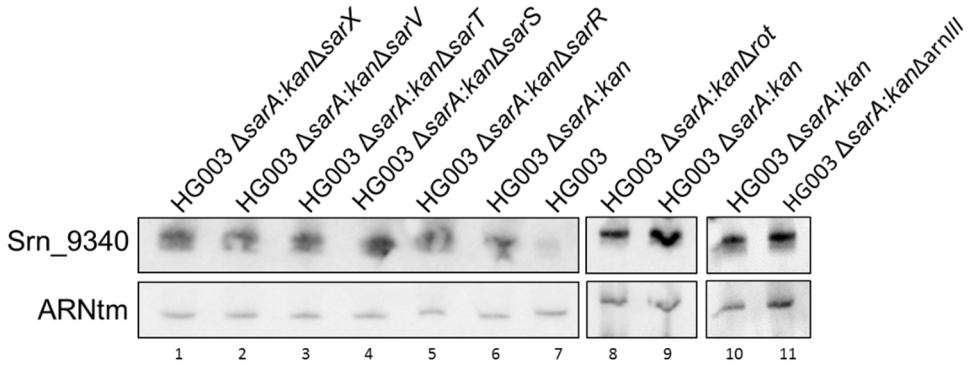


Figure 55 : Expression de *srn_9340* dans des souches ne produisant ni SarA ni l'un des autres facteurs de transcription étudiés. Les ARN ont été extraits à 5 heures de croissance. L'ARNtm sert de contrôle de dépôt.

Nous retrouvons dans la souche HG003Δ*sarA* (puits 6) l'augmentation de la quantité d'ARN Srn_9340 d'un facteur 3,5 par rapport à la souche sauvage HG003 (puits 7). Dans les autres souches doublement mutées pour *sarA et sarR, sarS, sarT, sarV* ou *sarX* (puits 1 à 5),

la quantité de Srn_9340 est maintenue au niveau de la quantité observée dans la souche HG003Δ*sarA*. Nous n'observons aucune différence également entre les souches HG003Δ*sarA* (puits 9) et HG003Δ*sarA*Δ*rot* (puits 8). Les puits 10 (HG003Δ*sarA*) et 11 (HG003Δ*sarA*Δ*arnIII*) de la Figure 55 ne présentent pas de différence notable dans la quantité d'ARN Srn_9340.

Ainsi, le fait de ne pas observer de différence entre la souche HG003Δ*sarA* et les souches doublement mutées semblent signifier que SarR, SarS, SarT, SarV, SarX et Rot n'agissent pas sur l'expression de *srn_9340.* En ce qui concerne l'ARNIII, nous avons vu que la délétion du gène transcrit en ARNIII entraîne une diminution de la quantité de Srn_9340. Par contre, en délétant en plus de l'ARNIII le gène *sarA*, nous n'observons pas de différences d'expression par rapport à la souche HG003Δ*sarA*. Nous pouvons supposer que la modification de l'expression de *srn_9340* passerait par une régulation indirecte par l'ARNIII.

Nous avons donc identifié l'implication de trois régulateurs dans l'expression de *srn_3610_sprC* et *srn_9340* : SarA, SarR et l'ARNIII. Pour l'ARNIII, nous constatons une différence dans l'observation de son rôle de régulateur. Pour *srn_9340*, nous avons vu que l'effet est visible uniquement dans la souche mutante HG003Δ*arnIII* alors que pour *srn_3610_sprC,* nous observons l'effet de l'ARNIII dans la souche doublement mutée HG003Δ*sarA*Δ*arnIII*. L'explication pourrait dans un premier temps d'ordre technique. En effet, la révélation d'un Northern Blot n'est pas assez résolutive pour de petites variations de quantité d'ARN. Il existe peut-être alors des variations d'expression de *srn_3610_sprC* dans la souche HG003Δ*arnIII*, non visible en Northern Blot. Pour *srn_9340,* la régulation, sans doute indirecte, de l'ARNIII dans la souche HG003Δ*sarA*Δ*arnIII* est peut-être masquée par la forte augmentation de la quantité d'ARN due à la délétion de *sarA.*

La forte répression de *srn_9340* par SarA dans la souche HG003 est caractérisée, il reste à savoir si SarA agit, comme pour *srn_3610_sprC*, par fixation sur le promoteur de *srn_9340*.

B. Action de SarA sur la transcription de *srn_9340*

Pour s'assurer que l'absence de sarA dans la souche HG003Δ*sarA* est responsable du phénotype de dérépression de *srn_9340*, l'expression de *srn_9340* a été étudiée dans la souche HG003Δ*sarA* complémentée. Pour cela, l'ARN a été détecté par Northern Blot parmi les ARN endogènes extraits dans les souches ayant servies dans l'expérience de complémentation pour *srn_3610_sprC* (Figure 35). Les souches avec un niveau sauvage de SarA sont situées aux puits 1 et 2, celles surexprimant *sarA* aux puits 3 et 4, celles délétée du gène *sarA* aux puits 5 et 6 et celles fonctionnellement complémentée aux puits 7 et 8 (Figure

56). Les niveaux d'expression de *sarA* dans les différentes souches avaient été préalablement validés (Figure 35).

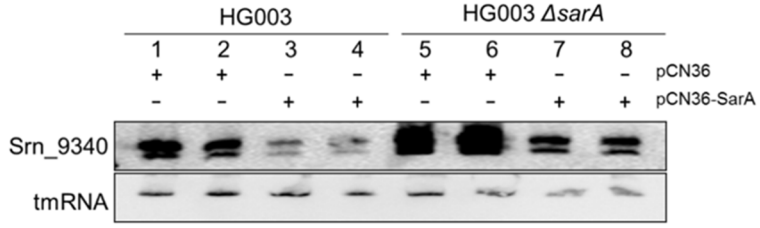


Figure 56 : Expression de *srn_9340* dans les différentes souches utilisées lors des expériences de complémentation fonctionnelle de SarA. Les ARN Srn_9340 et ARNtm sont révélés avec des sondes spécifiques sur un gel de polyacrylamide 8%.

Le niveau basal de Srn_9340 dans la souche exprimant uniquement la copie chromosomique de *sarA* est révélé dans les puits 1 et 2. Lorsque nous surexprimons *sarA*, nous constatons une diminution de deux fois la quantité de Srn_9340 (puits 3 et 4) par rapport à la quantité de Srn_9340 dans les souches sauvages. La quantité de Srn_9340, dans les souches où le gène *sarA* est absent (puits 5 et 6), est augmentée environ d'un facteur 3. Dans les puits correspondant à la complémentation fonctionnelle de la souche HG003Δ*sarA* par le plasmide pCN36-SarA (puits 7 et 8), la quantité d'ARN Srn_9340 revient à un niveau proche de celui de la souche sauvage. L'augmentation de l'expression de *srn_9340* dans la souche HG003Δ*sarA* est corrigée par la production de SarA (via le plasmide pCN36-SarA), SarA est donc bien acteur dans la régulation de *srn_9340.*

L'action de SarA sur l'expression du gène *srn_9340* a été confirmée, nous cherchons maintenant à savoir si le promoteur de ce dernier est le lieu de la régulation de l'expression de *srn_9340* par SarA. Pour cela, des expériences de gène rapporteur (ß-lactamase) ont été effectuées.

L'activité du promoteur de *srn_9340* est étudiée dans un contexte sauvage HG003 et mutant pour *sarA* (HG003Δ*sarA*). Pour cela, nous avons amplifié le promoteur de *srn_9340* de la souche HG003 (144 pb) qui a été placé en amont du gène *blaZ* (codant la ß-lactamase) présent sur le plasmide pCN41c (Figure 57).

Pour les expériences de gène rapporteur, les protéines ont été extraites au bout de 5 heures de croissance dans les différentes souches. L'activité de la ß-lactamase mesurée est le témoin de l'activité du promoteur. Dans la souche HG003, le niveau de l'activité de la ß-lactamase est de 47,84 UA/mg ±18,99 et dans la souche HG003Δ*sarA*, elle est de 575,59 UA/mg ±73,59. Ainsi, P_{srn_9340} est 12 fois plus actif dans la souche HG003Δ*sarA* par rapport à la souche HG003. L'absence de SarA entraîne une activité significativement augmentée de P_{srn_9340}, le promoteur de *srn_9340* est donc le siège de la régulation par SarA.

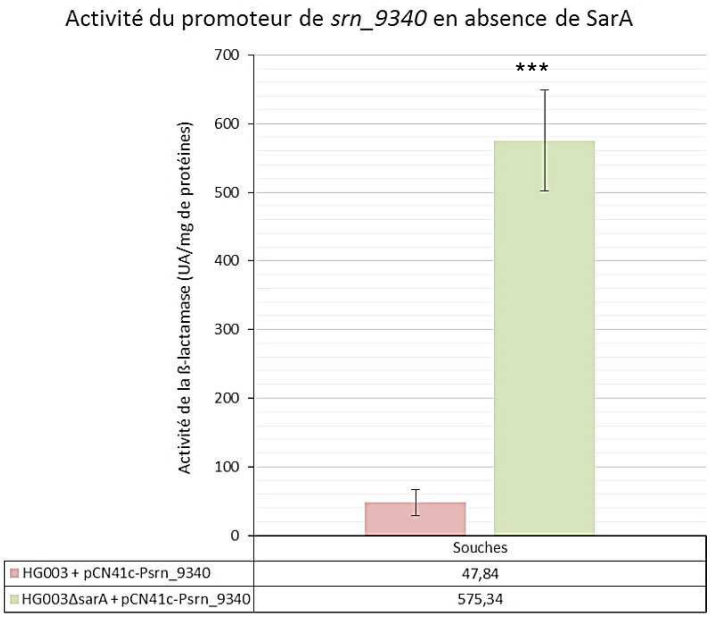


Figure 57 : Activité du promoteur P*srn_9340* en absence de SarA. La moyenne des activités de la ß-lactamase de trois expériences indépendantes a été représentée (Test de Mann-Whitney ; *** *P*<0,001)

Il est intéressant de noter ici que pour *srn_9340*, nous avons un plus grand différentiel d'expression du gène entre la souche HG003 et HG003Δ*sarA* avec la technique de gène rapporteur (différence d'un facteur 12) par rapport au Northern Blot (différence d'un facteur 3,5). Les Northern Blot révèlent deux phénomènes biologiques : l'activité du promoteur et la stabilité des ARN transcrits. Les expériences de gène rapporteur, quant à elles, ne montrent que l'activité du promoteur.

Des expériences de mesure de demi-vie des ARN de la souche Newman avaient été préalablement réalisées au laboratoire. Nous avons utilisé ces ARN pour déterminer les demi-vies des ARN Srn_3610_SprC et Srn_9340 dans la souche de *S. aureus* Newman. Srn_3610_SprC a une demi-vie de 20 minutes et Srn_9340, une demi-vie de 5 minutes (Figure 58).

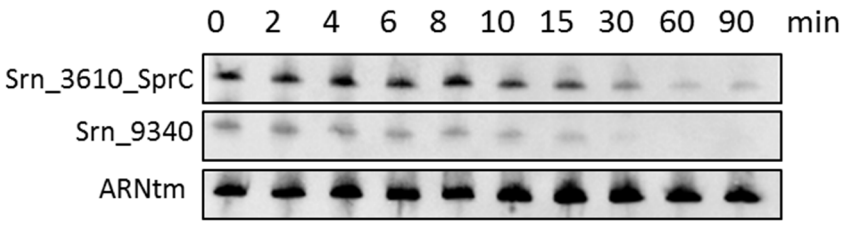


Figure 58 : Détermination de la demi-vie des ARN Srn_3610_SprC et Srn_9340 dans la souche *S. aureus* Newman. Le mécanisme de transcription a été arrêté par l'ajout de rifampycine. Les ARN totaux ont été extraits à différents temps après traitement à la rifampycine. Les ARN d'intérêt ont été spécifiquement révélés.

Nous constatons que dans la souche Newman, Srn_3610_SprC est plus stable que Srn_9340. Au vu de la grande similarité entre le génome de la souche Newman et celui de la souche HG003, la stabilité des ARN d'intérêt devrait être du même ordre de grandeur dans les deux souches. Nous supposons donc que la différence d'expression observée entre les Northern Bot et les expériences de gènes rapporteurs pour Srn_9340 pourrait être liée à la fois à la faible stabilité de Srn_9340 ainsi qu'au seuil de sensibilité faible du Northern Blot.

De manière générale, nous avons montré que l'action de SarA sur l'expression de *srn_9340* implique le promoteur de celui-ci. Nous aimerions savoir si, comme *srn_3610_sprC*, SarA se fixe sur le promoteur de *srn_9340*.

C. Fixation de SarA sur le promoteur de *srn_9340*

L'identification de *srn_9340* grâce à la séquence protégée de P_{sprC} par SarA et la vérification que SarA est bien acteur dans la régulation de ces ARN et laissent supposer que, comme pour *srn_3610_sprC*, SarA fixe le promoteur de *srn_9340.* Pour valider cette hypothèse, des expériences de retards sur gel ont été effectuées entre *srn_9340* et la protéine SarA purifiée.

Nous avons isolé un fragment d'ADN de 225pb (*srn_9340*$_{P225}$) correspondant à 165pb du promoteur et à 60 pb du gène de *srn_9340.* Ce fragment radiomarquée par du [γ-^{32}P] ATP est mis en présence de différentes quantités de SarA (10 à 1000 fmol) pendant 30 minutes à 37°C (lignes 2 à 10 ; Figure 59). Après migration sur gel de polyacrylamide natif, nous observons qu'en présence de 150 fmol de SarA (puits 5), une fraction de la sonde *srn_9340*$_{P225}$ est retardée dans sa migration. A 500 fmol de SarA (puits 7), la quasi-totalité de la sonde *srn_9340*$_{P225}$ est retardée. Au-delà de 500 fmol de protéine, et en ajoutant encore plus de protéine SarA, un deuxième retard de migration apparaît (« super-retard »). Lorsque nous ajoutons 1 pmol de SarA purifiée, la totalité de la sonde *srn_9340*$_{P225}$ est « super-retardée ».

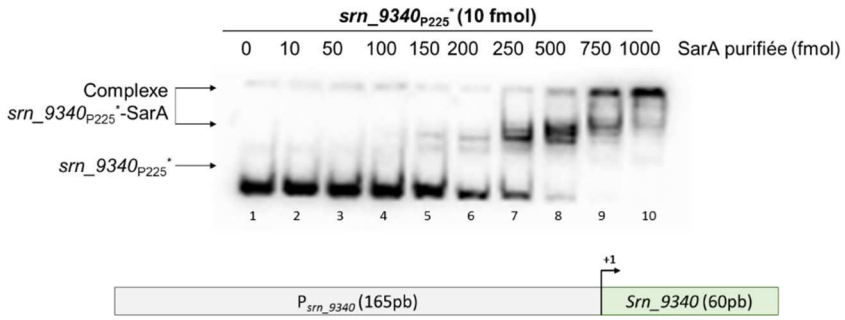


Figure 59 : SarA fixe la sonde *srn_9340*P225 *in vitro*. La sonde *srn_9340*$_{P225}$ (10 fmol) est mise en présence d'une quantité croissante de SarA (de 10 à 1000 fmol : puits 2 à 10).

Ces expériences de retards sur gel ont permis de déterminer la constante de dissociation (K_D) de la protéine SarA sur la sonde *srn_9340*$_{P225}$. Nous avons différencié le K_D du premier retard de celui du « super-retard ». Nous avons obtenu ainsi un K_{D1} de 10,9±0,2 nM pour la 1ère fixation de SarA sur *srn_9340*$_{P225}$ et le second K_{D2} de 38,9±3,5 nM pour le « super-retard ». Au regard du K_D de *sprC*$_{P267}$ (86±3 nM) et du K_{D1} de *srn_9340*$_{P225}$ (10,9±0,2 nM), nous constatons que SarA est 7,9 fois plus affin pour *srn_9340*$_{P225}$ par rapport à *sprC*$_{P267}$.

Contrairement à *srn_3610_sprC* où nous constations un seul retard de migration, deux retards de migration sont présents ici. Ce phénomène a déjà été observé et les auteurs avaient proposé que ce 2ème retard (à forte concentration en protéines) soit dû à une agglomération de SarA sur la sonde ADN (Andrey et al., 2010). Le fait que nous n'ayons vu le « super-retard » que pour *srn_9340*$_{P225}$ est probablement lié à la différence d'affinité de SarA pour *srn_9340*$_{P225}$ par rapport à *sprC*$_{P267}$. Si nous tenons compte des différences de K_D, les expériences de retard sur gel pour *sprC*$_{P267}$ n'auraient pas été faites avec des concentrations suffisantes de SarA pour visualiser un « super-retard » avec *sprC*$_{P267}$.

Afin de valider la spécificité de la fixation de SarA sur *srn_9340*$_{P225}$, nous avons mis la sonde *srn_9340*$_{P225}$ radiomarquée en présence d'1 pmol de SarA purifiée (Figure 60 ; puits 2) permettant de visualiser un « super-retard ». L'addition de la même quantité de compétiteur spécifique (1X de sonde *srn_9340*$_{P225}$ froide ; puits 3) que de sonde radiomarquée entraîne une diminution de la quantité de sonde « super-retardée » et l'apparition du premier retard de migration. A l'ajout de 10X de compétiteur spécifique (puits 4), le « super-retard » n'est plus présent au bénéfice du premier retard. A 50X et 100X (puits 5 et 6), la sonde radiomarquée est complètement libre de la fixation de SarA. La sonde *srn_9340*$_{P225}$ froide entre en compétition avec la sonde radiomarquée pour la fixation avec SarA. L'ajout de compétiteur non spécifique 16S$_{225}$ à 1X ou 10X (puits 7 et 8) ne déplace pas la sonde « super-retardée ». A 50X et 100X de 16S$_{225}$ (puits 9 et 10), la sonde radiomarquée est partagée entre les deux états retardés. Le compétiteur non spécifique n'entre en compétition que pour le « super-retard ». Cette observation confirmerait la fixation non spécifique de SarA pour des concentrations fortes. De plus, au regard du premier retard de migration, ces résultats montrent la spécificité de la fixation de la protéine SarA sur *srn_9340*$_{P225}$.

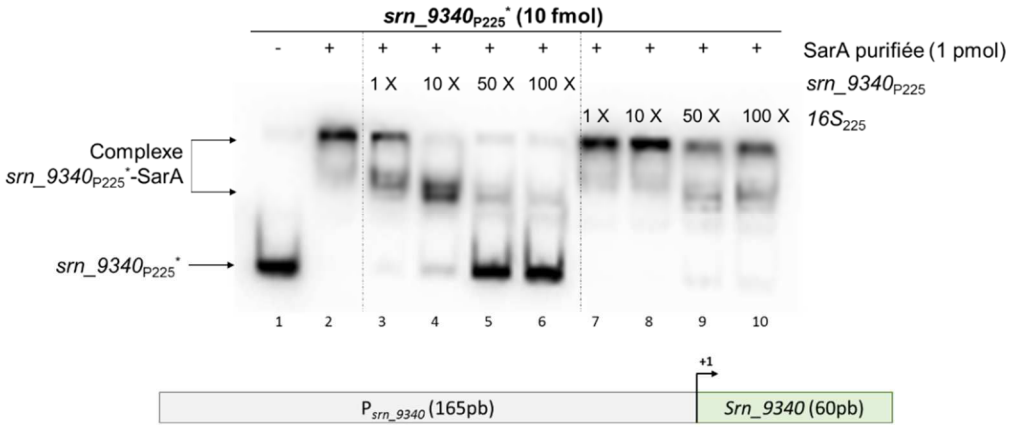


Figure 60 : SarA fixe de manière spécifique la sonde *srn_9340*$_{P225}$ *in vitro*. Des quantités croissantes de compétiteurs spécifique (*srn_9340*$_{P225}$ non marquée) et non spécifique (16S$_{225}$) sont ajoutés à la réaction de fixation (10 fmol de *srn_9340*$_{P225}$* / 4 pmol de SarA).

La fixation spécifique de SarA sur *srn_9340*$_{P225}$ est confirmée, il est maintenant possible de localiser le site de fixation de ce facteur de transcription sur notre sonde.

D. Détermination de la zone de fixation de SarA sur *srn_9340*

Pour caractériser la zone de fixation de SarA sur *srn_9340*$_{P225}$, des expériences d'empreinte à la DNase I ont été réalisées (Figure 61).

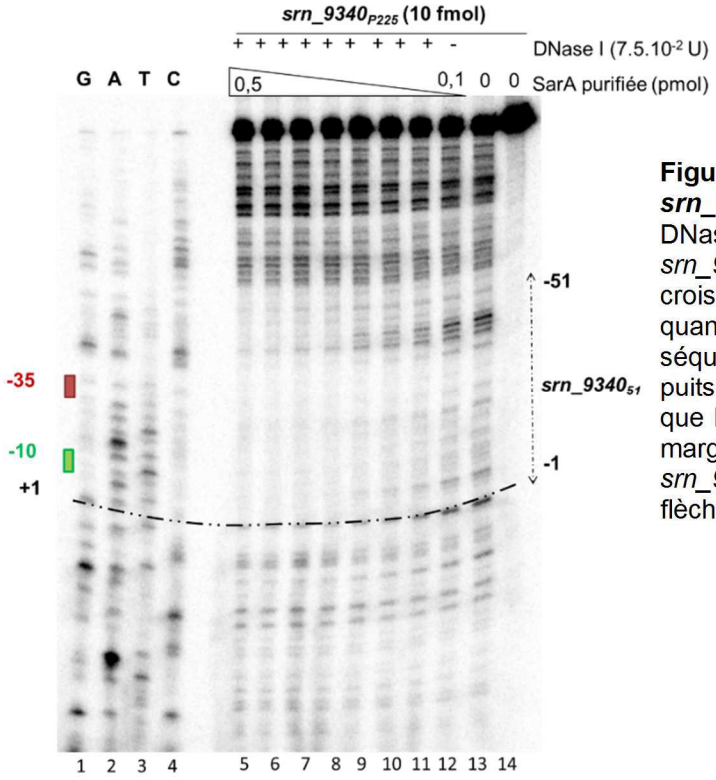


Figure 61 : SarA fixe le promoteur de *srn_9340*. Les expériences d'empreinte à la DNase ont été réalisées avec 10 fmol de *srn_9340*$_{P225}$ radiomarquée, une quantité croissante de SarA (0,1-0,5 pmol) et une quantité de DNase I (7,5.10^{-2} U). Le séquençage de la sonde est effectué dans les puits G, A, T et C. Les boîtes -10 et -35 ainsi que le +1 de transcription ont été placées en marge du séquençage. La zone protégée *srn_9340*$_{51}$ est représentée par une double-flèche.

Après avoir mis en contact *srn_9340*$_{P225}$ avec une quantité croissante de SarA (0,1-0,5 pmol de SarA purifiée), 7,5.10^{-2} unités de DNase I est ajoutée pour couper les fragments d'ADN une seule fois statistiquement. La présence de SarA (puits 6 à 13) entraîne une protection de la sonde *srn_9340*$_{P225}$ contre la dégradation par la DNase I entre les nucléotides -51 et -1 par rapport au +1 de transcription (localisé grâce au séquençage : lignes G, A, T et C). Cette séquence recouvre l'intégralité des boîtes -35 et -10, zones de fixation de la sous-unité σ^A de l'ARN polymérase. Cette zone de 51 pb se situe entièrement dans le promoteur de *srn_9340* et il nous reste à confirmer la fixation de SarA sur cette séquence et d'essayer de réduire la zone d'interaction entre le promoteur P_{srn_9340} et la protéine SarA.

E. Analyse de la séquence de *srn_9340* fixée par SarA

Afin de valider que SarA fixe effectivement la séquence de 51 nucléotides protégés dans l'expérience précédente, des nouvelles expériences de retards sur gel ont été effectuées. Pour cela, une sonde de 174pb (*srn_9340*$_{P225\Delta51}$) a été créée en délétant de la sonde *srn_9340*$_{P225}$ la séquence de 51 pb.

Dans les puits 1 à 5 de la Figure 62A, des expériences de retard sur gel avec la sonde complète *srn_9340*$_{P225}$ sont réalisées. Au puits 1, la bande obtenue correspond à la migration de *srn_9340*$_{P225}$ sans ajout de protéine. Lorsque nous ajoutons SarA (de 0,05 à 0,5 pmol ; puits 2 à 5), la sonde *srn_9340*$_{P225}$ est retardée à partir de 0,2 pmol de protéine SarA et la totalité de la sonde est retardée à 0,5 pmol de SarA. En ce qui concerne les puits 6 à 10 correspondant aux expériences de retard sur gel avec la sonde *srn_9340*$_{P225\Delta51}$, nous constatons que quelle que soit la quantité de protéine SarA ajoutée à la sonde *srn_9340*$_{P225\Delta51}$ radiomarquée (lignes 6 à 10), aucun retard de migration n'est observé. (Figure 62A). Nous concluons donc que les 51 nucléotides sont nécessaires à la fixation de SarA *in vitro*.

Une sonde de 51 pb (*srn_9340*$_{51}$) a été construite et mise en contact avec une quantité croissante de SarA (de 0,25 à 2 pmol ; puits 2 à 5 de la Figure 62B). La présence de SarA entraîne un retard de la migration de *srn_9340*$_{51}$ (à partir de 0,25 pmol de SarA). Plus nous ajoutons SarA, plus la proportion de sonde *srn_9340*$_{51}$ retardée dans sa migration est importante. Ces résultats montrent que ces 51 nucléotides sont suffisants pour la fixation de SarA *in vitro*.

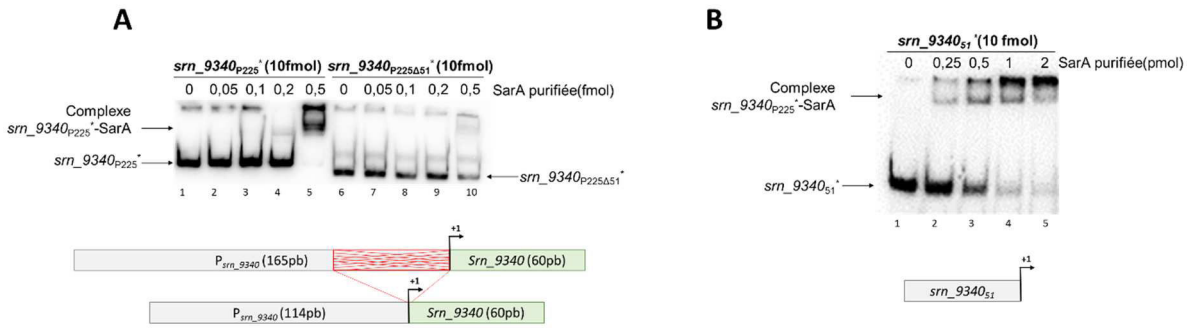


Figure 62 : Les 51 nucléotides sont nécessaires et suffisants à la fixation de SarA sur le promoteur de *srn_9340.* (A) 10 fmol de *srn_9340*$_{P225}$ ou *srn_9340*$_{P225\Delta51}$ (sonde délétée de *srn_9340*$_{51}$) sont mis en contact avec la protéine SarA (0,05-0,5 pmol). **(B)** Des quantités croissantes de la protéine SarA (0,25 à 2 pmol) sont ajoutées à 10 fmol de *srn_9340*$_{51}$.

Nous avons montré que les 51 pb identifiées sont nécessaires et suffisantes pour la fixation de SarA sur le promoteur de *srn_9340.* Pour déterminer plus précisément la zone de fixation de SarA sur *srn_9340*$_{51}$, trois sondes chevauchantes de 30pb ont été préparées pour recouvrir l'intégralité de ces 51 nucléotides (Figure 63).



Figure 63 : Identification de 23 nucléotides fixés par SarA. 10 fmol de sonde de 30nt (Srn_9340$_{30Gauche}$, Srn_9340$_{30Milieu}$ et Srn_9340$_{30Droite}$) sont mis en contact avec la protéine SarA purifiée (0,5-4 pmol).

Ces trois sondes ont été mises en contact avec 4 ou 20 pmol de SarA purifiée (puits 2 et 3 de chaque panel). Nous constatons qu'à l'ajout de 20 pmol, toutes les sondes sont retardées dans leur migration. Cette quantité est très forte et nous supposons que cette fixation est due à une agglomération de SarA et n'est donc pas spécifique. A 4 pmol de SarA, une bande retardée apparaît pour les sondes *srn_9340*$_{30Milieu}$ et *srn_9340*$_{30Droite}$, la quantité de sonde retardée observée est moins importante qu'avec 20 pmol de protéine et implique un complexe ADN-protéine de plus petite taille (migration moins retardée). Aucun retard de migration de *srn_9340*$_{30Gauche}$ n'est observé avec 4 pmol de SarA.

En conclusion, nous pouvons dire que la fixation de SarA se fait préférentiellement avec les sondes *srn_9340*$_{30Milieu}$ et *srn_9340*$_{30Droite}$. Ainsi, en superposant les séquences de ces deux sondes, 23 nucléotides communs ont été identifiés.

Pour tester l'hypothèse où les 23 nt identifiés sont nécessaires à la fixation de SarA sur *srn_9340*$_{P225}$, nous avons décidé de construire une sonde *srn_9340*$_{P225\Delta23}$ correspondant à la sonde *srn_9340*$_{P225}$ délétée de ces 23 pb. Cette sonde délétée a été utilisée dans des expériences de retard sur gel où la sonde complète *srn_9340*$_{P225}$ est mise en compétition avec *srn_9340*$_{P225\Delta23}$ (Figure 64).

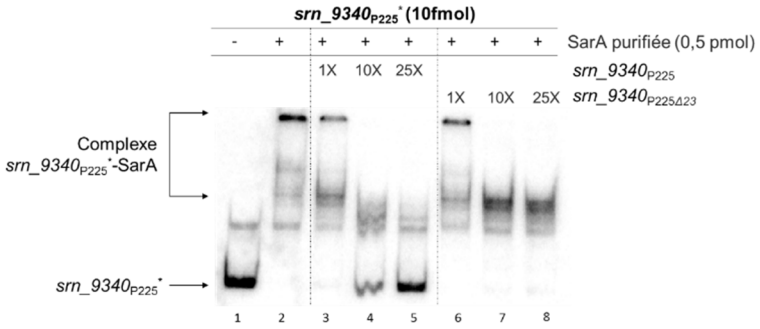


Figure 64 : Les 23 nt sont nécessaires à la fixation de SarA. 0,5 pmol de SarA est fixé à 10 fmol de *srn_9340*$_{P2225}$* (puits 2 à 8). La sonde *srn_9340*$_{P225}$ non radiomarquée (de 1 à 25X : puits 3 à 5) ou la sonde *srn_9340*$_{P267\Delta23}$ (de 1X à 25X : puits 6 à 8) est ajoutée à la réaction de fixation *srn_9340*$_{P225}$*-SarA.

Dans cette expérience de retard sur gel, impliquant *srn_9340*$_{P225}$* et la protéine SarA (Figure 64 ; puits 2), nous ajoutons la même quantité de sonde froide *srn_9340*$_{P225}$ (puits 3), et nous n'observons aucun changement. Cependant, lorsque le compétiteur froid est 10X ou 25X plus présent que la sonde radiomarquée (puits 4 et 5), nous observons un déplacement de la fixation de SarA depuis la sonde radiomarquée vers la sonde froide, la sonde *srn_9340*$_{P225}$* est libérée. Dans le cas où nous remplaçons la sonde *srn_9340*$_{P225}$ froide par la sonde *srn_9340*$_{P225\Delta23}$ froide (puits 6 à 8) dans cette expérience de compétition, l'ajout d'une même quantité de sonde *srn_9340*$_{P225\Delta23}$ (1X) n'entraîne pas de modification : la sonde *srn_9340*$_{P225}$ a un profil identique au puits 2 et 3 (fixation de SarA sur *srn_9340*$_{P225}$ sans compétiteur). Lorsque nous augmentons la quantité de *srn_9340*$_{P225\Delta23}$ dans la réaction de fixation, la totalité de la sonde *srn_9340*$_{P225}$ est retardée au niveau du premier retard. Contrairement à la sonde complète, la délétion des 23 nucléotides ne permet pas d'entrer en compétition avec la sonde radiomarquée pour la fixation spécifique de SarA.

L'ensemble des expériences précédentes indiquent que le facteur de transcription SarA réprime l'expression de *srn_9340,* comme pour *srn_3610_sprC*. Nous avons montré que la protéine SarA fixe *in vitro* les promoteurs de *srn_3610_sprC* et *srn_9340* permettant ainsi de réprimer leur expression. Nous avons voulu étudier l'hypothèse d'une fixation *in vivo* de SarA sur ces promoteurs.

V. Etude de la fixation de SarA *in vivo*

A. SarA fixe les promoteurs de *srn_3610_sprC* et *srn_9340 in vivo*

Afin de vérifier que la fixation de SarA sur les promoteurs de *srn_3610_sprC* et *srn_9340* observée *in vitro* est une réalité *in vivo*, des expériences d'immunoprécipitation de la chromatine (ChIP) ont été réalisées.

Le ChIP est une technique qui permet de vérifier la fixation d'une protéine d'intérêt sur une région spécifique du génome à un moment précis de croissance. Au moment choisi pour l'expérience, du formaldéhyde est ajouté à la culture bactérienne. Le formaldéhyde a pour effet de lier de manière covalente les protéines à l'ADN, la situation chromosomique est alors figée. Ensuite, les bactéries sont détruites, le chromosome bactérien et les protéines associées sont récupérés. L'ADN est par la suite fragmenté par sonication. Des paramètres de sonication ont été déterminés afin que, dans notre cas, l'ADN soit fragmenté en morceaux d'environ 500pb. Une immunoprécipitation de la protéine d'intérêt est effectuée grâce à un anticorps spécifique. A cette étape, si la protéine est liée à des fragments d'ADN, ces derniers ont alors été précipités avec la protéine ciblée. Après dissociation de la liaison ADN-Protéine, une recherche de la présence des promoteurs cibles est réalisée par PCR quantitative.

Suite aux difficultés rencontrées pour immunoprécipiter SarA avec l'anticorps polyclonal anti-SarA (disponible au laboratoire), nous avons décidé de produire la protéine SarA fusionnée à une étiquette composée de 6 histidines (6His). L'insertion de 6 codons pour l'histidine à l'extrémité 3' du cadre ouvert de lecture (ORF) permet la transcription et la production d'une protéine SarA-6His. Le gène *sarA-6His* porté par le plasmide pCN38 a été transformé dans la souche HG003Δ*sarA*. Parallèlement, le plasmide vide a été introduit dans cette dernière.

En suivant le protocole présenté Figure 65, nous récupérons une partie de l'échantillon soniqué (fraction INPUT) correspondant à l'ensemble du chromosome fragmenté, cettte fraction de l'échantillon sert de contrôle positif. Ensuite, le reste de la fraction est divisé en trois fractions différentes (ChIP, DIG et Pol). L'immunoprécipitation d'une protéine, différente pour chaque fraction, par un anticoprs qui lui est spécifique est effectuée. L'immunoprécipitation est réalisée soit avec les anticorps anti-6His (ciblant SarA-6His, fraction ChIP), anti-σ^{70} (ciblant σ^{A}, homologue de σ^{70} d'*E. coli ;* fraction Pol) ou anti-digoxigénine (fraction DIG). L'anticorps anti-digoxigénine est utilisé en tant que contrôle négatif, la digoxigénine n'étant pas une protéine présente chez *S. aureus*.

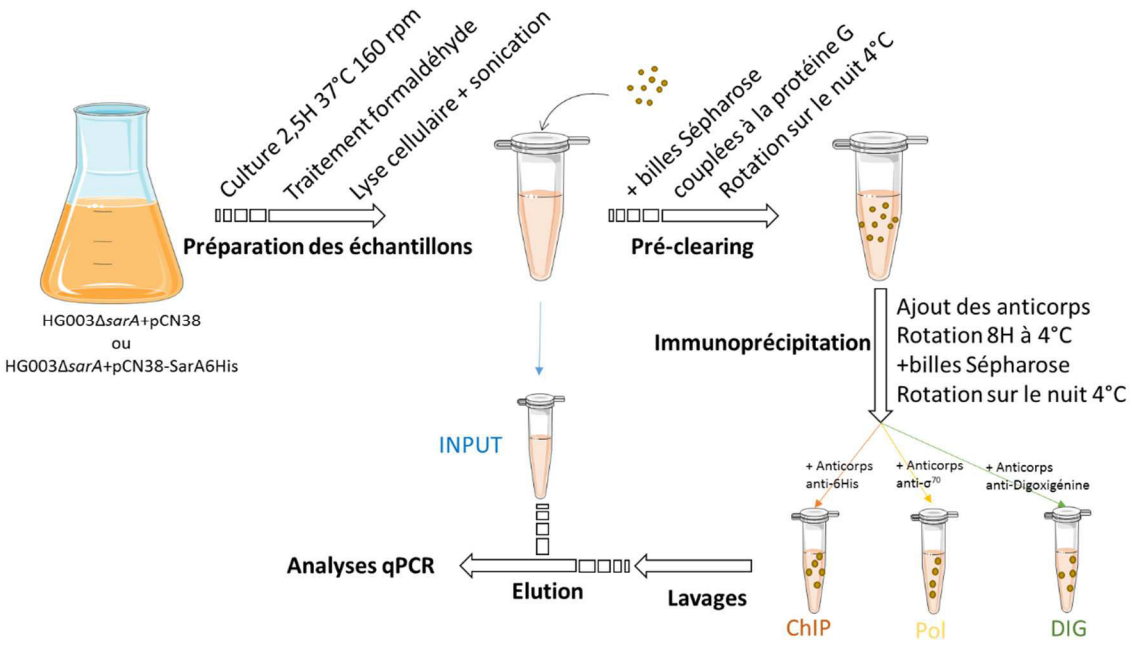


Figure 65 : Protocole d'immunoprécipitation de la chromatine.

Après dissociation des complexes ADN-Protéines et élution des fragments d'ADN, les promoteurs de *srn_3610_sprC*, *srn_9340* et *hla* sont recherchés par PCRq dans les fractions INPUT, ChIP et DIG (Figure 66). La technique de PCRq nous permet de déterminer, par la valeur du Ct, la quantité de fragments d'ADN (choisis par des amorces spécifiques) présent dans les différentes souches pour les différentes fractions. La souche HG003Δ*sarA*+pCN38 nous permet de normaliser les valeurs obtenues pour la souche HG003Δ*sarA*+pCN38-SarA6His et sert donc de référence pour l'expérience.

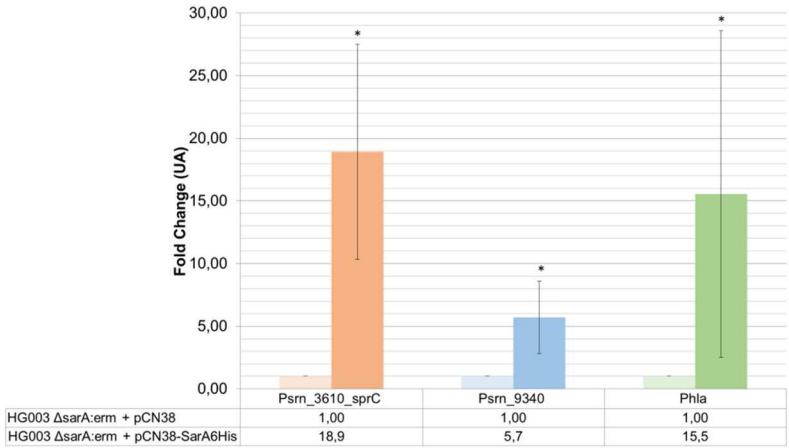


Figure 66 : SarA fixe *in vivo* les promoteurs de *srn_3610_sprC, srn_9340* et *hla*. L'immunoprécipitation a été réalisée contre la protéine SarA-6His avec un anitcorps anti-6His. La quantité d'ADN correspondant à ces promoteurs a été mesurée par PCR quantitative dans les souches HG003Δ*sarA* pCN38 (couleurs claires) et HG003Δ*sarA* pCN38-SarA6His (couleurs foncées). Après normalisation, les enrichissements de trois expériences indépendantes ont été représentés (test unilatéral de Mann-Whitney : * : p-value<0,05)

La quantité des promoteurs amplifiés de la souche HG003Δ*sarA*+pCN38 a été normalisée à la valeur de 1. Dans la souche HG003Δ*sarA*+pCN38-SarA6His, la quantité de promoteurs de *srn_3610_sprC*, *srn_9340* et *hla* est augmentée significativement et correspond à un enrichissement de la présence de ces promoteurs en présence de SarA-6His (respectivement aux valeurs de 18,9, 5,7 et 15,5).

En conclusion, l'enrichissement des promoteurs $P_{sprC,}$ P_{srn_9340} et P_{hla} en présence de la protéine SarA fusionnée à l'étiquette 6 histidines permet de conclure que cette dernière est capable de fixer *in vivo* les promoteurs étudiés. De plus, nous avons vu qu'*in vitro* les sites de fixations de SarA sur les 2 promoteurs, englobent ou chevauchent le site de fixation de l'ARN polymérase (core-enzyme + facteur σ^A). Nous avons donc voulu savoir si la fixation de SarA entraîne une gêne pour la fixation de la sous-unité σ^A.

B. SarA empêche la fixation de l'ARN polymérase *in vivo*

Pour étudier l'hypothèse que la fixation de SarA sur les promoteurs de *srn_3610_sprC* et *srn_9340* empêche la fixation de l'ARN polymérase, nous avons immunoprécipité la sous-unité σ^A présente dans les extraits par l'anticorps anti-σ^{70} de *E. coli* capable, d'après le fournisseur, de reconnaître la sous-unité σ^A chez *S.aureus*. Cette étude a été réalisée dans la souche HG003Δ*sarA*+pCN38 (souche ne produisant pas SarA) ou dans la souche HG003Δ*sarA*+pCN38-SarA6His (souche produisant exclusivement SarA étiqueté par 6His en C-terminale.

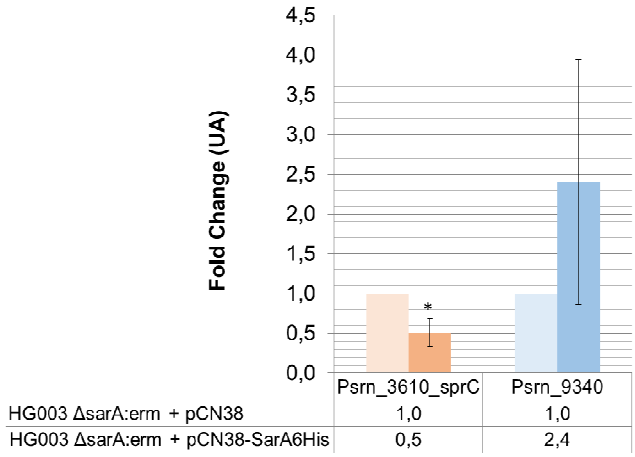


Figure 67 : SarA gêne la fixation de l'ARN polymérase pour le promoteur *srn_3610_sprC*. L'immunoprécipitation a été réalisée avec un anitcorps anti-σ^{70}. La quantité d'ADN correspondant aux promoteurs *srn_3610_sprC* et *srn_9340* a été mesurée par PCR quantitative dans les souches HG003Δ*sarA* pCN38 (couleurs claires) et HG003Δ*sarA* pCN38-SarA6His (couleurs foncées). Après normalisation, les enrichissements de trois expériences indépendantes ont été représentés (test unilatéral de Mann-Whitney : * : p-value<0,05)

En présence de la protéine SarA-6His, nous observons un moindre enrichissement du promoteur de *srn_3610_sprC* (diminution de 2 fois par rapport à la souche HG003Δ*sarA*+pCN38) (Figure 67). Cela signifie qu'il y a moins de promoteurs de *srn_3610_sprC* liés avec la sous-unité σ^A en présence de SarA-6His par rapport à une souche où SarA n'est pas produite. SarA semble donc gêner la fixation de la sous-unité σ^A pour P_{sprC}. Par contre, aucune différence significative n'est observée pour le promoteur de *srn_9340*. Nous ne pouvons donc pas conclure sur l'action de SarA sur la fixation de l'ARN polymérase dans le cas de *srn_9340.*

Si les sites de fixations de SarA et σ^A ne sont pas responsables de ces différences entre *srn_3610_sprC* et *srn_9340*, l'organisation génomique ainsi que la technologie d'immunoprécipitation de la chromatine le sont peut-être. En effet, une différence majeure entre *srn_3610_sprC* et *srn_9340* vient de l'environnement génomique : *srn_3610_sprC* est isolé de part et d'autre et les gènes environnants semblent peu transcrits. Un regroupement d'ARNt, fortement transcrits, est positionné à 408 pb en amont de *srn_9340*. Moins de 500 pb séparent donc la fin du dernier gène transcrit en ARNt (ARNt-sérine) et le site de fixation de SarA sur *srn_9340*. Or, l'étape de sonication préalable à l'immunoprécipitation a été optimisée pour donner des fragments de 500 pb en moyenne. L'enrichissement par PCRq de *srn_9340* pourrait donc être le résultat de trois types d'immunoprécipitation de σ^A. La première correspond à la fixation de σ^A sur le promoteur de *srn_9340* (Figure 68 ①). La seconde est la fixation de σ^A sur le promoteur du gène de l'ARNt-sérine sans fixation de σ^A sur le promoteur de *srn_9340* (Figure 68 ②). Enfin, des fragments d'ADN peuvent avoir été immunoprécipités à partir d'une double fixation de σ^A sur *srn_9340* et le gène de l'ARNt (Figure 68 ③). Le fait qu'il soit possible d'immunoprécipiter la région promotrice de *srn_9340*, sans que σ^A ne lui soit lié, ne permet pas d'étudier l'influence de SarA sur la fixation de σ^A. Ce phénomène montre les limites des expériences de ChIP pour certaines fixations ADN-protéines.

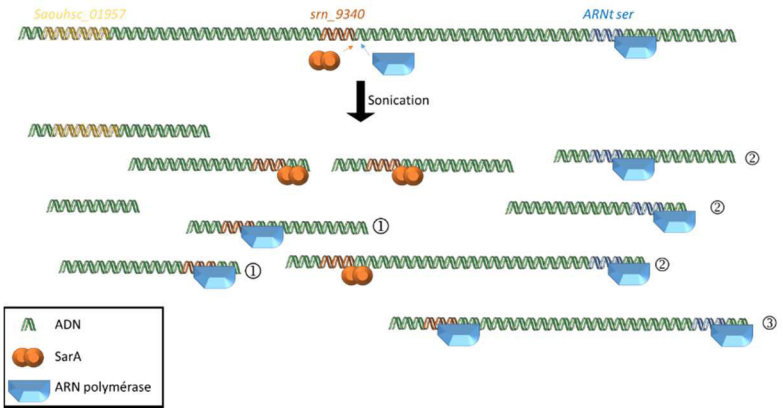


Figure 68 : Représentation des différentes possibilités de fixation de SarA ou de l'ARN polymérase sur le promoteur de *srn_9340*.

Afin de vérifier qu'il n'existe pas de différences entre les promoteurs de *srn_3610_sprC* et *srn*_9340 quant au mode d'action de SarA des études complémentaires de la fixation de l'ARN polymérase *in vitro* ont été effectuées.

C. L'ARN polymérase et SarA se fixent sur les promoteurs de *srn_3610_sprC* et *srn_9340* de manière mutuellement exclusive

La première étape a été de confirmer que l'ARN polymérase d'*E. coli* est capable de fixer *in vitro* $sprC_{P267}$ et srn_9340_{P225}. Pour cela, nous avons mis en contact les sondes $sprC_{P267}$ et srn_9340_{P225} avec de l'ARN polymérase d'*E. coli* (Holoenzyme contenant la sous unité σ^{70} de New England Biolabs).

Le puits 1 de la Figure 69 (panels A et B) correspond la migration de la sonde sans ARN polymérase. Nous voyons qu'en présence de concentrations croissante d'ARN

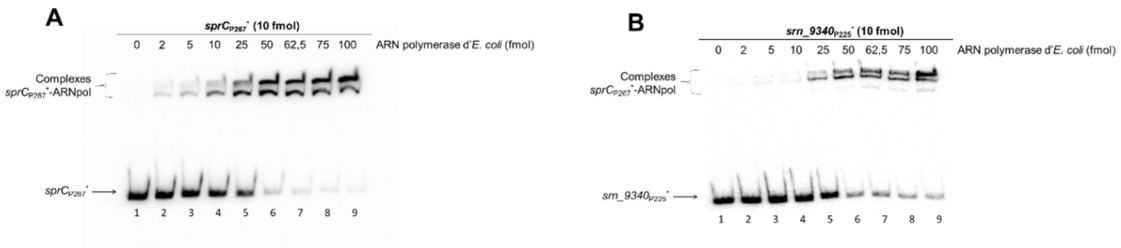


Figure 69 : L'ARN polymérase d'*E. coli* est capable de fixer les promoteurs de *srn_3610_sprC* (A) et *srn_9340* (B). La sonde $sprC_{P225}$ ou srn_9340_{P225} (10 fmol) est mise en présence d'une quantité croissante d'ARN polymérase d'*E. coli* (de 2 à 100 fmol : puits 2 à 9).

polymérase d'*E. coli* (puits 2 à 10), les sondes $sprC_{P267}$ et srn_9340_{P225} sont retardées avec dans les deux cas, la présence de deux bandes de très haut poids moléculaire (l'ARN polymérase a une masse moléculaire d'environ 470kDa). L'affinité de l'ARN polymérase d'*E. coli* est donné par la constante de dissociation K_D lequel est de 0,76 nM pour $sprC_{P267}$ et de 1,85 nM pour srn_9340_{P225}. Ainsi, l'ARN polymérase d'*E. coli* est capable de fixer $sprC_{P267}$ et srn_9340_{P225} avec de meilleures affinités que celles de SarA pour les mêmes séquences d'ADN.

Comme nous l'avons montré précédemment avec les expériences d'empreinte à la DNase I, les sites de fixation de SarA sur les promoteurs de *srn_3610_sprC* et *srn_9340* englobent les sites de fixation de l'ARN polymérase suggérant que SarA et l'ARN polymérase se fixeraient de manière mutuellement exclusive. Pour tester cette hypothèse, nous avons réalisé une réaction de retard entre 10 fmol de promoteur (*srn_3610_sprC* ou *srn_9340*), l'ARN polymérase d'*E. coli* contenant le facteur σ^{70} (50 à 100 fmol) avec ou sans ajout de protéine SarA (0,5 pmol).

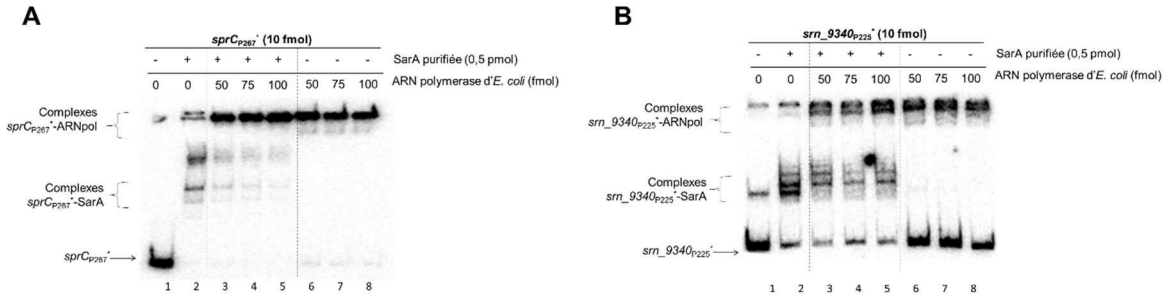


Figure 70 : SarA et l'ARN polymérase fixent de manière mutuellement exclusive les promoteurs de *srn_3610_sprC* et *srn_9340*. 10 fmol de sonde *sprC*P267 **(A)** ou *srn_9340*P225 **(B)** sont mis en contact avec des quantités croissantes d'ARN polymérase d'*E. coli* (de 50 à 100 fmol : puits 6 à 8) et de protéine SarA purifiée (0,5 pmol : puits 3 à 5).

A la Figure 70 (panels A et B), les expériences de retard sur gel pour les promoteurs *srn_3610_sprC* et *srn_9340* sont présentées. Dans le puits 2, nous observons le retard des promoteurs en présence de SarA. Nous constatons de multiples bandes correspondant à la fixation de SarA. Aux puits 6 à 8, nous observons également le retard des sondes en présence de l'holoenzyme d'*E. coli*. Dès 50 fmol d'holoenzyme, la quasi-totalité de la sonde *sprC*P267 est retardée. Pour *srn_9340*P225, une fraction de la sonde reste libre quelle que soit la quantité d'ARN polymérase ajoutée. Les puits 3 à 5 correspondent à l'ajout de l'holoenzyme dans le mélange de retard (sonde ADN avec SarA). Pour *sprC*P267 comme pour *srn_9340*P225, nous avons la présence des bandes correspondant aux complexes SarA-ADN ainsi que celles correspondant aux complexes ARN polymérase-ADN.

Avec l'ajout de SarA et de l'ARN polymérase, nous nous attendions à trois cas de figures différents :

1- SarA se fixe seule sur la sonde : un retard correspondant à ce complexe est présent

2- L'ARN polymérase se fixe seule sur la sonde : un retard à haut poids moléculaire apparaît

3- Les deux protéines se fixent : un « super-retard » est observé.

Du fait de la faible migration du complexe ARN polymérase-Sonde, il nous est apparu difficile, voire impossible de faire la différence entre le « super-retard » hypothétique correspondant au troisième cas et le retard ARN-polymérase-sonde. Néanmoins, nous avons vu que malgré l'ajout simultané de SarA et l'ARN polymérase, le complexe SarA-sonde est toujours présent. Cette observation prouve que des sondes *sprC*P267 ou *srn_9340*P225 ont seulement fixé SarA alors qu'au vu des quantités de polymérase utilisée toutes les sondes devraient avoir accroché l'holoenzyme. La seule explication que nous ayons à cette observation est que la fixation de SarA a empêché celle de l'ARN polymérase et donc que la

fixation de SarA et de l'ARN polymérase sur *sprC*$_{P267}$ ou *srn_9340*$_{P225}$ est mutuellement exclusive.

Pour finaliser l'étude de la régulation de SarA sur l'activité de l'ARN polymérase, nous nous sommes intéressés à la transcription *in vitro* de *srn_3610_sprC* et *srn_9340* en présence de SarA.

D. SarA et la transcription *in vitro* de *srn_3610_sprC* et *srn_9340*

Dans les expériences de transcription *in vitro,* nous avons utilisé deux fragments d'ADN, *sprC*$_{456}$ et *srn_9340*$_{337}$ correspondant aux séquences promotrices, géniques et terminatrices de *srn_3610_sprC* et *srn_9340*. La réaction de transcription a été réalisée avec 100 fmol d'ARN polymérase d'*E. coli* ainsi que les nucléosides ATP, GTP et CTP et le nucléoside radiomarqué [α-^{32}P] UTP.

En absence de SarA (Figure 71 ; puits de gauche), nous observons trois transcrits de tailles différentes pour *srn_3610_sprC* (panel A) et de deux transcrits pour *snr_9340* (panel B). Comme *srn_3610_sprC* ne produit qu'un seul transcrit *in vivo*, nous supposons que les différents transcrits sont le résultat d'arrêts de transcription précoces. Cependant, l'utilisation d'une ARN polymérase d'*E. coli* a permis la transcription *in vitro* de gènes issus de *S. aureus* et dans notre cas des gènes *srn_3610_sprC* et *srn_9340.*

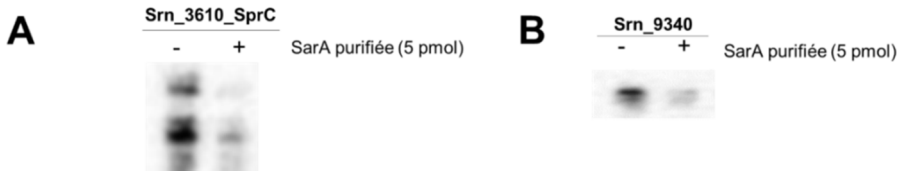


Figure 71 : SarA réprime la transcription *in vitro* de *srn_3610_sprC* et *srn_9340*. Les expériences de transcription *in vitro* ont été effectuées 10 fmol de sonde *sprC*$_{456}$ **(A)** ou *srn_9340*$_{337}$ **(B)** et 100 fmol d'ARN polymérase de *E. coli*. 5 pmol de SarA purifiée sont ajoutés quand indiqué (+).

Pour tester l'effet répresseur de SarA sur la transcription de *srn_3610_sprC* et *srn_9340,* 5 pmol de SarA purifiée sont ajoutés à la réaction de transcription *in vitro.* Lorsque SarA est ajoutée à la réaction de transcription (Figure 71 ; puits de droite), nous constatons une diminution importante de la quantité de transcrits produits aussi bien pour *srn_3610_sprC* que pour *srn_9340.* Dans les deux cas, l'addition de SarA à la réaction de transcription *in vitro* a eu pour effet diminuer la quantité de transcrits produits. Nous pouvons conclure que SarA est capable de réprimer la transcription *in vitro* des deux gènes étudiés.

Au vu de l'ensemble des résultats obtenus, nous pouvons maintenant intégrer la régulation de *srn_3610_sprC* et *srn_9340* dans un schéma général de la régulation dans *S. aureus*.

VI. Schéma général de la régulation de l'expression de *srn_3610_sprC* et *srn_9340*

L'ensemble des résultats obtenus précédemment sur l'expression de *srn_3610_sprC* et *srn_9340* dans différents contextes mutants et ceux obtenus *in vitro* avec SarA ou l'ARN polymérase nous permettent de proposer un schéma plus général de la régulation des deux gènes d'intérêt (Figure 72). SarR est connu pour être un répresseur de la transcription de *sarA* et d'*agr*. Nous avons montré les actions positives indirectes du facteur de transcription SarR et de l'ARN régulateur ARNIII sur l'expression de *srn_3610_sprC* et *srn_9340*. Par ailleurs, le facteur de transcription SarA est capable d'activer sa propre transcription et celle *d'agr*. De plus, nous avons montré que SarA, est un répresseur direct fort de l'expression de *srn_3610_sprC* et *srn_9340*.

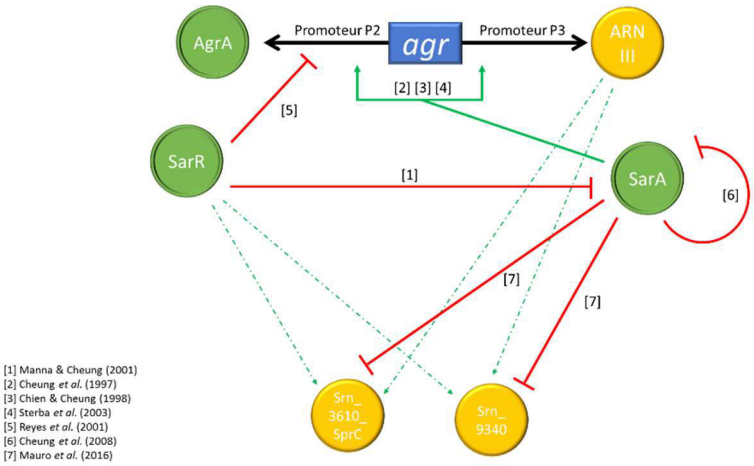


Figure 72 : Interrégulation de *srn_3610_sprC* et *srn_9340* par SarA, SarR et ARNIII. Les activations et les répressions sont représentées respectivement en vert et en rouge. Les actions directes sont identifiées par des traits pleins. Les traits pointillés correspondent aux régulations indirectes.

Nous avons donc focalisé notre attention sur le mécanisme de répression de SarA sur les deux ARN régulateurs. SarA est capable de fixer *in vitro* et *in vivo* le promoteur de *srn_3610_sprC* et *srn_9340* au niveau des régions spécifiques de ceux-ci : les boîtes -10 et -35. Ces boîtes sont le siège de la fixation du facteur σ^A de l'ARN polymérase. Nous avons pu montrer que la fixation sur ces régions promotrices de SarA et σ^A est mutuellement exclusive.

Enfin, Le Pabic *et al.* ont montré que la transcription de *srn_3610_sprC*, forte aux premières minutes de la phagocytose, est rapidement réprimée après internalisation des bactéries (Le Pabic et al., 2015). Nous proposons donc le modèle suivant pour la régulation de l'expression des gènes *srn_3610_sprC* et *srn_9340*.

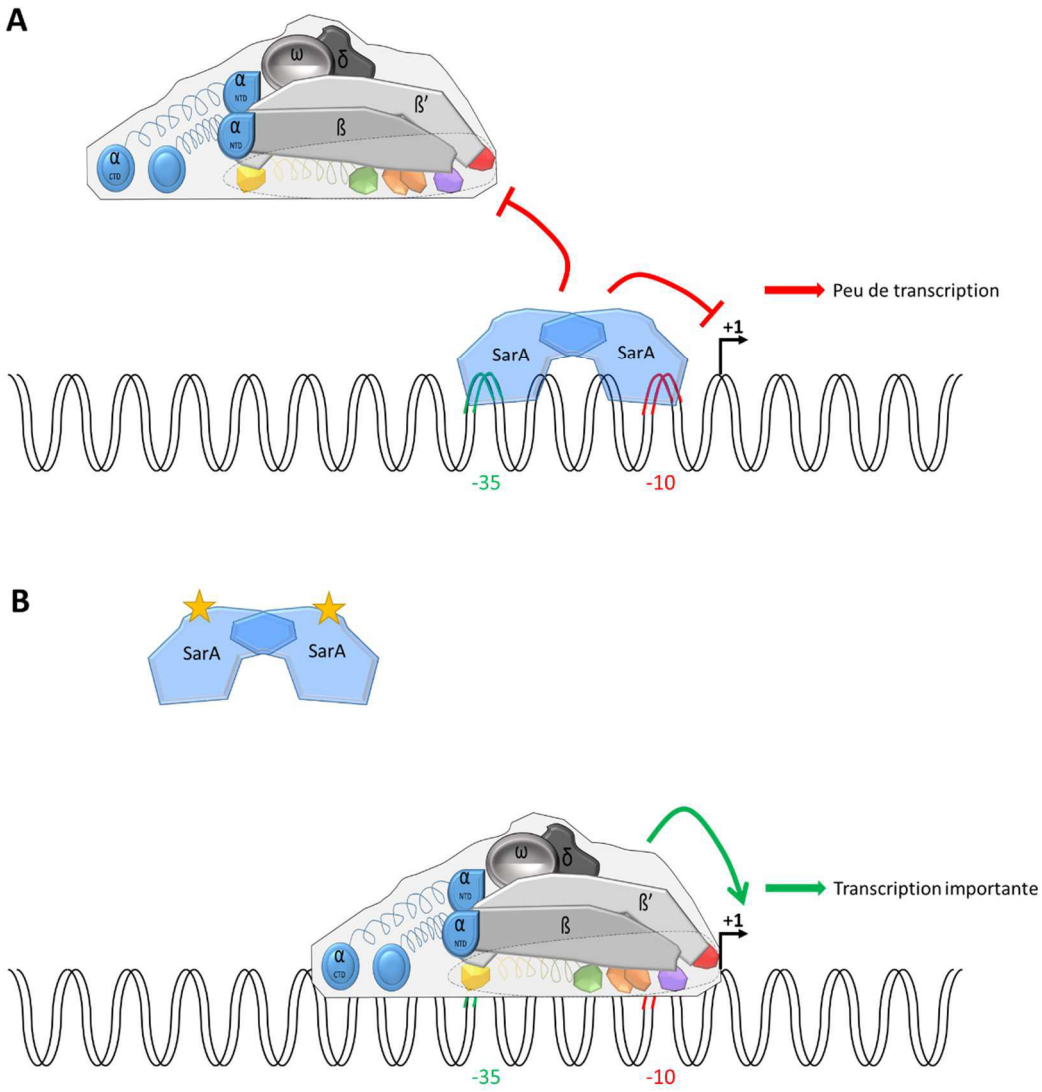


Figure 73 : Schéma général de la répression de la transcription de *srn_3610_sprC* et *srn_9340* par le facteur de transcription SarA.

Dans les conditions conventionnelles de croissance effectuées au laboratoire, le niveau d'expression de ces gènes est très faible. SarA est fixée sur leur promoteur empêchant la transcription (Figure 73A).

L'arrivée d'un signal (comme le contact cellule-hôte lors de la phagocytose par exemple) entraînerait le décrochage de SarA du promoteur permettant à l'ARN polymérase de se fixer et de transcrire fortement les deux gènes. (Figure 73B)

En conclusion, l'expression de *srn_3610_sprC* et *srn_9340* est fortement réprimée par SarA mais il a été montré que l'expression de *srn_3610_sprC*, pouvait être forte dans certaines conditions environnementales. Nous avons donc recherché des signaux permettant la dérépression de *srn_3610_sprC* et *srn_9340* par SarA.

PARTIE II :

Recherche du signal de dérépression de la transcription de *srn_3610_sprC* et *srn_9340*

RESULTATS ET DISCUSSIONS

Comme nous l'avons montré précédemment, l'expression de *srn_3610_sprc* et *srn_9340* est très fortement réprimée dans les conditions conventionnelles de croissance. Cette répression fait intervenir la protéine SarA. Notre hypothèse est que dans certaines conditions, SarA se décroche des promoteurs de *srn_3610_sprC* et *srn_9340* permettant l'expression de ces petits ARN, de manière transitoire ou prolongée. Srn_3610_SprC et Srn_9340, fortement transcrits, pourraient alors remplir leur(s) fonction(s). Ce chapitre résumera l'ensemble de différentes expériences menées pour identifier le ou les stimuli capables de permettre la levée de répression de SarA sur l'activité des promoteurs de *srn_3610_sprC* et *srn_9340.*

I. Etude de la levée de répression de SarA

A. Expression de *srn_3610_sprC* et *srn_9340* au cours du temps

Avant même d'introduire différents stress aux bactéries, nous avons souhaité étudier l'expression de *srn_3610_sprC* et *srn_9340* au cours de la croissance. Cette étude a été réalisée sur les souches HG003 sauvage et HG003Δ*sarA*. Les ARN totaux ont été extraits toutes les heures pendant 9 heures (Figure 74).

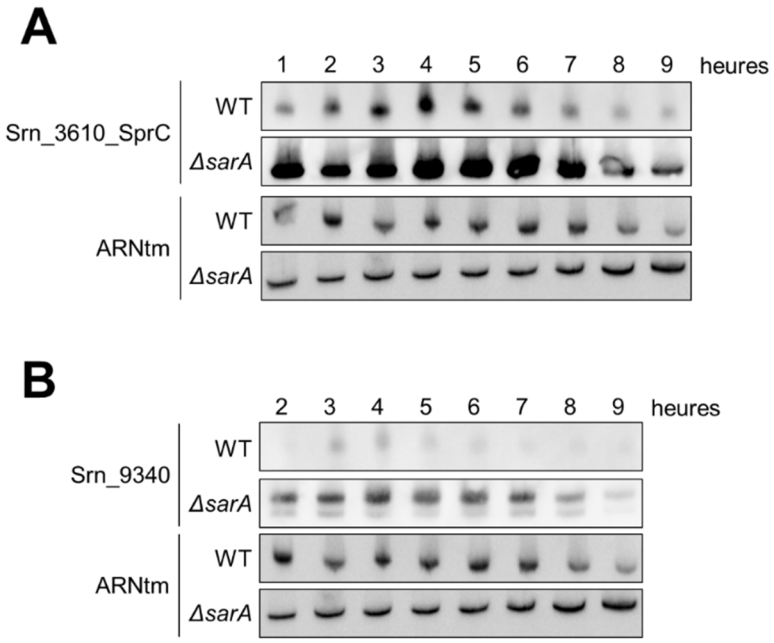


Figure 74 : Expression de *srn_3610_sprC* (A) et *srn_9340* (B) dans les souches HG003 (WT) et HG003Δ*sarA* au cours du temps. Les ARN ont extraits depuis des cultures bactériennes en milieu BHI.

Dans la souche sauvage, la quantité de Srn_3610_SprC et de Srn_9340 atteint un maximum vers 4 heures de croissance. Pour la souche HG003Δ*sarA*, nous constatons de légères fluctuations de la quantité des ARN Srn_3610_SprC et Srn_9340 tout au long de la croissance avec un profil similaire à la souche HG003. Le pic d'expression, dans la souche HG003Δ*sarA*, est également atteint à 4 heures de croissance pour les deux gènes. De plus, l'augmentation des quantités d'ARN Srn_3610_SprC et Srn_9340, observées dans les études précédentes à 5 H de culture (Figure 30 et Figure 53), est visible tout au long de la croissance (souche HG003Δ*sarA* vs HG003). Le fait que les fluctuations d'expression observées dans la souche HG003Δ*sarA* soient similaires à celles observées dans la souche HG003 sauvage (avec maximum d'expression vers 4-5 heures) indique qu'elles sont indépendantes de SarA.

Par ailleurs, les fluctuations, observées au cours de la croissance dans la souche sauvage ne semblent pas non plus liées à la dissociation de SarA des promoteurs étudiés (*srn_3610_sprC* ou *srn_9340*). En effet, le décrochage de SarA doit conduire à une augmentation de la quantité de transcrits Srn_3610_SprC et de Srn_9340 d'un facteur respectivement de 10 ou 3,5 fois par rapport à la quantité initiale. La quantité d'ARN dans la souche HG003 aurait alors été identique à celle dans la souche HG003Δ*sarA*. Or ici, nous n'observons pas ce phénomène (Figure 74). Il n'y a donc pas de différence majeure dans l'expression de *srn_3610_sprC* et *srn_9340* au cours de la croissance pour la souche HG003. Dans les conditions conventionnelles de croissance, SarA réprime la transcription des gènes *srn_3610_sprC* ou *srn_9340*.

Afin de suivre l'activité des promoteurs de *srn_3610_sprC* et *srn_9340* au cours de la croissance et dans différentes conditions de culture, nous avons mis en place des expériences de gène rapporteur. Nous avons choisi d'utiliser la protéine rapportrice *gfp* dont l'activité sera le témoin de l'activité de ces deux promoteurs.

B. Utilisation du gène rapporteur *gfp*, témoin de l'expression de *srn_3610_sprC* et *srn_9340*

Afin de faciliter l'étude de l'expression des gènes *srn_3610_sprC* et *srn_9340*, les promoteurs de 144pb de *srn_3610_sprC* (P_{sprC}) ou *srn_9340* (P_{srn_9340}) ont été insérés en amont du gène *gfp* codant une protéine fluorescente (Green Fluorescent Protein) (Figure 75).

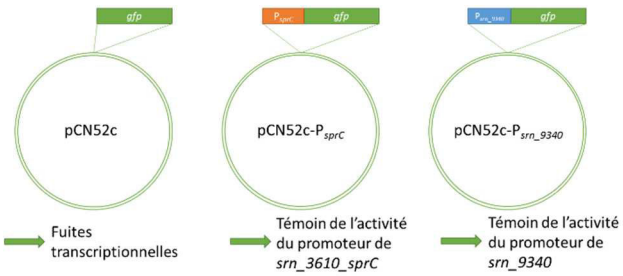


Figure 75 : Représentation des trois plasmides utilisés pour l'expérience de gène rapporteur *gfp*.

La fluorescence mesurée, qui peut se faire sans manipulation des bactéries (pas d'extraction protéique comme pour les expériences de ß-lactamase), sera donc en corrélation avec l'activité des deux promoteurs.

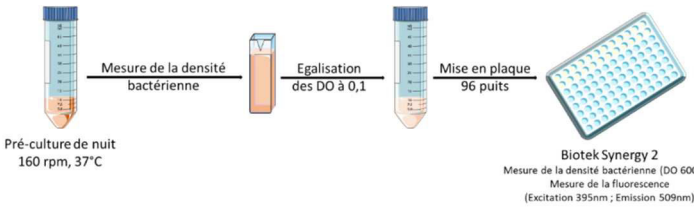


Figure 76 : Protocole d'expérience de gène rapporteur *gfp*.

Les expériences ont été réalisées dans un spectromètre thermostaté à 37°C (Biotek) (Figure 76). Les souches HG003 et HG003Δ*sarA* ont été transformées avec le plasmide pCN52c (plasmide vide), pCN52c-P_{sprC} (promoteur de *srn_3610_sprC* placé en amont du gène *gfp*) ou le pCN52c-P_{srn_9340} (promoteur de *srn_9340* placé en amont du gène *gfp*). La fluorescence mesurée pour les souches contenant pCN52c correspond à l'autofluorescence des bactéries ainsi qu'aux fuites transcriptionnelles qui pourraient être responsables de la production de GFP. Les valeurs d'autofluorescence pour une souche donnée doivent être soustraites à celles mesurées en présence de promoteurs dans la même souche.

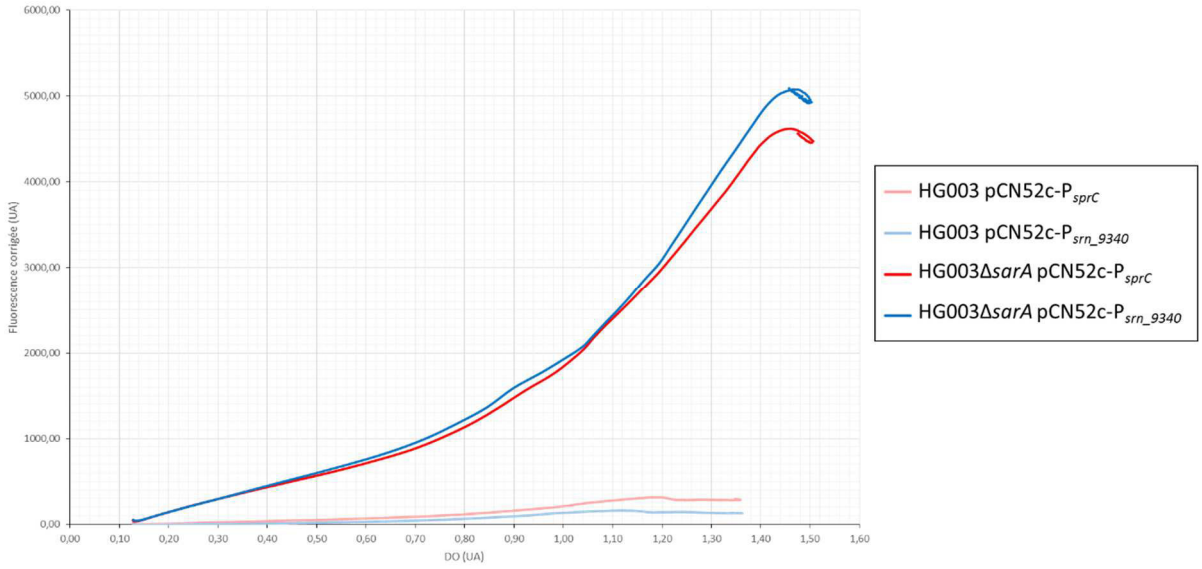


Figure 77 : Expression des promoteurs de P_{sprC} et de P_{srn_9340} dans la souche HG003 et de HG003ΔsarA. La fluorescence émise dans les souches HG003 et *HG003ΔsarA* est représentée respectivement par des courbes de couleurs claires et foncées (rouge pour *srn_3610_sprC* et bleue pour *srn_9340*).

La Figure 77 présente les courbes de fluorescence (corrigées par rapport aux valeurs obtenues pour les fuites transcriptionnelles) en fonction de la croissance bactérienne (mesurée par la densité optique à 600 nm) dans les souches HG003 et HG003Δ*sarA*. Comme attendu, l'absence de SarA (Figure 77, courbes foncées) entraîne une augmentation de l'activité

transcriptionnelle des promoteurs P_{sprC} et P_{srn_9340}. La différence d'activité entre la souche HG003 (Figure 77, courbes claires) et HG003Δ*sarA* (Figure 77, courbes foncées) est de 10 fois pour le promoteur de *srn_3610_sprC* et de 15 fois pour celui de *srn_9340.* Ces résultats sont en accord avec ceux obtenus dans les expériences de gène rapporteur précédentes (ß-lactamase) (Figure 37 et Figure 57).

Ces résultats nous ont permis de vérifier qu'il était possible de différencier l'activité des promoteurs dans les souches HG003 et HG003Δ*sarA* lorsqu'ls étaient fusionnés à la GFP. Ainsi, nous avons validé l'utilisation de cette technologie pour rechercher des signaux qui entraineraient une augmentaiton de la fluorescence observée donc d'une dérépression transcriptionnelle des promoteurs étudiés. De plus, les niveaux de fluorescence observée en contexte sauvage (HG003) bien que faibles sont suffisants pour la poursuite de l'étude qui cherche à identifier le signal permettant une augmentation de la fluorescence.

L'utilisation de la fluorescence produite par la protéine GFP pour tracer l'activité des promoteurs P_{sprC} et P_{srn_9340} étant possible, nous avons donc décidé de rechercher, dans les souches portant le plasmide rapporteur, un signal capable d'engendrer une levée de répression de l'expression de *srn_3610_sprC* et *srn_9340*. Cette levée de répression transcriptionnelle se traduirait, dans notre expérience de gène rapporteur, par une augmentation du niveau de fluorescence. Cette augmentation pourrait aboutir à des niveaux de fluorescence identiques à ceux observés dans la souche HG003Δ*sarA*. Bien évidemment, si une diminution de la fluorescence est observée dans la souche HG003Δ*sarA*, elle correspondrait à une diminution de l'activité transcriptionnelle du promoteur concerné (fixation d'un nouveau répresseur ? décrochage d'un activateur de la transcription ?). Les différentes possibilités d'évolution de fluorescence ont été résumées à la Figure 78.

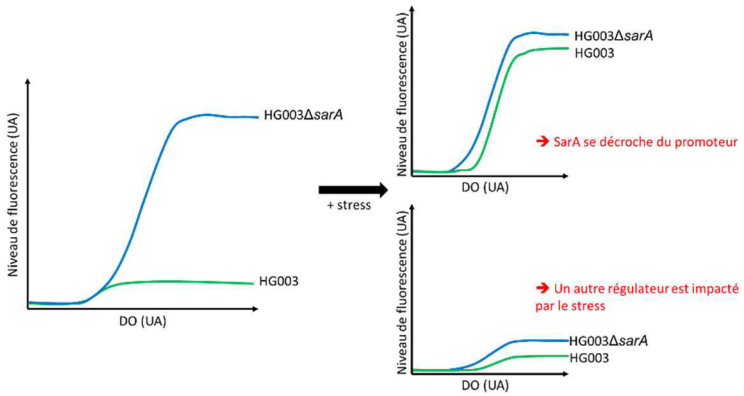


Figure 78 : Représentation graphique des hypothèses de régulation transcriptionnelle de *srn_3610_sprC* et *srn_9340.*

Nous avons alors souhaité rechercher les conditions qui permettent une levée de répression de l'expression de *srn_3610_sprC* et *srn_9340*. Pour cela nous avons choisi d'étudier l'activité des deux promoteurs dans différents contextes mimant différents environnements que la bactérie pourrait rencontrer.

C. Expression de P_{sprC} dans différents milieux de culture

Le premier signal étudié est un signal dû à un changement dans la composition du milieu. Nous avons choisi de comparer les milieux Brain Heart Infusion (BHI ; milieu riche), Luria-Bertoni (LB ; milieu intermédiaire), Trypton Soy Broth (TSB ; milieu intermédiaire) et Mueller-Hinton (MH ; milieu pauvre) (Figure 79). Les mesures de DO et de fluorescence ont été effectuées en Biotek. Il est important de noter ici que la densité optique maximale mesurée, d'une culture bactérienne, dans le Biotek est différente de celle mesurée pour des bactéries mises en culture dans des erlenmeyers. Néanmoins, quel que soit le type de culture effectué, les différentes étapes de croissance sont présentes (phase de latence, exponentielle, post-exponentielle et stationnaire) (Figure 80).

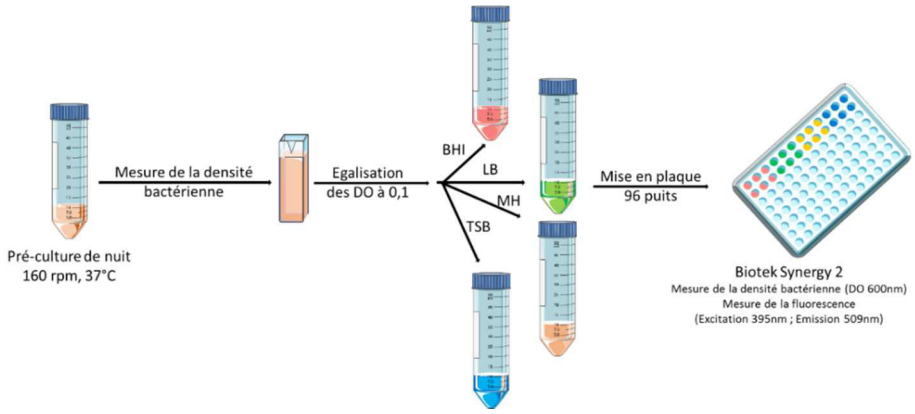


Figure 79 : Protocole d'expérience de gène rapporteur *gfp* dans différents milieux de culture.

Dans la Figure 80, nous confirmons que la croissance bactérienne est plus faible dans les milieux MH, TSB et LB par rapport au milieu BHI. En effet, dans le milieu riche BHI, la DO maximale (DO max) atteinte, qui témoigne de la biomasse, est d'environ 1,2 alors que pour les milieux intermédiaires (LB et TSB), la DO max est comprise entre 0,8 et 1. Pour le milieu MH, la DO max pour la souche HG003 est de 0,5. Il est intéressant de noter que dans le milieu MH, la souche HG003Δ*sarA* atteint une plus grande densité optique maximale (environ 0,75) que la souche sauvage HG003. La délétion de *sarA* serait donc un avantage pour la bactérie dans un milieu pauvre.

De manière générale quel que soit le milieu de culture, le niveau de fluorescence rapporté à la DO est toujours plus élevé dans les souches HG003Δ*sarA* pCN52c-P_{sprC} que

dans les souches HG003 pCN52c-P_{sprC} (Figure 80). Dans les conditions de culture utilisées, le niveau de transcription de *srn_3610_sprC* est toujours faible dans la souche HG003. SarA est donc capable de réprimer la transcription de *sprC* dans les milieux testés. De plus, le fait que la fluorescence des souches HG003Δ*sarA* n'est pas impactée par les différents milieux, signifie que la composition des milieux ne modifie pas la transcription de *srn_3610_sprC* dans la souche HG003Δ*sarA*. Cette observation montre alors que, dans les conditions de culture testées, l'activité transcriptionnelle de *srn_3610_sprC* n'est pas modifiée par un autre régulateur.

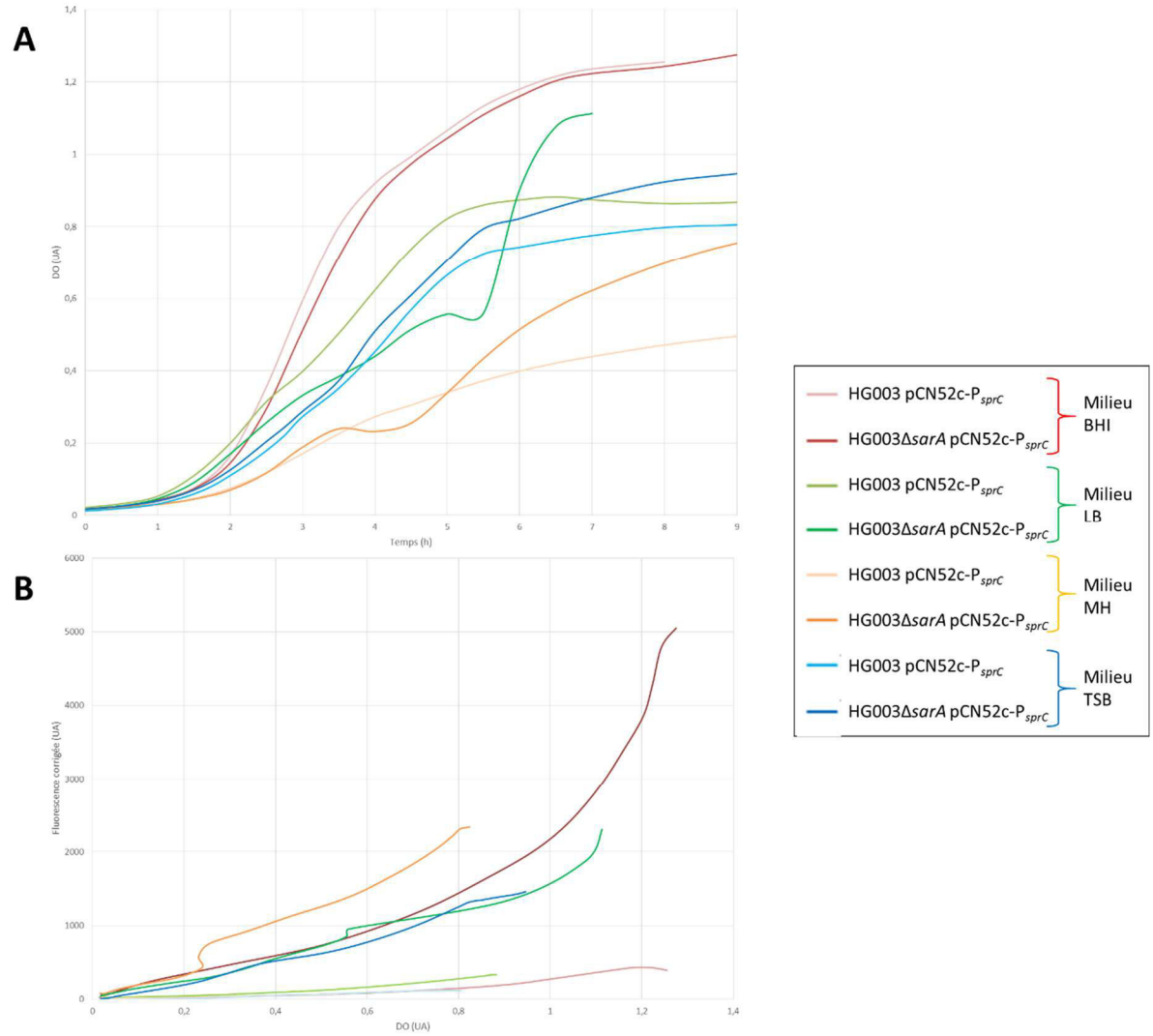


Figure 80 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* dans différents milieux de culture. (A) Croissance des souches dans les différents milieux en fonction du temps. **(B)** Fluoresence émise par les différentes souches selon le milieu utilisé en fonction de la densité de la population bactérienne.

Ce résultat montre que la répression de l'expression de *srn_3610_sprC* par SarA est indépendante des milieux de culture dans lesquels les souches ont été cultivées.

D. Expression de P_{sprC} à différentes températures

Lors d'une infection bactérienne à *S. aureus*, cette dernière va devoir survivre puis se développer dans différentes conditions de températures, c'est pourquoi les souches d'intérêt ont été cultivées à différentes températures. Après une préculture de nuit à 37°C, nous avons choisi d'étudier l'expression de P_{sprC} à 20°C (température ambiante), 37°C (température intracorporelle) et 42°C (température en cas de fièvre). Les souches ont été mises en pré-culture dans le milieu BHI à 37°C, puis les densités bactériennes sont égalisées et les souches mises en culture à 20°C, 37°C ou 42°C (Figure 81).

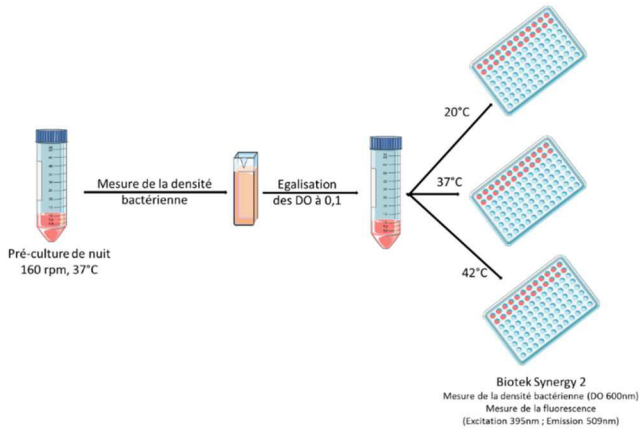


Figure 81 : Protocole d'expérience de gène rapporteur *gfp* à différentes températures de culture.

A 20°C et 37°C, la croissance des souches HG003 est légèrement meilleure que celles des souches HG003Δ*sarA* ; à l'inverse, à 42°C, la souche HG003Δ*sarA* a une vitesse de croissance légèrement supérieure à la souche HG003. Cette observation pourrait indiquer que l'absence de SarA serait favorable à une croissance à 42°C alors qu'à des températures inférieures, son absence serait défavorable pour la bactérie (Figure 82).

La croissance des souches à 37°C est optimale avec une DO allant jusqu'à 1,3 atteinte à 7 heures de croissance. A 42°C, la croissance bactérienne est impactée, elle est à la fois ralentie et la biomasse finale est inférieure (par rapport à ce qui est observé à 37°C). En effet, la DO maximale est autour 0,7. Enfin, à 20°C, la croissance bactérienne est nettement ralentie, la phase stationnaire n'ayant pas été atteinte au bout de 10 heures de cultures (Figure 82A).

La température est donc un élément essentiel pour la croissance bactérienne. *S. aureus* est une bactérie mésophile avec une température optimale à 37°C (température intracorporelle de ses hôtes). Elle peut néanmoins survivre à une gamme importante de température 7-48°C et nos résultats sont en accord avec cette observation. Néanmoins, la biomasse finale est impactée pour les souches mises en culture à 42°C. Nous pourrions l'expliquer par l'accumulation dans le cytoplasme de protéines mal repliées ce qui entraînerait une toxicité pour la bactérie.

La Figure 82B présente les mesures de la fluorescence des souches (corrigées par rapport aux souches porteuses du plasmide vide) en fonction de la DO. L'avantage de cette représentation est de prendre en compte le nombre total de bactérie pour une valeur de fluorescence donnée.

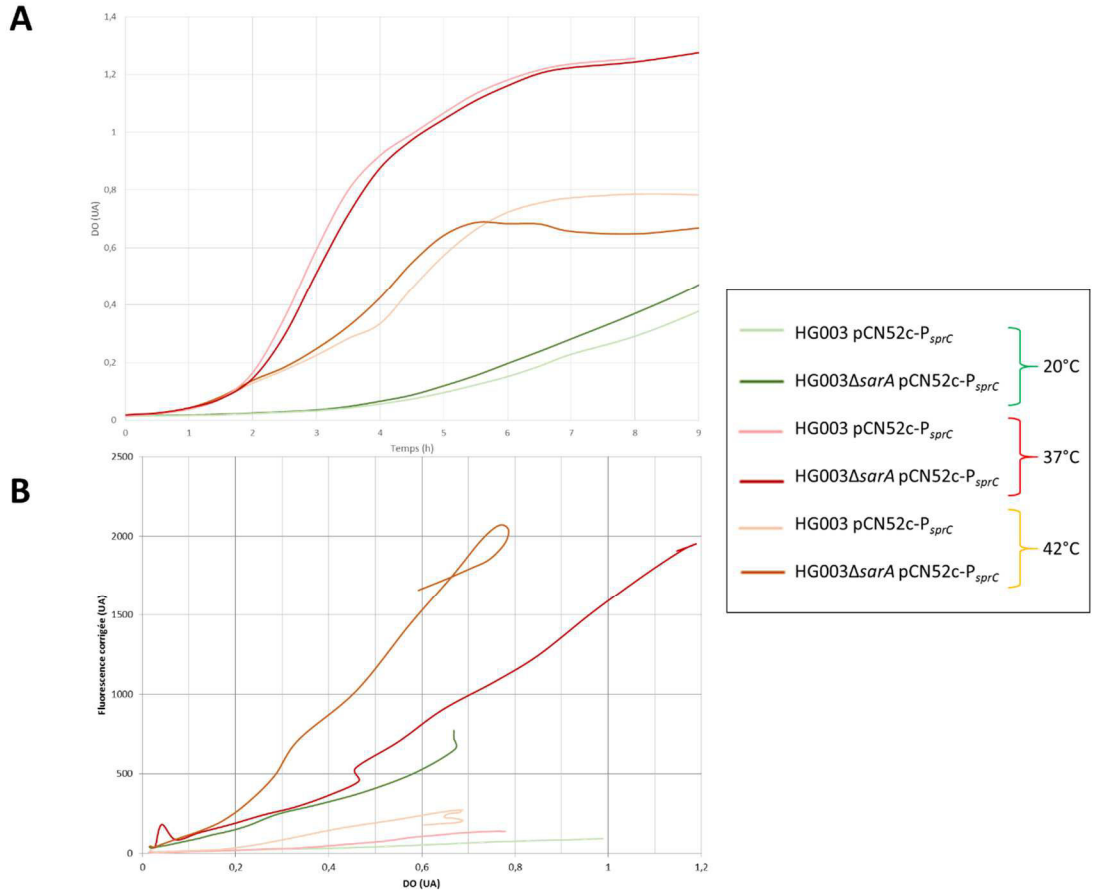


Figure 82 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* à différentes températures de croissance (20°C, 37°C et 42°C). (A) Croissance des souches aux différentes températures en fonction du temps. **(B)** Fluoresence émise par les différentes souches selon la température choisie en fonction de la densité de la population bactérienne.

Globalement, le différentiel de niveau de fluorescence entre les souches HG003 et HG003Δ*sarA* est identique quelle que soit la température étudiée. Néanmoins, à la température de 42°C, le niveau global de fluorescence est plus important qu'à 20°C et 37°C (qui ont des niveaux de fluorescence comparables). Cette observation est probablement due à la biomasse finale, laquelle est impactée par les hautes températures. Par exemple, si nous comparons les niveaux de fluorescence à la DO 0.4, ils sont d'environ 900 pour la croissance à 42°C et de 450 à 37°C. Or, les bactéries ne se situent pas au même stade de leur croissance. A DO=0,4, à 42°C nous sommes à une DO correspondant à la moitié de la DO obtenue en phase stationnaire alors qu'à 37°C, nous nous situons vraiment au début de la phase exponentielle. Pour être dans la même phase de croissance à 37°C qu'à 42°C il faudrait

comparer les niveaux de fluorescence pour la DO=0,65 (moitié DO de la phase stationnaire à 37°C). Dans ce cas, la fluorescence est d'environ 900. Ces observations montrent qu'il faut être vigilant dans l'analyse des résultats de fluorescence/DO.

Même si le graphique représentant la fluorescence par rapport à la DO est une bonne alternative pour comparer deux souches ayant la même biomasse finale, il pose problème si les biomasses finales sont différentes, les phases de croissance ne sont donc pas prises en considération avec cette représentation graphique. Néanmoins, nous arrivons tout de même à conclure grâce à la comparaison des valeurs obtenues pour la souche HG003 par rapport à celles obtenues pour la souche HG003Δ*sarA*.

Il ne faut pas oublier que nous souhaitons observer une variation importante de la fluorescence dans l'une des deux souches. Or, l'activité transcriptionnelle de *srn_3610_sprC* n'est pas impactée par la température. La température n'est ni le signal de levée de répression attendu ni une condition impactant la transcription de *srn_3610_sprC* dans la souche HG003Δ*sarA.*

E. Expression de P_{sprC} en condition de stress osmotique

La mesure et l'adaptation à l'osmolarité environnante par la bactérie sont nécessaires à sa survie. Lors de l'infection, le système immunitaire de l'hôte répond à la présence de bactéries et recrute sur le site d'infection des cellules phagocytaires. Ces cellules ont pour but de phagocyter les bactéries, c'est-à-dire de les internaliser pour les dégrader. Il a été montré que parfois l'internalisation n'était pas suffisante pour la destruction de la bactérie. L'osmolarité dans une cellule phagocytaire est très différente de l'osmolarité du milieu extérieur. C'est pourquoi, une fois dans la cellule hôte, *S. aureus* peut être sujet à un potentiel osmotique différent au sien. Nous avons tenté de produire un stress osmotique par ajout de NaCl dans le milieu de culture. Les souches HG003 et HG003Δ*sarA* ont été mises en culture dans le milieu BHI pendant 3 heures, puis nous avons ajouté du NaCl à une concentration finale de 1,5M ou de l'eau stérile comme contrôle (Figure 83).

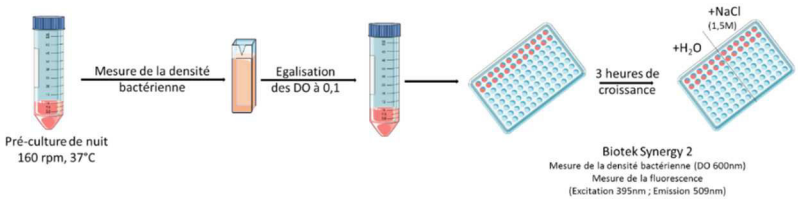


Figure 83 : Protocole d'expérience de gène rapporteur *gfp* en condition de stress osmotique

La Figure 84A présente la croissance bactérienne des souches HG003 et HG003*ΔsarA* dans le milieu BHI +/- NaCl. Dans les deux conditions, nous retrouvons un léger ralentissement

de la croissance pour les souches HG003Δ*sarA* par rapport aux souches HG003. SarA semble donc nécessaire pour obtenir une croissance optimale quelle que soit les conditions étudiées. L'ajout de NaCl impacte la croissance de la bactérie d'une manière globale avec une biomasse finale à environ 0,6 contre une DO finale à 0,9 pour les cultures ayant reçu l'eau. Comme attendu, l'augmentation du potentiel osmotique du milieu (par ajout de NaCl 1,5M finale) est délétère à la croissance de *S. aureus*.

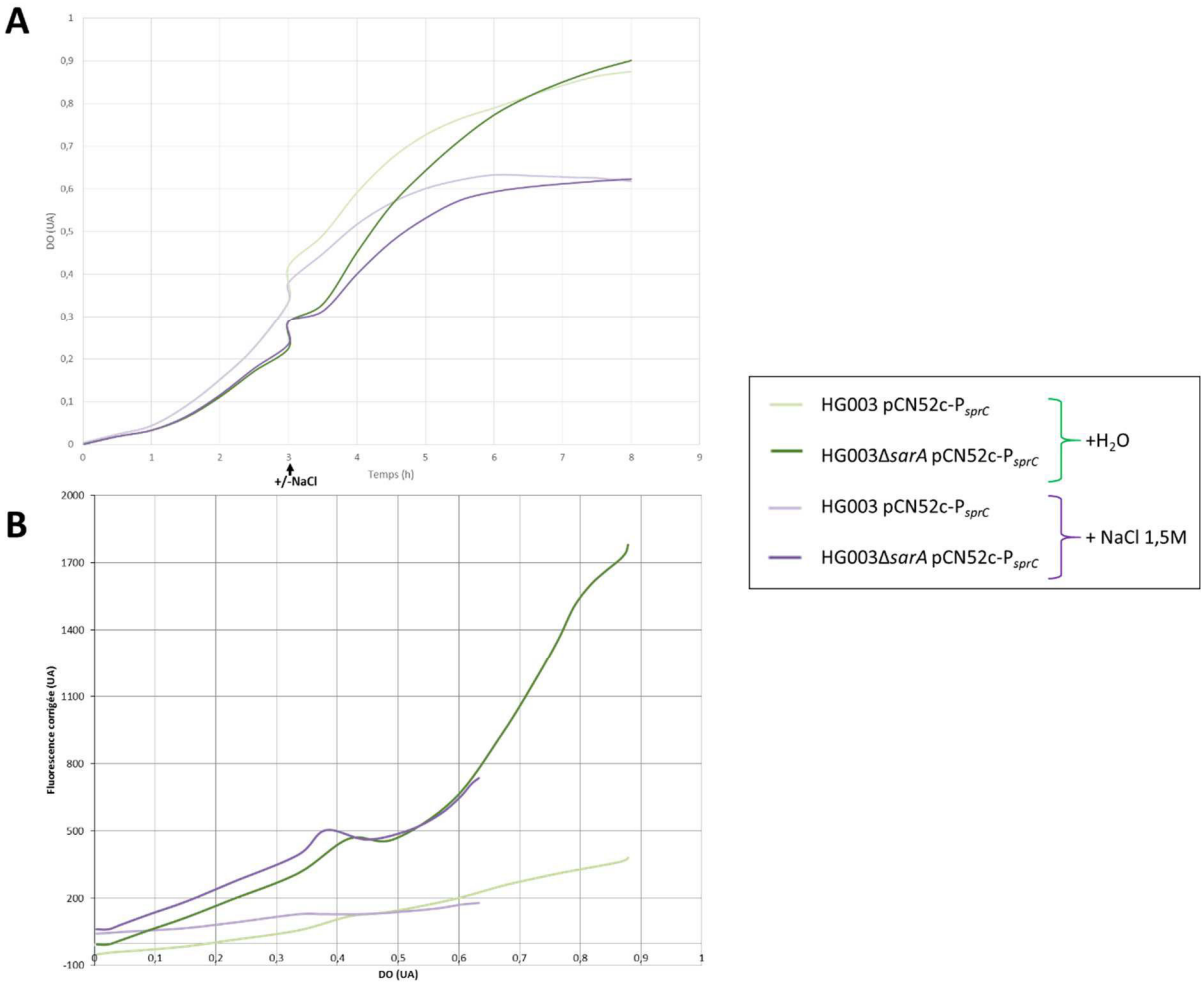


Figure 84 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* en condition de stress osmotique (+ NaCl 1,5M). (A) Croissance des souches sous stress osmotique en fonction du temps. **(B)** Fluoresence émise par les différentes souches après ajout d'eau stérile ou de NaCl 1,5M en fonction de la densité de la population bactérienne.

L'activité du promoteur P$_{sprC}$ est légèrement augmentée dans les souches HG003 et HG003Δ*sarA* lorsque du NaCl est ajouté à 3 heures de croissance. Néanmoins, comme pour l'étude de l'impact des températures, le fait que la biomasse maximale soit différente entre les cultures sans stress et avec stress, implique un biais dans la technique. Il en résulte ce léger différentiel entre les deux conditions de culture. De manière générale, nous pouvons dire que les niveaux de fluorescence avec ou sans ajout de NaCl se superposent. Le stress osmotique

n'est donc pas le signal permettant le décrochage de SarA sur les promoteurs de *srn_3610_sprC* et *srn_9340*. En absence de modification du niveau de fluorescence dans la souche HG003Δ*sarA* en condition de stress osmotique, nous pouvons conclure que la transcription de *sprC* en absence de SarA n'est pas sensible au stress osmotique.

F. Expression de P_{sprC} dans des conditions favorisant le biofilm

Afin de se préserver des agressions extérieures (antibiotiques, cellules immunitaires, etc.), *S. aureus* a la possibilité de s'agréger pour former un biofilm. Au laboratoire, nous pouvons mettre en culture des bactéries dans des conditions mimant les premières étapes de la formation du biofilm. Pour cela, nous plaçons les bactéries dans un milieu dont la composition est faible en nutriment et réalisons une croissance sans appliquer d'agitation.

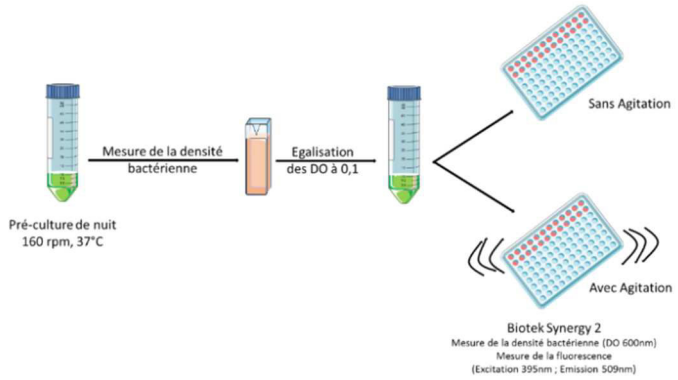


Figure 85 : Protocole d'expérience de gène rapporteur *gfp* en condition favorisant la création de biofilm.

Nous avons voulu étudier l'activité du promoteur P_{sprC} lors d'une croissance bactérienne dans des conditions mimant la première étape de la formation du biofilm bactérien. Le milieu utilisé pour cette expérience est le milieu LB. Au matin, les densités bactériennes des cultures de HG003 et HG003Δ*sarA* sont égalisées à 0,1. La fluorescence et la densité bactérienne sont analysées par un Biotek dont la plaque est mise sous agitation (contrôle) ou sans agitation.

L'absence d'agitation pour les cultures entraîne une diminution générale de la croissance bactérienne des souches HG003 et HG003Δ*sarA* (phase stationnaire atteinte vers DO 0,4) pouvant être expliquée par un manque d'aération du milieu (Figure 86A). Cependant, les niveaux de fluorescence entre les deux conditions de culture sont, pour chaque souche, superposables (Figure 86B).

L'application ou non de l'agitation n'entraîne donc pas de variation du niveau de fluorescence aussi bien pour les souches HG003 que pour les souches HG003Δ*sarA*. L'absence d'agitation dans un milieu intermédiaire ne permet pas de modifier l'activité du promoteur P_{sprC}. Ces conditions de cultures ne miment donc pas un signal de dérépression de SarA sur l'expression de *srn_3610_sprC*.

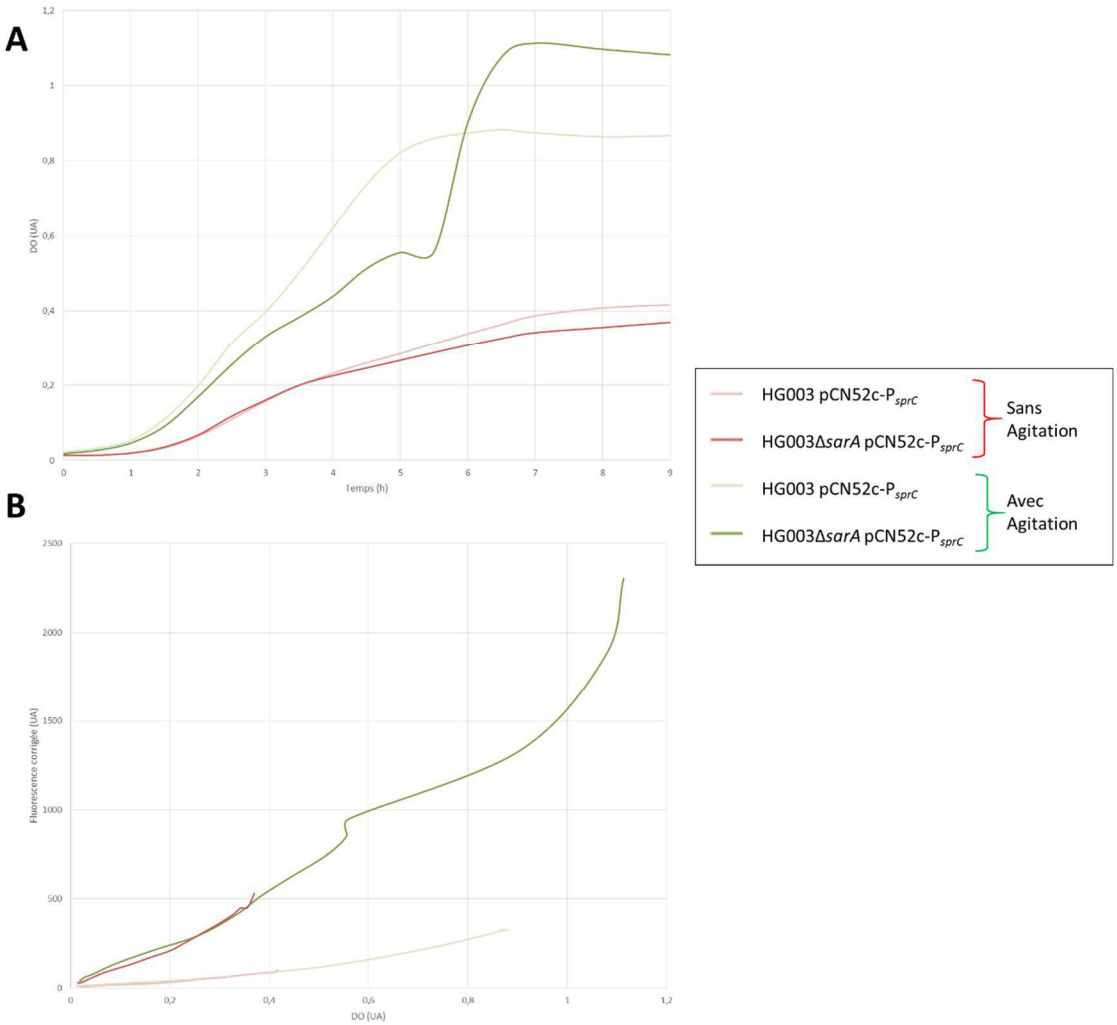


Figure 86 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* dans les conditions favorisant le biofilm. (A) Croissance des souches avec ou sans agitation en fonction du temps. **(B)** Fluoresence émise par les différentes souches avec ou sans agitation en fonction de la densité de la population bactérienne.

G. Expression de P$_{sprC}$ lors d'un stress d'autolyse

Le Pabic *et al.* ont montré que, lors des premières minutes de la phagocytose, Srn_3610_SprC est fortement produit. Néanmoins, l'expression de son gène est très vite réprimée pendant la phagocytose. Nous savons également que Srn_3610_SprC cible l'ARNm de l'autolysine, protéine impliquée entre autres dans le contact cellule hôte/pathogène (Le Pabic et al., 2015). Nous avons alors voulu tester l'expression de *srn_3610_sprC* dans des conditions d'autolyse.

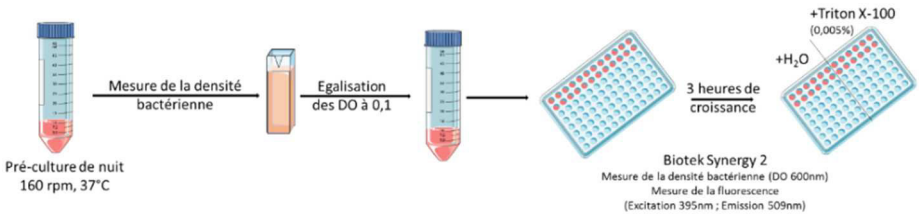


Figure 87 : Protocole d'expérience de gène rapporteur *gfp* lors d'un stress d'autolyse

L'induction du stress d'autolyse est réalisée après 3 heures de croissance en milieu BHI par l'ajout du détergent (Triton X-100) à un pourcentage final de 0,005% (Figure 87). Le Triton X-100 a pour effet de perméabiliser les membranes et donc de fragiliser l'intégrité de la bactérie.

La concentration de Triton X-100 utilisée dans l'expérience a été déterminée comme étant la première concentration sub-inhibitrice de la croissance. À la vue des courbes de croissance présentées Figure 88, nous constatons, comme attendu, que le détergent, à cette concentration, n'a pas d'effet significatif sur la croissance de souches HG003 et HG003Δ*sarA*. Par ailleurs, nous constatons, comme précédemment un léger défaut de vitesse de croissance de la souche HG003Δ*sarA* par rapport à la souche HG003.

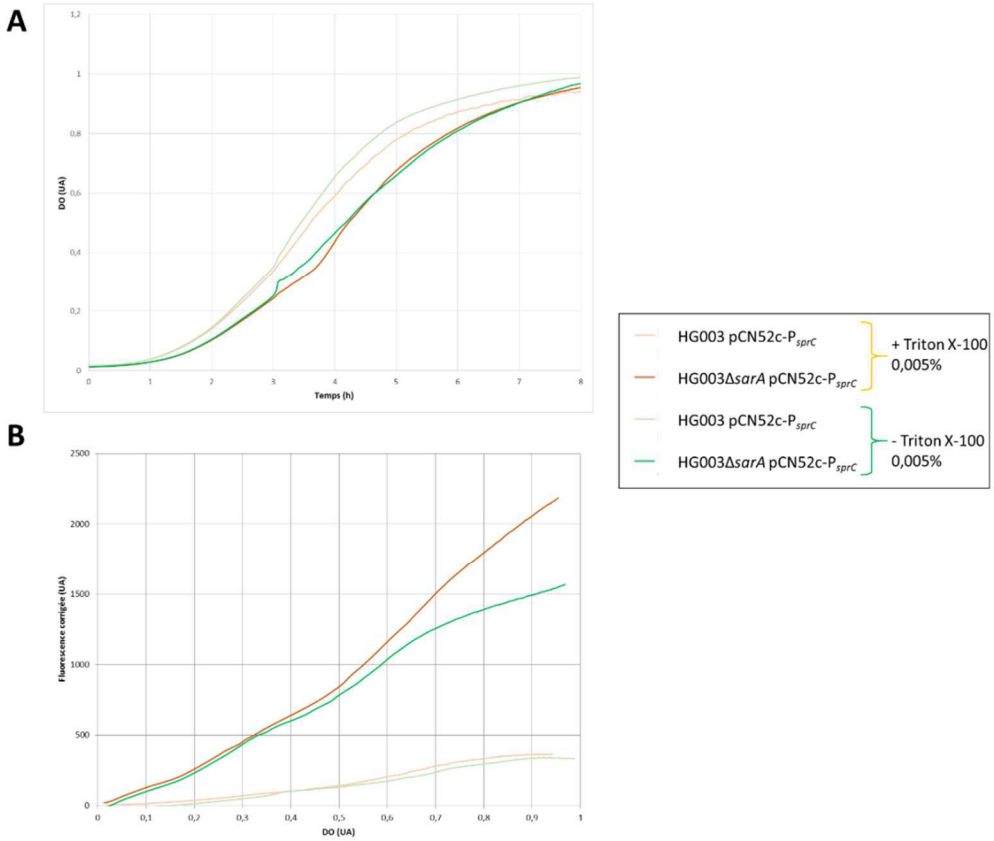


Figure 88 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. (A) Croissance des souches avec ou sans ajout de Triton X-100 0,005% en fonction du temps. **(B)** Fluoresence émise par les différentes souches sous stress d'autolyse en fonction de la densité de la population bactérienne.

Pour la souche HG003, nous n'observons pas de différence de fluorescence avec ou sans stress. Par contre, dans la souche HG003Δ*sarA*, l'ajout de Triton X-100 entraîne une diminution importante du niveau de fluorescence par rapport à la même souche cultivée en absence de stress. L'activité du promoteur P_{sprC} dans la souche HG003Δ*sarA* est donc impactée lors de l'ajout de Triton X-100. Pour vérifier cette observation, nous avons décidé d'étudier l'expression du promoteur dans de nouvelles expériences de gène rapporteur. Nous avons utilisé le deuxième gène rapporteur à disposition : la ß-lactamase. La mesure de l'activité de la ß-lactamase dans trois expériences indépendantes permettra l'analyse statistique.

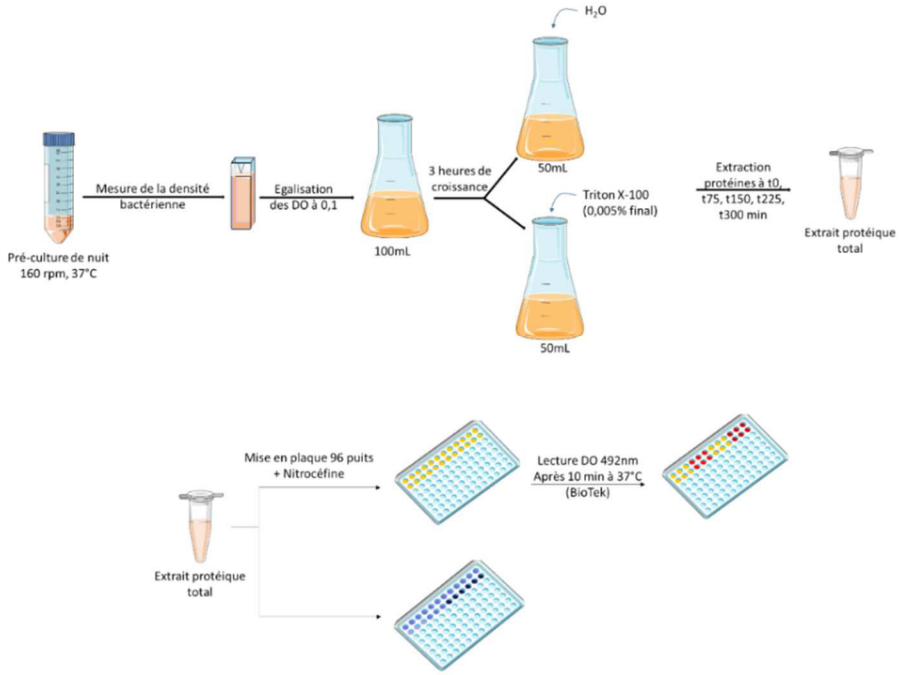


Figure 89 : Protocole d'expérience de gène rapporteur *blaZ* (ß-lactamase) lors d'un stress d'autolyse

Les cultures bactériennes de HG003 et HG003Δ*sarA* sont préparées à DO 0,1 puis misent sous agitation à 37°C pendant 3 heures. Les cultures sont ensuite divisées dans deux erlenmeyers. Dans le premier est ajouté du Triton X-100 à la concentration de 0,005%, dans le second le même volume d'eau stérile que celui de Triton X-100 utilisé pour réaliser le stress d'autolyse. Le dosage de la quantité de ß-lactamase produite ainsi que la quantité de protéines totales ont été réalisés pour déterminer l'activité du promoteur de *srn_3610_sprC* (Figure 89).

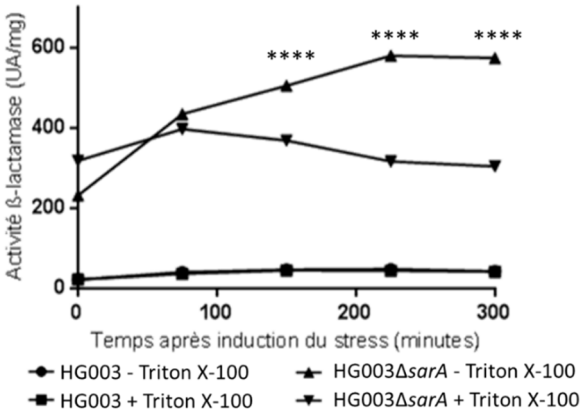


Figure 90 : Activité du promoteur *srn_3610_sprC* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. Les extraits protéiques totaux ont été extraits à 0, 75, 150, 225 et 300 minutes après application ou non du Triton X-100. L'activité de la ß-lactamase a été mesurée pour chacun des extraits. Les moyennes des trois expériences ont été représentées (test de Tukey : **** : p-value<0,001%).

Les activités ß-lactamase mesurées dans la souche HG003 ayant reçu ou non du Triton X-100, ne sont pas significativement différentes et ce résultat est valable pour tous les temps de prélèvement après l'induction du stress (les courbes se superposent au point de n'en voir qu'une sur le graphe Figure 90). Par contre, on observe que l'ajout de Triton X-100 entraîne une diminution significative de l'activité enzymatique de la ß-lactamase dans la HG003Δ*sarA* et ce à partir de 150 minutes après son application (test de Tukey : p-value<0,001%). Les valeurs d'activité ß-lactamase dans la HG003Δ*sarA* ayant subi le stress d'autolyse sont toujours très supérieures à celles observées dans la souche HG003 mais la différence d'expression n'est plus de 10 fois mais de X fois. Les expériences de gène rapporteur *blaZ* ont permis de confirmer que l'activité du promoteur P_{sprC} est significativement diminuée dans la souche HG003Δ*sarA* après un stress d'autolyse (Triton X-100).

En conclusion, les données obtenues avec les expériences de gènes rapporteurs *gfp* et ß-lactamase sont concordantes et montrent un impact du Triton X-100 sur l'activité du promoteur P_{sprC} dans la souche HG003Δ*sarA*. Il semble donc que l'agression de la membrane par le détergent entraîne une diminution de l'activité du promoteur P_{sprC}, en absence de SarA. Nous pouvons supposer que l'excès de Srn_3610_SprC est délétère pour la bactérie en cas d'environnement perméabilisant sa paroi et qu'une répression ne passant pas par SarA doit avoir lieu. Une autre hypothèse serait, qu'en présence de Triton X-100, un facteur de transcription autre que SarA ne soit pas recruté sur le promoteur au détriment de l'expression de *srn_3610_sprC.*

Au vu des similarités de la répression de SarA sur *srn_3610_sprC* et *srn_9340*, nous avons voulu étudier l'impact de l'ajout de Triton X-100 sur l'activité du promoteur de *srn_9340.*

H. Expression de P_{srn_9340} lors d'un stress d'autolyse

Nous venons d'identifier un stress impactant l'activité du promoteur de *srn_3610_sprC* dans des conditions où SarA n'est pas fixée sur ce promoteur. Nous avons voulu savoir si nous retrouvons ce phénomène pour le promoteur de *srn_9340*. Nous avons donc utilisé les souches HG003 ou HG003Δ*sarA* portant le plasmide pCN52c-P_{srn_9340} auxquelles nous avons appliqué ou non le détergent Triton X-100 à une concentration sub-inhibitrice de 0,005%. L'expression du gène *gfp* porté par le plasmide pCN52c a été placée sous le contrôle du promoteur de 144 pb de *srn_9340*. La fluorescence émise par la bactérie sera donc le reflet de l'activité du promoteur P_{srn_9340}.

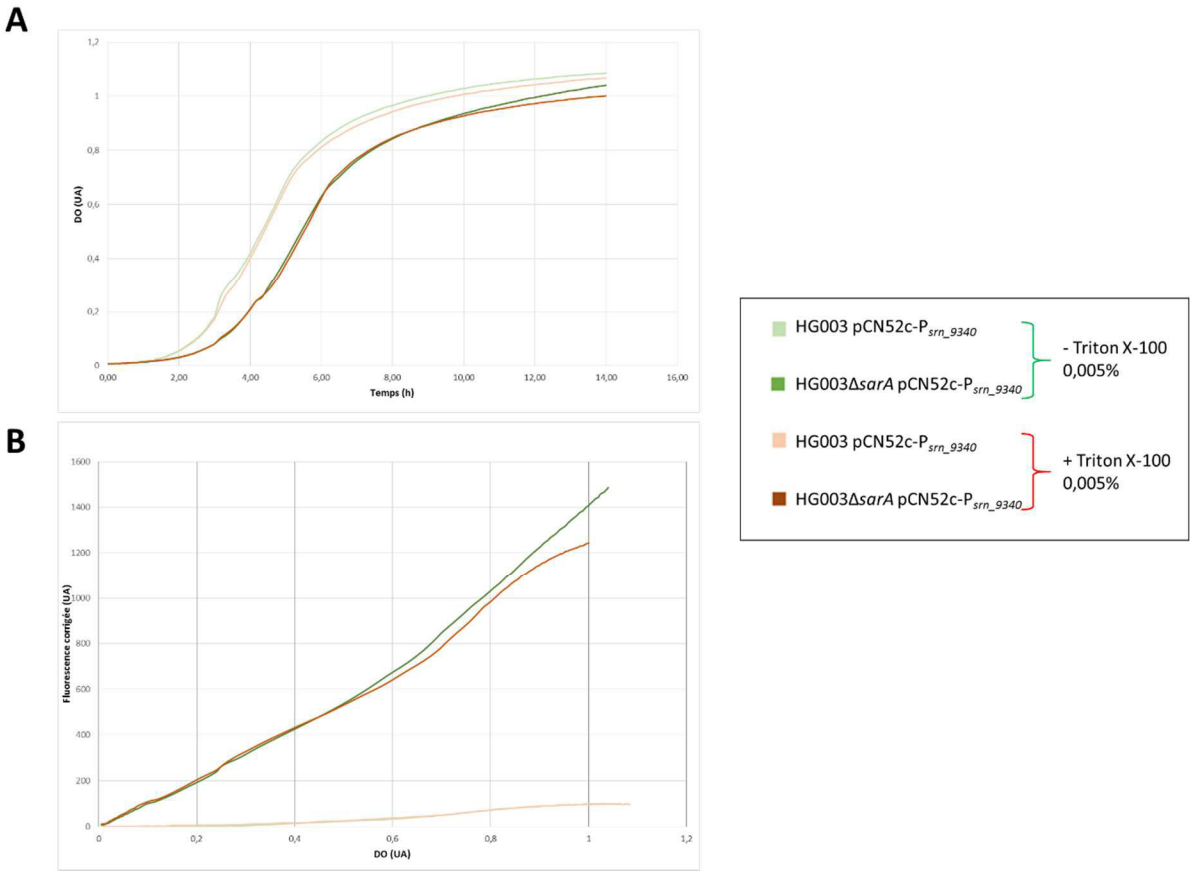


Figure 91 : Activité du promoteur *srn_9340* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. (A) Croissance des souches avec ou sans ajout de Triton X-100 0,005% en fonction du temps. **(B)** Fluoresence émise par les différentes souches sous stress d'autolyse en fonction de la densité de la population bactérienne.

Comme précédemment observé, la vitesse de croissance des souches HG003 est meilleure que celle des souches HG003Δ*sarA*. Nous voyons aussi que l'application du stress d'autolyse (Triton X-100 à 0,005%) ne modifie pas la croissance des bactéries.

La comparaison des niveaux de fluorescence des souches HG003 stressées ou non montre que l'ajout de Triton X-100 n'influence pas l'activité du promoteur de *srn_9340*. En

revanche, pour les souches HG003Δ*sarA*, la présence du détergent entraîne une diminution de la fluorescence émise. Nous avons effectué des tests de statistiques sur trois expériences indépendantes identiques à celle présentée ci-dessus. Pour pouvoir comparer les différentes souches, nous avons choisi de regarder le niveau à quatre points de croissance (à DO 0,3 ; 0,5 ; 0,7 et 1 UA). Nous constatons que la délétion de *sarA* entraîne une augmentation significative des niveaux de fluorescence, témoin de l'augmentation de l'activité du promoteur de *srn_9340* dans la souche HG003Δ*sarA.* Nous constatons que l'ajout de Triton X-100 entraîne une diminution de l'activité du promoteur de *srn_9340*. Les tests statistiques nous indiquent que cette diminution est significative uniquement à une DO=1 (p-value = 0,05 en test unilatéral Mann-Withney). Ainsi, la présence du détergent impacterait moins l'expression de *srn_9340* que celle de *srn_3610_sprC*.

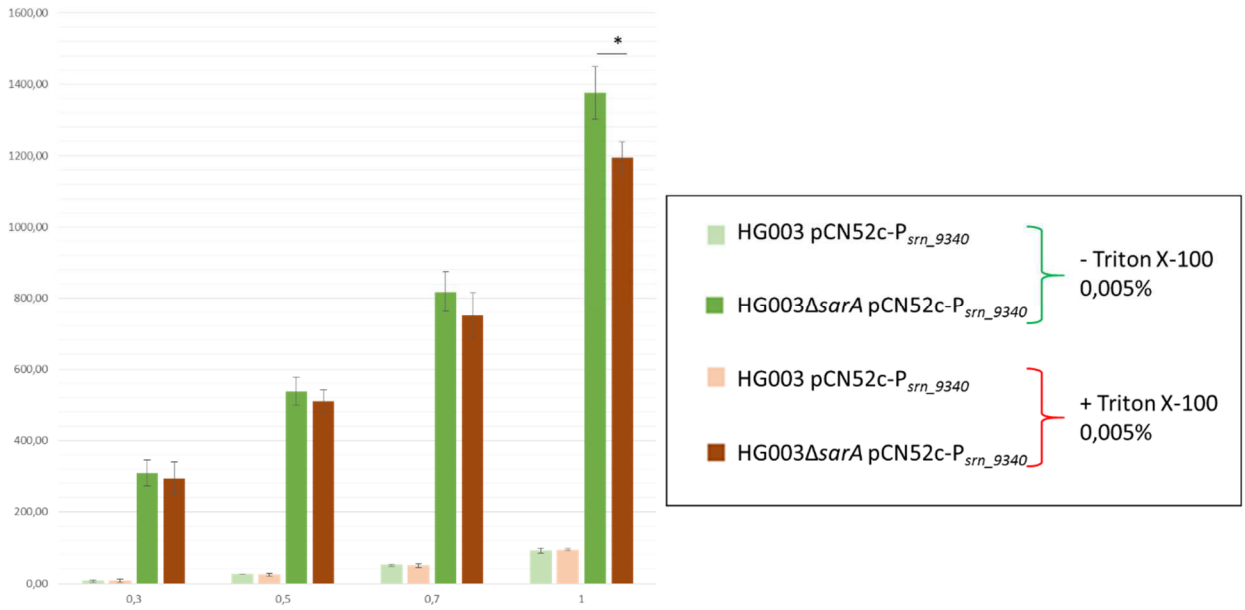


Figure 92 : Activité du promoteur *srn_9340* dans les souches HG003 et HG003Δ*sarA* lors d'un stress d'autolyse. La moyenne, de trois expériences indépendantes, de mesure de fluorescence à DO_{600nm} 0,3 ; 0,5 ; 0,7 et 1 des différentes souches a été représentée (test de Mann-Whitney : * : p-value<0,05%).

En conclusion, parmi les différents stress testés, le détergent Triton X-100 est le seul à avoir modifié l'activité du promoteur de *srn_3610_sprC* et *srn_9340*. Ce stress dit d'autolyse permet de fragiliser la membrane bactérienne. La diminution de l'activité de ces promoteurs est indépendante de SarA et pourrait être due à l'action d'un nouveau répresseur lequel en condition de stress d'autolyse agirait négativement sur l'expression de *srn_3610_sprC* et *srn_9340.* Il est également possible, qu'en condition de stress d'autolyse un activateur transcriptionnel se décroche des promoteurs de *srn_3610_sprC* et *srn_9340*. Dans tous les cas, il semble qu'en cas de fragilisation de la membrane bactérienne *S. aureus* a besoin de réprimer l'expression de *srn_3610_sprC* et *srn_9340*.

Nous avons vu ici la difficulté de trouver un signal permettant la dérépression de l'expression de *srn_3610_sprC* et *srn_9340.* Nous avons donc décidé de rechercher le signal de dérépression de SarA indirectement. Nous avons voulu connaître les cibles protéiques dont la production serait impactée positivement ou négativement par une expression constitutive forte de *srn_3610_sprC* ou *srn_9340*. L'identification des cibles protéiques pouvant nous donner des informations sur le rôle de Srn_3610_SprC et Srn_9340 dans la cellule.

II. Détermination de cibles

Pour déterminer les cibles protéiques en fonction de la surexpression de *srn_3610_sprC* et *srn_9340*, nous avons modifié l'expression de ces gènes en remplaçant leur promoteur par un promoteur constitutif.

A. Construction de vecteur de surexpression de *srn_3610_sprC* et *srn_9340*

Le problème majeur de l'expression des promoteurs de *srn_3610_sprC* et *srn_9340* est qu'ils sont très fortement réprimés par SarA. Nous aurions pu étudier les cibles de Srn_3610_SprC et Srn_9340 dans une souche HG003Δ*sarA*, où l'expression de leur gène est très forte. Mais comme SarA est un régulateur majeur chez *S. aureus*, nous ne pourrions pas distinguer les cibles spécifiques de nos ARN régulateurs de celles de SarA. Aussi, nous avons choisi de remplacer le promoteur endogène de *srn_3610_sprC* et *srn_9340* par un promoteur dont l'activité ne serait pas réprimée par SarA. Pour choisir le promoteur, nous nous sommes basés sur l'étude de l'expression de quatre promoteurs fortement actifs chez différentes bactéries (Biswas et al., 2008). Nous avons choisi le promoteur P_{veg} issu de *B. subtilis* (Figure 93). Deux raisons nous ont permis de faire ce choix : le promoteur P_{veg} vient d'une bactérie à Gram + proche de *S. aureus* et la séquence de ce promoteur est riche en guanine et cytosine. En choisissant un promoteur riche en GC, nous diminuons la probabilité que SarA se fixe sur ce promoteur et intervienne dans l'expression du gène placé sous son contrôle.

P_{veg} : **TTGACA**AAAATGGGCTCGTGTTG**TACAAT**AAATGT

Figure 93 : Séquence du promoteur P_{veg}

Pour pouvoir exprimer fortement *srn_3610_sprC* et *srn_9340*, nous avons choisi d'insérer ces gènes sur des plasmides. Ainsi, le gène et le terminateur de *srn_3610_sprC* ou *srn_9340* ont été amplifiés par PCR et respectivement intégré, après digestion et ligation, dans le plasmide pCN38 ou pCN36 (Tableau 12). Pour positionner le promoteur P_{veg} en amont du gène, nous avons opté pour la technique de Gibson. Cette technologie a l'avantage d'éviter d'utiliser des sites de restriction qui se situerait entre le promoteur et le gène. En effet, la

présence d'un site de restriction rajoute 6 nucléotides supplémentaires à l'extrémité 5' de l'ARN étudié.

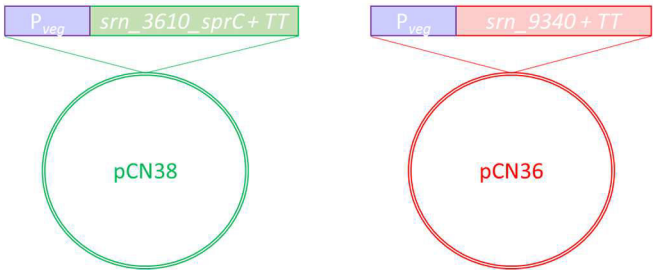


Figure 94 : Représentation des plasmides utilisés pour l'expérience de surexpression de *srn_3610_sprC* et *srn_9340*.

Au final, nous obtenons deux plasmides : pCN38-P$_{veg}$SprC et pCN36-P$_{veg}$Srn_9340 (Figure 94). L'expression de ces promoteurs doit donc produire, dans la souche sauvage HG003, de manière constitutive et forte les trois transcrits (celui de *srn_3610_sprC* et les deux de *srn_9340*). La souche HG003 a donc été transformée avec l'un ou l'autre des plasmides et les ARN totaux ont été extraits à 5 heures de croissance. Les Northern Blot réalisés (Figure 95) permettent de vérifier la quantité d'ARN Srn_3610_SprC ou Srn_9340 présent dans les bactéries portant les plasmides.

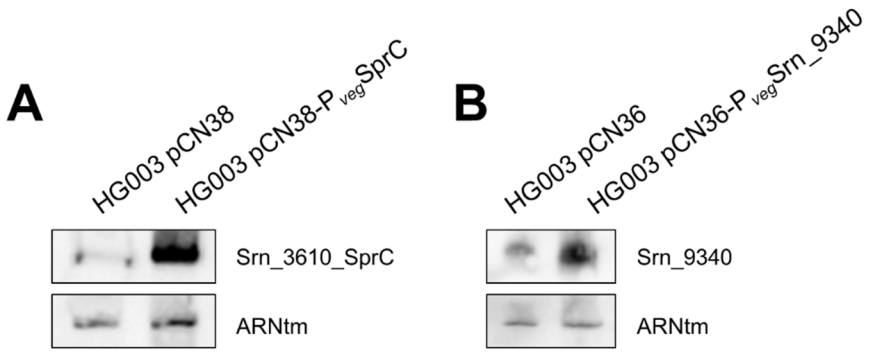


Figure 95 : Validation de la surexpression de *srn_3610_sprC* (A) et *srn_9340* (B) par l'activité du promoteur P$_{veg}$.

La comparaison des quantités d'ARN Srn_3610_SprC présentes dans les souches HG003 pCN38 (plasmide vide) et HG003 pCN38-P$_{veg}$SprC (Figure 95A) montre une augmentation dans la souche HG003 pCN38-P$_{veg}$SprC par rapport à la souche HG003 pCN38. Cette expérience indique que le promoteur P$_{veg}$, issu de *B. subtilis*, est actif chez *S. aureus* et que celui-ci permet une forte production de Srn_3610_SprC. L'outil est donc validé pour réaliser une surexpression de *srn_3610_sprC.*

En ce qui concerne Srn_9340, la Figure 95B présente la quantité d'ARN dans les souches HG003 pCN36 (plasmide vide) et HG003 pCN36-P$_{veg}$Srn_9340. Nous observons une augmentation de la quantité d'ARN Srn_9340 dans la souche portant le plasmide avec pCN36-P$_{veg}$Srn_9340 par rapport à la souche ayant le plasmide vide pCN36. Comme *srn_3610_sprC*,

l'expression de *srn_9340* est fortement augmentée lorsque son gène est placé en amont du promoteur P_{veg}.

Nous notons aussi que dans les deux cas, la taille des transcrits issu du plasmide (puits de droite, Figure 95 A et B) est comparable à la taille des transcrits endogènes (puits de gauche : HG003 pCN38 ou pCN36). La transcription obtenue par la fixation de l'ARN polymérase sur le promoteur $P_{veg,}$ permet de produire fortement des transcrits identiques aux transcrits endogènes.

Malheureusement, je n'ai pas pu, dans le cadre de ma thèse, poursuivre cette étude. Cependant, ces outils serviront à étudier l'impact de la surexpression de *srn_3610_sprC* et/ou *srn_9340* sur le protéome de *S. aureus*.

En conclusion, nous n'avons pu pas identifier de condition permettant le décrochage de SarA de l'ADN (Figure 96). Par contre, la forte transcription de *srn_3610_sprC* et *srn_9340* est possible (1) en absence de SarA et (2) lors de la phagocytose. La phagocytose pourrait être un signal pour une levée de répression transitoire de SarA. Au moment de la phagocytose un signal de l'hôte est probablement perçu par la bactérie et conduit à la dérépression de *srn_3610_sprC* et *srn_9340* sans doute via le décrochage de SarA.

Nous proposons d'ajouter au modèle précédemment présenté (Figure 73) les caractéristiques suivantes :

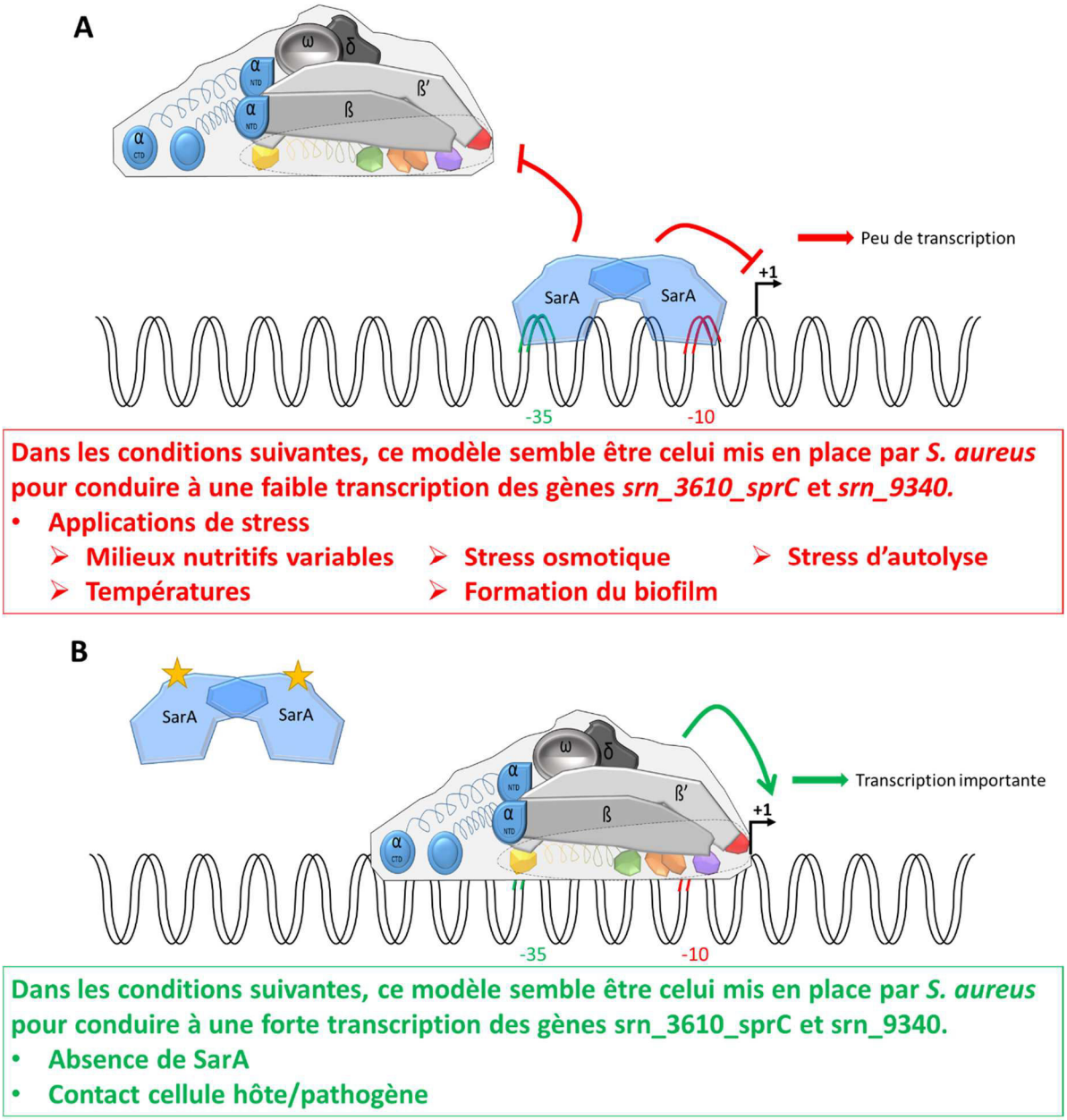


Figure 96 : Schéma général des conditions de régulation de *srn_3610_sprC* et *srn_9340* selon (A) que l'expression de ces gènes est réprimée ou (B) qu'elle est activée.

PARTIE III :

Identification de gènes régulés par SarA

RESULTATS ET DISCUSSIONS

La caractérisation des séquences de fixation de SarA sur les promoteurs de *srn_3610_sprC* et *srn_9340* a fourni de nouvelles informations sur la fixation de SarA sur l'ADN. Dans cette partie, nous allons d'abord déterminer des séquences consensus de la fixation de SarA pour ensuite proposer un modèle de répression ou d'activation générale de ce facteur de transcription.

I. Recherche de séquences consensus de la fixation de SarA à partir de cibles identifiées

Depuis sa découverte en 1992, le facteur de transcription SarA a fait l'objet de très nombreuses recherches. Il a ainsi été décrit comme un régulateur majeur de l'expression de gènes chez *S. aureus.* Afin de mieux déterminer les cibles directes de SarA, plusieurs études ont essayé de définir une séquence consensus de la fixation de SarA sur l'ADN.

La première étude, où une séquence consensus est présentée, date de 1998 (Chan et al., 1998). Dans cet article, les auteurs ont montré que SarA est un répresseur de la production de protéase. Par alignement de séquences nucléotidiques de régions promotrices de sept gènes codant pour des protéases (toutes réprimées par SarA), ils ont mis en évidence une séquence consensus de 29 nt : 5'-aaAtNTaatTtgtattNaAtattTtaNtN-3' (Consensus1 Chan). Néanmoins, cette séquence est à utiliser avec précaution car la fixation de SarA sur ces promoteurs n'a pas été démontrée. Il n'est donc pas impossible que certaines régulations soient indirectes.

En 1999, suite à l'étude de la fixation de SarA sur plusieurs promoteurs (*agr*, *hla*, *spa*,…), Chien *et al.* proposent une séquence de 26 pb : 5'-ATTTgTATtTAATATTTataTAAtTg-3' (Consensus2 Chien) (Chien et al., 1999). Dans cette étude, des expériences d'empreinte à la DNase I ont été réalisées sur les promoteurs *hla* et *spa* en présence de la protéine SarA. Le consensus2 Chien est retrouvé dans les séquences protégées de la dégradation par la DNase I ce qui valide ce dernier.

Enfin, une dernière étude a porté sur la caractérisation *in vitro*, la fixation de SarA sur l'ADN (Sterba et al., 2003). Pour cela, Sterba *et al.* ont utilisé la technologie SELEX. Cette technique consiste à mettre la protéine SarA en présence d'oligonucléotides de 98 pb contenant dans leur centre des séquences nucléotidiques aléatoires. Les oligonucléotides liés à SarA sont ensuite purifiés et séquencés. Après alignement des séquences obtenues, une séquence de 7 nt est identifiée : 5'-ATTTTAT-3' (Consensus3 Sterba).

Trois séquences consensus ont donc été proposées par différentes équipes de recherche : partagent-elles une séquence commune ?

Pour répondre à cette question, nous avons aligné la séquence de 29 nt identifiée par Chan *et al.,* et celle de 26 nt de Chien *et al.* . Pour cela, nous avons utilisé le logiciel Clustal Oméga en spécifiant le caractère « ADN » des séquences (Figure 97).

```
Consensus1Chan    aaAtaTaatTtgtatttaAtattTtaata----
Consensus2Chien   -------ATTTgTATtTAATATTTataTAAtTg
                          ****************  ***
Consensus4                ATTTGTATTTAATATTTwwATA
```

Figure 97 : Détermination d'une région commune entre les séquences consensus de Chan *et al.* et Chien *et al.* . Les nucléotides adénine, thymine, guanine et cytosine sont représentés par leur initiale ; la lettre w (Weak) correspond à une adénine ou thymine (alignement réalisé par Custal Oméga).

Les deux séquences analysées ont une très grande similarité et le logiciel a réussi à les aligner. En effet, une séquence de 22 nt est fortement conservée (environ 90% d'identité). Ce résultat montre que l'alignement de zones de fixation de SarA (putative ou vérifiée) permet d'obtenir cette séquence consensus : 5'-ATTTGTATTTAATATTTwwATA-3' (consensus4). Cette séquence, très riche en adénine et thymine, est caractérisée par un %GC de 4,5%. Il parait logique d'avoir ce type de séquence car la protéine SarA est connue pour se lier de manière privilégiée sur des séquences riches en adénine et thymine.

```
Consensus4            ATTTGTATTTAATATTTwwATA
Consensus3Sterba      ------ATTTTAT---------
                            **** **
```

Figure 98 : Alignement de la séquence consensus déterminée par Sterba *et al.* et la séquence consensus1. (alignement réalisé par Custal Oméga).

Lorsque nous recherchons les 7 nt de la séquence consensus3 de Sterba *et al.* dans la séquence consensus4, nous constatons que la séquence ATTTTAT se positionne au centre de la séquence consensus4 (Figure 98, alignement de séquence par Clustal Oméga).



De plus, en analysant la séquence consensus4, nous constatons qu'il est possible de retrouver la séquence ATTTTAT (avec un saut dans la séquence) à deux autres endroits de la séquence consensus4.

En conclusion, l'analyse des trois séquences proposées dans la littérature permet de révéler le caractère riche en adénine et thymine de la séquence de fixation de SarA.

L'étude de la régulation directe de gènes par SarA a été réalisée pour d'autres cibles. Néanmoins, force est de constater, que même si la fixation de la protéine SarA sur des promoteurs cibles est confirmée par des expériences de retard sur gel, peu d'études ont

déterminé la séquence fixée par SarA par des expériences d'empreinte à la DNase I. Pour la suite de notre étude, nous nous intéresserons uniquement aux cibles de SarA dont un site de fixation a été réellement identifié.

SarA est capable d'activer ou de réprimer l'expression de nombreux gènes. Pour nous affranchir de cette dualité, nous avons séparé ici les cibles directes réprimées par SarA de celles dont l'expression est activée.

Nous avons répertorié les séquences de fixation de SarA pour cinq gènes cibles réprimés par la protéine (Tableau 4).

Tableau 4 : Sites de protection par SarA de gènes dont l'expression est réprimée.

>Spa AATTATAAATATAGATTTTAG	**Chien *et al.*, 1999**
>SarV ATTTGTAATAGATAAAATGCTAAGTTTGATATATAATAAATTTAAG	**Manna *et al.*, 2004**
>SarA AATTAAAAAAATTATAAAAATTTTTC	**Cheung *et al.*, 2008**
>Srn_3610_SprC GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG	**Mauro *et al.*, 2016**
>Srn_9340 AATATTATATATACGCTTGAAAAAATAAAATAATAATATTATAATATAATC	**Mauro *et al.*, 2016**

- Chien *et al.* ont déterminé par des expériences d'empreinte à la DNase I une région de 144 pb protégée par SarA, cette séquence se situe dans la région promotrice de *spa* (Chien et al., 1999). Les auteurs ont ensuite aligné cette séquence avec leur séquence consensus de la fixation de SarA (consensus2 Chien) et révélé une séquence de 21 nucléotides susceptible de fixer la protéine.
- SarA a été identifiée comme répresseur de l'expression du facteur de transcription *sarV* (Manna et al., 2004). Par des expériences d'empreinte à la DNase I, une séquence de 46 nt protégée par SarA de la dégradation de la DNase I a pu être identifiée.
- SarA réprime l'expression de son propre gène. Le site de fixation de SarA sur son propre promoteur a été caractérisé par Cheung *et al.*(Cheung et al., 2008a). Cette séquence de 26 nucléotides est positionnée au niveau du promoteur P3 de *sarA*. Dans cet article, les auteurs proposent, grâce à des homologies de séquences avec la séquence consensus (consensus2 Chien), deux sites putatifs de fixation de SarA sur le promoteur P1 (données non intégrées dans le tableau).
- Les deux dernières correspondent à la zone de protection de la dégradation de la DNase I par SarA pour les deux ARN régulateurs, *srn_3610_sprC* et *srn_9340*, présentés dans la partie I de ce manuscrit (Mauro et al., 2016).

Un motif nucléotidique commun entre ces différents sites de fixation a été recherché par le logiciel GLAM2 (plateforme MEME : http://meme-suite.org/) (Figure 99).

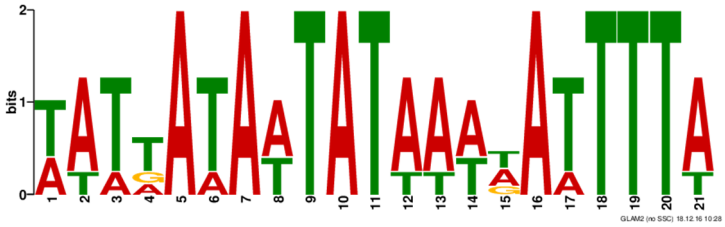


Figure 99 : Motif commun (consensus5) entre les différents sites de fixation de SarA pour des gènes réprimés. L'analyse a été effectuée par le logiciel GLAM2.

Une séquence de 21 nucléotides a été identifiée (5'-wwwdAwAwTATwwwdAwTTTw-3'). Nous constatons que ce motif est majoritairement composé d'adénine et de thymine. Seules deux positions sont susceptibles d'être une guanine (positions 4 et 15). Deux zones sont fortement conservées : la première entre les nucléotides 5 à 11 (AtAaTAT) et la seconde entre les nucléotides 16 à 21 (AtTTTa). De manière intéressante, nous retrouvons, dans cette dernière, les 6 premiers nucléotides de la séquence consensus déterminée par Sterba *et al.* (5'-ATTTTAT-3') (Sterba et al., 2003).

De la même manière que précédemment, nous avons recensé les gènes cibles dont l'expression est, cette fois-ci, activée par SarA et dont une zone de fixation pour cette protéine est avérée. Toutes les zones identifiées et décrites dans le Tableau 5 sont situées dans les régions promotrices des gènes nommés.

Tableau 5 : Sites de protection par SarA de gènes dont l'expression est activée.

>Agr GTAAATTTTTTTATGTTAAAATATTAAATACAAATTACATTTAAC	**Chien & Cheung, 1998**
>Hla TTTTTTATTTAATAGTTAATTAATTG	**Chien *et al.*, 1999**
>FnbA TCTGATGACTTGAATACAATTTATAGGTATATTCCAAATAATAAAATTATCAATTAACA TAAAATAATGACAATCTTAACTTTTCATTAACTCGCTTTTTT	**Wolz *et al.*,2000**
>Bap ATCATGTATAATATATATAGATGCAACTTTAATTCAATTTTTTGAAATGTACATAATTA GATATATATTAACTTTAATA	**Trotonda et al., 2005**
>Tst CAAACACTTTTTAATTAATATATATTTAAACAATAATTTAGAGATGGTTAATTGATTCA TTTAAATAATATTTATACATTC	**Andrey et al., 2010**
>FmtA ATTTGTATTTTTTACTATAAAGTACT	**Zhao *et al.*, 2012**

- Dès 1998, Chien et Cheung ont montré l'effet activateur de SarA sur l'expression des deux promoteurs du système *agr*. Ils ont réussi à déterminer une zone de 45 pb

protégée par SarA lors d'expériences d'empreinte à la DNase I (Chien and Cheung, 1998).

- L'année suivante, la mise en place par Chien *et al.* de la séquence consensus avait permis de localiser, parmi les 97 nt du gène *hla* protégées par SarA, une séquence de 26 nucléotides proche de la séquence consensus (Consensus2 Chien) (Chien et al., 1999).
- Les séquences protégées par SarA pour les gènes *tst, bap, fmtA et fnbA* identifiées respectivement par Andrey *et al.,* Trotonda *et al.,* Zhao *et al.* et Wolz *et al.* sont discutables (Andrey et al., 2010; Trotonda et al., 2005; Wolz et al., 2000; Zhao et al., 2012). En effet, pour les quatre expériences d'empreinte à la DNase I, le nombre de nucléotides protégés par SarA est proche de 100. Nous pensons que même si SarA est capable de fixer cette séquence, il semble peu probable que le dimère protège de manière spécifique l'intégralité de ces nucléotides. Ces séquences contiennent sans doute un site initial de fixation de SarA qui entraînerait ensuite une agglomération de SarA autour de la fixation initiale.

Nous avons recherché un motif commun pour l'ensemble de ces sites protégés par SarA grâce au logiciel GLAM2.

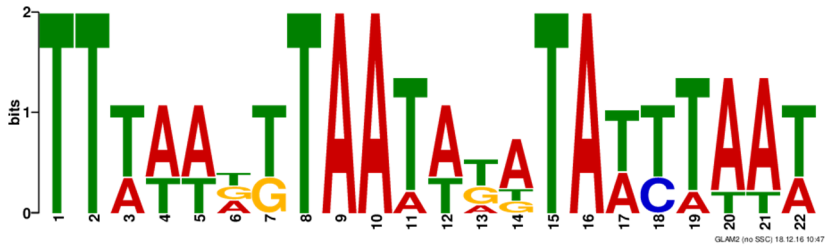


Figure 100 : Motif consensus (consensus6) entre les différents sites de fixation de SarA pour des gènes activés. L'analyse a été effectuée par le logiciel GLAM2

En analysant le motif présenté en Figure 100, nous constatons que la séquence de 22 nt (consensus6) est toujours composée majoritairement de thymines et d'adénines. Néanmoins, nous n'observons pas réellement de grandes séquences fortement représentées. Les séquences fortement conservées sont formées de 2 ou 3 nt ce qui ne permet pas l'obtention d'une séquence consensus forte. La séquence consensus6 du fait de sa faiblesse nous semble alors inadéquate pour une recherche globale des cibles directes de SarA. De plus, nous avons remarqué que si les séquences utilisées, pour la détermination d'un motif commun, sont de tailles variables, alors le motif obtenu ne permet pas de rassembler le maximum de données. La séquence consensus identifiée est alors, comme dans le cas du consensus6, peu informative.

Afin de pouvoir déterminer une séquence consensus forte, nous avons décidé d'utiliser les résultats que nous avions obtenus lors de l'étude de la régulation directe de *srn_3610_sprC* et *srn_9340* par SarA.

Après avoir confirmé la fixation *in vitro* et *in vivo* de SarA sur *srn_3610_sprC* et *srn_9340*, nous nous sommes intéressés à la séquence commune entre $sprC_{47}$ et srn_9340_{51} (sites de protection de la dégradation de la DNase I par SarA). L'alignement des séquences de protection de SarA, obtenues après empreintes à la DNase I, a été effectué par le logiciel Clustal Omega (Figure 101).



Figure 101 : Alignement des séquences *sprC*$_{47}$ (bleue) et *srn_9340*$_{51}$ (verte). Les boîtes -10 et -35 sont respectivement encadrées en rouge et en orange. Les flèches pleines correspondent à la séquence palindromique. Les flèches en pointillés indiquent les séquences similaires à la séquence consensus déterminée par Sterba *et al.*

Sur cet alignement, nous retrouvons, souligné par un astérisque, les 29 nucléotides qui nous avaient permis de révéler *srn_9340* (Tableau 2). Les séquences de 22 nucléotides pour *srn_3610_sprC* et de 23 nucléotides pour *srn_9340*, révélées par les expériences de retard sur gels avec des sondes de 30pb chevauchantes, sont représentées en gras sur la Figure 101. Nous pouvons voir que ces séquences se chevauchent pour recouvrir l'intégralité de la région promotrice des deux gènes à partir de la boîte -35.

En analysant ces 42 nucléotides, nous observons une séquence palindromique située en amont du +1 de transcription. Cette séquence est intéressante car il est établi que les facteurs de transcription agissant en dimère (comme SarA) ont une préférence pour la fixation au niveau de palindrome (Stenz et al., 2011). Nous nous sommes ensuite interrogés sur la possible similarité entre ces séquences et les séquences consensus déterminées précédemment.

```
SprC          ----------GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG
Srn_9340      AATATTATATATACGCTTGAAAAAATAAAATAATAATATTATAATATAATC----------
Consensus4    ----------TATWWAAATATTAAATACAAAT--------------------------
                        *** ** *****  ***
```

Figure 102 : Alignement de la séquence consensus4 avec les séquences *sprC*$_{47}$ et *srn_9340*$_{51}$. (alignement réalisé par Custal Oméga).

En comparant la séquence consensus4 avec les séquences de *srn_3610_sprC* et *srn_9340* protégées par SarA, nous retrouvons une séquence de 17 nucléotides très similaire avec une identité de plus de 75% (Figure 102).

```
Srn_9340          AATATTATATATACGCTTGAAAAAATAAAATAATAATATTATAATATAATC---------------
SprC              ------------------GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG
Consensus3Sterba ------------------------ATAAAAT---------ATAAAAT-------------------
                                         *******         **** **
```

Figure 103 : Alignement de la séquence consensus3 déterminée par Sterba *et al.* avec les séquences *sprC*$_{47}$ et *srn_9340*$_{51}$. (alignement réalisé par Custal Oméga). La séquence écrite en rouge correspond à un ajout personnel d'une seconde séquence consensus.

En alignant la séquence consensus3 déterminée par Sterba *et al.* avec les séquences SprC$_{47}$ et Srn_9340$_{51}$, nous constatons que cette séquence est retrouvée parfaitement une fois (Figure 103, en noir) et partiellement une deuxième fois (Figure 103, séquence rouge, un mésappariement sur les 7 nt de la séquence) dans les deux séquences étudiées. SarA se fixe en dimère sur l'ADN, il est alors logique de retrouver deux petites séquences consensus de fixation de SarA.

Ces résultats suggèrent que l'identification d'une séquence consensus de fixation de SarA sur l'ADN est possible. Nous avons donc choisi de rechercher une nouvelle séquence consensus en prenant en compte les données de certaines cibles directes de SarA.

Pour initier la mise en place d'une nouvelle séquence consensus, nous avons décidé de prendre les séquences de *srn_3610_sprC*, *srn_9340* et *sarV*, trois gènes régulés négativement par SarA et pour lesquels un site de fixation avait été déterminé suite à des expériences d'empreinte à la DNase I. La taille des zones protégées pour ces trois gènes est similaire (autour de 50 pb).

```
>Srn_3610_SprC

AAGTGTATAATATCAATTTACTACCTCGCAATATAATTTTATACAATGTATAATAAGATTTAAAGATGTAGCTCA
TCAACATTTTCATAGGGGATTTTTAAAATTAAGTATTGAAAAATAAAATATTTATATTATAATATAAATAAGTCA
ACGACCATGCGTGGACAGTAAAACGAAGTACATGCGAATGTAGTAGCCGTTTACTATGAGACTAGGCATTATAAT

>Srn_9340

ATCAAATGTGAATATGTTTTAAGCGTTGTATGTACATCTGCATGACCTAATCGTTTTGAAATGTAATGTTTAGCT
GTTACTTAAAGAAAATCAAATATTATATATACGCTTGAAAAAATAAAATAATAATATTATAATATAATCAATCGG
CATCTACGTGCGGATTATAAAGTAAAGCGTAGCATAACCAGGAGTTAGTCGATTTTTAATAAATTCGACTAAATT

>SarV

AAAGATATAAATCTAAACAAGATATAGCCAGCAATTTAATATTTGTAATAGATAAAATGCTAAGTTTGATATATA
ATAAATTTAACTAATTGTATAATAATATGAATTACAAACATCTAAGAAGAAACATAGGAGGCATCATATTATGAG
TAATAAAGTTCAACGTTTTATAGAAGCAGAAAGGGAGTTAAGTCAGTTAAAGCACTGGTTAAAAACAACACATAA
```

Figure 104 : Séquences de *srn_3610_sprC*, *srn_9340* et *sarV* (promoteur et gène). Les boîtes -10 et -35 sont respectivement écrites en rouge et en vert. Le +1 transcription est en gras. Le site de fixation du ribosome est écrit en violet et la phase codante du gène est en bleu. Les sites de fixation de SarA déterminés par des expériences d'empreinte à la DNase sont encadrés en rouge.

Nous avons choisi de faire nos études sur des séquences de 225 pb qui englobent les régions promotrices et géniques de ces trois gènes. Nous espérons que l'utilisation des séquences de tailles identiques nous donnera un motif fortement représenté. Nous avons ensuite intégré ces séquences au logiciel GLAM2 pour déterminer un motif commun aux trois

gènes. Nous avons décidé de partir « à l'aveugle », c'est-à-dire de ne pas utiliser directement les sites de protection par SarA, pour ne pas introduire de biais dans l'étude.

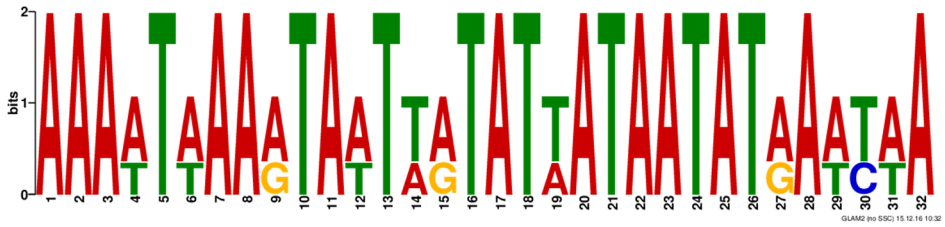


Figure 105 : Motif commun (consensus7) entre les différentes séquences de *srn_3610_sprC*, *srn_9340* et *sarV*. L'analyse a été effectuée par le logiciel GLAM2.

Ci-dessus, est présenté le motif commun aux trois séquences (consensus7 : AAAwTwAArTAwrTATwATAATATrAwywA). Ce motif est composé de 32 nucléotides, très riche en adénine et thymine et est assez conservé (70% d'identité).

>Srn_3610_SprC

AAGTGTATAATATCAATTTACTACCTCGCAATATAATTTTATACAATGTATAATAAGATTTAAAGATGTAGCTCA
TCAACATTTTCATAGGGGATTTTTAAAATTAAGTATTGAAAAATAAAATATTTATATTATAATATAAATAAGTCA
ACGACCATCCGTGGACAGTAAAACGAAGTACATGCGAATGTAGTAGCCGTTTACTATGAGACTAGGCATTATAAT

>Srn_9340

ATCAAATGTGAATATGTTTTAAGCGTTGTATGTACATCTGCATGACCTAATCGTTTTGAAATGTAATGTTTAGCT
GTTACTTAAAGAAAATCAAATATTATATATACGCTTGAAAAAATAAAATAATAATATTATAATATAATCAATCGG
CATCTACGTGCGGATTATAAAGTAAAGCGTAGCATAACCAGGAGTTAGTCGATTTTTAATAAATTCGACTAAATT

>SarV

AAAGATATAAATCTAAACAAGATATAGCCAGCAATTTAATATTTGTAATAGATAAAATGCTAAGTTTGATATATA
ATAAATTTAAGTAATTGTATAATAATATGAATTACAAACATCTAAGAAGAAACATAGGAGGCATCATATTATGAG
TAATAAAGTTCAACGTTTTATAGAAGCAGAAAGGGAGTTAAGTCAGTTAAAGCACTGGTTAAAAACAACACATAA

Figure 106 : Séquences de *srn_3610_sprC*, *srn_9340* et *sarV* (promoteur et gène). Les boîtes -10 et -35 sont respectivement écrites en rouge et en vert. Le +1 transcription est en gras. Le site de fixation du ribosome est écrit en violet et la phase codante du gène est en bleu. Les motifs obtenus à la Figure 105 sont soulignés par une vague. Les sites de fixation de SarA déterminés par des expériences d'empreinte à la DNase sont encadrés en rouge.

Lorsque nous localisons ce motif sur ces séquences, nous constatons que pour *srn_3610_sprC* et *srn_9340,* le motif bio-informatique (consensus7) est complètement contenu dans les séquences protégées déterminées expérimentalement. En ce qui concerne *sarV*, le motif est partiellement contenu dans la séquence de protection determinée expérimentalement. Il se place juste en aval de la boîte -10 (Figure 106). Le fait de retrouver, sans *a priori*, un motif de fixation au niveau des zones protégées de la dégradation de la DNase I est une preuve qui tend à penser que SarA, lors de sa fixation à l'ADN, se positionne sur un motif similaire pour chacune de ces 3 cibles.

Nous pouvons donc avec les séquences consensus4 (regroupant les données bibliographiques) et consensus7 (issu de nos données expérimentales) partir à la recherche d'autres cibles de SarA.

II. Utilisation des séquences consensus pour la recherche de cibles

Nous avons initié l'identification de cibles de SarA par la recherche de la séquence consensus4 dans l'ensemble du génome de la souche de *S. aureus* NCTC8325. Nous avons utilisé le logiciel Glam2Scan (serveur MEME) la séquence suivante : ATTTGTATTTAATATTTwwATA (consensus4).

Tableau 6 : Résultat de la recherche de la séquence consensus4 dans le génome de la souche NCTC8325 (en rouge sont représentées les cibles connues de SarA ; + : cible activée par SarA ; - : cible réprimée par SarA).

	Position 5'	Séquence	Position 3'	Régulation	Nom	Localisation
1	**2093774**	**ATTTGTATTTAATATTTTAACA**	**2093753**	**+**	**agr**	**promoteur**
2	960950	ATTTTTATTTAGTATTTAAATA	960971	?	N-acetyl-L,L-diaminopimelate aminotransferase	terminateur
3	1860968	ATTTGTATTAAAAATTTATATA	1860947	?	lukE	promoteur
4	222595	CTTTGTATTTCATATTTAAAAA	222574	?	protease PrsW / nickel ABC transporter substrate-binding protein	promoteur
5	1077762	ATTTTTATTTAATAGTTAATTA	1077741	?	srn_2480/2490	promoteur
6	**1847439**	**ATTTTTATTTAAAATGTATATA**	**1847418**	**+**	**splA**	**promoteur**
7	242153	ATTTGTATTTGATATTGAAATC	242174	?	tagB putative	gène
8	2182369	ATTTGTCTTGCATATTTATATA	2182390	?	Thymidine kinase	promoteur
9	2424055	ATTTATATTTAATATTTAT.TA	2424075	?	tcaR/membrane protein putative	promoteur
10	2097485	ATTTGCATTTAATGTTTTGGTA	2097506	?	fructokinase putative	gène
11	1981633	ATTTTTACTTAACATTTTGATA	1981612	?	ferritin putative	promoteur
12	593171	CTTTGTATTTGATAATTCTATA	593150	?	tandem five-TM protein	promoteur
13	1521369	ATTTGTATGTTATATTTTCTTA	1521348	?	Short-chain dehydrogenase/pyrroline-5-carboxylate reductase	promoteur
14	587660	TTTTATATTTAATAATTCTATA	587639	?	PtpA	promoteur
15	249437	ATCTGCATTTAATATTTCTACA	249416	?	srn_0670	gène
16	589072	TATTGTATTTAATAATTCTATA	589051	?	tandem five-TM protein	promoteur
17	596497	AATTGTATTTAATGGTGATATA	596518	?	oxidoreductase	terminateur
18	2133542	ATTTATATTTCAAATTATTATA	2133563	?	STAS domain putative	promoteur
19	**2335387**	**GTTTGTAATTCATATTATTATA**	**2335408**	**-**	**sarV**	**promoteur**
20	718056	ATTTTTATTTAATGCTTTAAAA	718077	?	permease putative	gène
21	2376452	ATTTGCATTTAAAATTTTAAGT	2376473	?	Inositol monophosphatase	gène

22	2575751	ATTTATATTTTATATATATATT	2575772	-	sarT	promoteur
23	1172685	CTTTGTAATTTATTTTTTAATA	1172706	?	DNA topoisomérase	promoteur
24	1332019	CTTTGTATTTGATATTCTTATC	1332040	?	Phosphate ABC transporter	gène
25	76371	ATTTGTATATAAAGTATTAATA	76392	-	sarS	promoteur

Parmi les 25 premiers résultats obtenus (Tableau 6), nous retrouvons 17 séquences situées dans des régions promotrices de gènes, 6 au coeur de gènes et 2 dans des régions terminatrices de gènes. Nous constatons donc que la séquence consensus4 est majoritairement retrouvée au sein de différents promoteurs de gènes cibles. Ces résultats renforcent l'idée que le consensus4 correspond bien à une zone de fixation par un facteur de transcription tel que la protéine SarA.

5 résultats correspondent à des cibles déjà connues pour être fixées ou régulées par SarA. La séquence consensus4 est donc retrouvée dans le promoteur de *sarV, sarT* et *sarS*, tous les trois sont des facteurs de transcription de la famille SarA. On retrouve également le consensus4 dans la région promotrice de la sérine protéase *splA* et enfin au cœur des promoteurs du système *agr.*

Au regard des autres candidats, nous retrouvons majoritairement des protéines membranaires (candidats n°4 ; 9 ; 12 ; 13 ; 16 ; 18 ; 20 et 24) et des protéines impliquées dans le métabolisme bactérien (candidats n°2 ; 4 ; 8 ; 10 ; 11 ; 14 ; 17 et 21). Retrouver de nombreuses protéines membranaires dans les cibles putatives de SarA n'est pas surprenant. En effet, SarA est caractérisée comme régulateur positif de la transcription de gènes codant pour des protéines membranaires (Trotonda et al., 2005; Wolz et al., 2000).

En ce qui concerne les autres candidats, nous retrouvons trois gènes transcrits en ARN régulateurs : *srn_2480*, *srn_2490* et *srn_0670*. Néanmoins, nous avons été déçus de ne pas retrouver les deux ARN régulateurs dont nous savons que leur expression est réprimée par SarA (*srn_3610_sprC* et *srn_9340*). Même en augmentant le nombre de résultats à 100 (données non montrées), aucun de ces deux gènes n'est identifié. L'utilisation de la séquence consensus4 ne permet pas l'identification de l'ensemble des cibles directes de SarA.

En prenant le motif décrit à la Figure 105 (consensus7 : AAAwTwAArTAwrTATwATAATATrAwywA) et en le recherchant dans le génome de *S. aureus* NCTC8325 grâce au logiciel GLAM2SCAN, nous souhaitions identifier des cibles déjà connues pour être régulées par SarA ou de nouvelles cibles dont l'expression pourrait être impactée par SarA.

Les 25 premiers résultats, donnés par GLAM2SCAN, sont présentés dans le Tableau 7.

Tableau 7 : Résultat de la recherche du motif consensus7 dans le génome de la souche NCTC8325 (en rouge sont représentées les cibles connues de SarA ; + : cible activée par SarA ; - : cible réprimée par SarA).

	Position 5'	Séquence	Position 3'	Régulation	Nom	Localisation
1	**1861583**	**TTATTTATATTATAATATAAATATTTTATTTT**	**1861614**	**-**	**Srn_3610_sprC**	**Promoteur**
2	**1863904**	**TTGATTATATTATAATATTATTATTTTATTTT**	**1863935**	**-**	**Srn_9340**	**Promoteur**
3	**2335392**	**TAATTCATATTATTATACAATTACTTAAATTT**	**2335423**	**-**	**sarV**	**Promoteur**
4	2002327	TTTATTTTATTATAAAATAATTATATTAAATT	2002296	?	Staphopain thiol protéinase	Promoteur
5	889101	TTTTATATATT.TAATAATAATATTTTATTTT	889131	?	Phosphatidyleth anolamine-binding protein	Promoteur
6	1768962	TTATTTATATTGTAATGTTTCTATTTGATTAT	1768993	?	Thiol reductase thioredoxin	Promoteur
7	1078411	ATGATTAAATTATTATAGCAATACTTTATTGT	1078380	?	Srn_2500	Terminateur
8	949592	AAATATATATTATTATTTTATCATTTAATTTG	949561	?		Région inter-génique
9	405923	TACTTTATCTAATAAGATAATTTTATTATTTT	405892	?	Tandem lipoprotein	Gène
10	963445	TTATTTATAATATTATATGCTTTTTTTATAAT	963414	?	AtlR	Promoteur
11	562956	TTATTGATATTATACTATAATT.TATTATTTA	562926	?	SIS domain protein	Terminateur
12	**2579355**	**TTATAACTATTATTATACAAGTAATTTAACTT**	**2579324**	**+**	**FnbB**	**Gène**
13	1079782	TTATAAGAATTATTATATAATTTTTTACTTTT	1079813	?	Superantigen-like protein	Gène
14	726032	TTGTTT.TATGCTAATATAAATACGTTAATTT	726002	?	EamA family transporter	promoteur
15	1994986	TAAATCATATTAACATGTTTATCTTTTAATTT	1995017	?	DNA ligase	Gène
16	800833	TATTTAATATTAGAAAATAAAAATTTTAAATT	800864	?	CspC	promoteur
17	2333955	TTGCTTTAATTATCACGCAATTACTTTATTTT	2333986	?	MFS transporter	Gène
18	1269980	TAATATGTATTAAATTATAATTATTATAAATT	1270011	?	Catalase	Promoteur
19	2002289	TTAACTAAATT.TAATATAATTATTTTATAAT	2002319	?	Staphopain thiol protéinase	Promoteur
20	1610529	TAATTTATTTTTTAATATTGTTACTTCTTCTT	1610560	?	MTA/SAH nucleosidase	Promoteur
21	586691	TACTTTATCTTTTAATCCAAAAATATTATTTT	586660	?	Transposase	Gène
22	222635	TTTTCCATATTTTAATATAAATACTTTGTTCG	222604	?	Membrane protein putative	Promoteur
23	394085	TTAAT.ATATTTTTATATGATTATTTCAATTA	394055	?	Superantigen-like protein	Promoteur
24	**76228**	**TTAAAAATAATATGATATTTTTATTAAAATTT**	**76259**	**-**	**SarS**	**Promoteur**
25	2774267	TCAATGATTTGATAATATTATTCTTTATATTT	2774236	?	Srn_5070	Promoteur

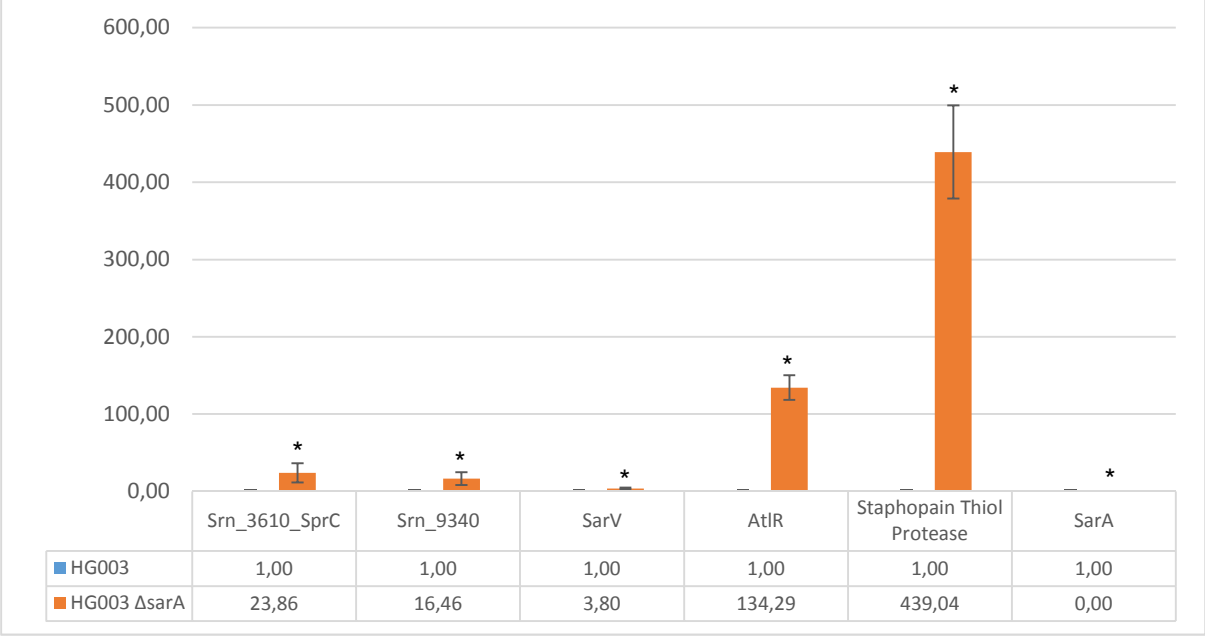
16 séquences correspondent à des régions promotrices de gènes, 6 régions géniques, 2 terminateurs et une région intergénique. Nous retrouvons donc une proportion importante de résultats issus de promoteurs, révélant de nouveaux la richesse en adénine et thymine de la séquence consensus, plus propice à être retrouvé dans les régions promotrices. Les 3 premières cibles sont les gènes ayant permis de définir la séquence à rechercher. En dehors

de ces 3 gènes, deux candidats sur les 22 restants sont déjà connus pour être régulés par SarA, il s'agit de FnbB et SarS.

Nous nous sommes interrogés sur l'éventualité d'une régulation de SarA sur d'autres candidats révélés par cette étude. Nous retrouvons deux séquences candidates pour le même gène, la staphopaine thiol protéase, *scpA* (n°4 et 19). Nous avons donc choisi d'étudier son expression dans un contexte mutant pour le gène *sarA*. Le second candidat dont nous voulions connaître la régulation par SarA est *atlR*, facteur de transcription répresseur de l'expression *d'atl* codant l'autolysine (candidat n°10). Nous savons que SprC est également répresseur de la traduction de l'autolysine *atl*. L'hypothèse où SarA régulerait deux répresseurs de l'expression d'*atl* est vraiment très intéressante.

Pour étudier l'impact de SarA sur l'expression de ces cibles déterminées précédemment, nous avons réalisé des RT-PCRq sur des ARN totaux des souches de *S. aureus* HG003 et HG003Δ*sarA* extraits après 4H30 de croissance en milieu TSB.

En utilisant des amorces spécifiques des ARN Srn_3610_SprC, Srn_9340, *sarV, atlR, scpA* et *sarA*, nous avons mesuré la quantité de chaque ARN présent dans la bactérie. Chaque valeur a été normalisée par la quantité d'ARN *hu*, produit de manière constitutive dans les deux souches étudiées. La quantité d'ARN dans la souche HG003 est ensuite ramenée à la valeur de 1 pour pouvoir détecter les variations dans la souche HG003Δ*sarA*.



	Srn_3610_SprC	Srn_9340	SarV	AtlR	Staphopain Thiol Protease	SarA
HG003	1,00	1,00	1,00	1,00	1,00	1,00
HG003 ΔsarA	23,86	16,46	3,80	134,29	439,04	0,00

Figure 107 : Expression des cibles identifiées par la recherche du motif dans le génome de la souche HG003. Les ARN des souches HG003 et HG003Δ*sarA* ont été extraits à 4h30 de croissance. La quantité des ARN a été mesurée par PCR quantitative et normalisée par rapport à la quantité d'ARN *hu*. La moyenne de trois expériences indépendantes (+/- écart-type) a été représenté. (test unilatéral de Mann-Whitney : * : p-value<0,05%)

Dans un premier temps, nous constatons que dans la souche HG003Δ*sarA*, il n'y pas plus aucun transcrits *sarA*. Nous confirmons ainsi que la souche HG003Δ*sarA* est valide pour l'étude de l'expression de gènes en absence de SarA.

L'expression des 3 gènes, ayant servi à identifier le motif, a également été mesurée. Nous observons que les trois transcrits sont plus présents dans la souche HG003Δ*sarA* que dans la souche HG003. L'ARN Srn_3610_SprC est augmentée d'environ 24 fois dans la souche mutée par rapport à la souche sauvage, pour Srn_9340, il y a une augmentation de 16 fois et pour l'ARN *sarV* 4 fois. Nous confirmons donc qu'en absence de SarA, la quantité de ces ARN est augmentée, SarA réprime leur expression. De plus, les données bibliographiques et nos expériences ont montré que cette répression se manifestait par fixation directe de SarA sur leur promoteur.

En ce qui concerne le répresseur de l'autolysine *atlR,* nous constatons une augmentation de 134 fois la quantité d'ARN dans la souche HG003Δ*sarA* par rapport à la souche HG003. L'absence de SarA entraîne la forte augmentation de l'ARN *atlR.* SarA est donc un répresseur très fort de l'expression d'*atlR*. SarA réprime donc deux répresseurs de l'expression de l'autolysine, *srn_3610_sprC* et *atlR*.

La quantité d'ARN *scpA* (Staphopain Thiol protéase) est environ 440 fois plus importante dans la souche HG003Δ*sarA* par rapport à la souche HG003. SarA réprime donc très fortement l'expression de ce gène. Cette observation est cohérente avec le fait que la souche HG003Δ*sarA* est incapable de former un biofilm. En effet, la Staphopain thiol protéase est une protéine qui, avec d'autres protéases, intervient dans la dispersion du biofilm bactérien.

Le biofilm bactérien est un processus entraînant l'agglomération de plusieurs bactéries entre-elles dans une structure tridimensionnelle afin de pouvoir mieux résister aux agressions extérieures (système immunitaire, action d'antibiotiques,…). Le biofilm est initié par la reconnaissance des cellules via des protéines membranaires, comme l'autolysine par exemple. Nous constatons donc ici que SarA réprime trois gènes dont l'expression défavorise soit le biofilm directement (avec la staphopain thiol reductase) soit l'expression de protéines impliqués dans celui-ci (avec Srn_3610_SprC et AtlR).

SarA est capable de fixer l'ADN que sous forme de dimère et nous supposons donc que chaque monomère fixe l'ADN sur une région spécifique. En analysant la séquence consensus7 (Figure 105), nous avions positionné deux sites consensus3Sterba. Cette observation a amené l'idée de l'existence de deux sites de fixation de SarA sur les promoteurs de ces gènes cibles. En analysant les séquences promotrices des différentes cibles de SarA, nous nous sommes aperçus que la plupart contenait deux petites séquences consensus

séparées d'une séquence aléatoire. Ce motif est le suivant : ATAAAAT-12nt-ATATA (Consensus8). Ce type de séquence ne peut être trouvé que « manuellement » du fait de la variabilité des 12 nt présents à l'intérieur de ces deux séquences. Nous avons ensuite recherché ce motif dans le génome de *S.aureus*.

Les 10 premiers résultats sont présentés dans le Tableau 8.

Tableau 8 : Résultat de la recherche de la séquence consensus8 dans le génome de la souche NCTC8325.

	Position du 5'	Séquence	Position du 3'	Régulation	Nom	Localisation
1	**1863908**	**ATAAAATAATAATATTATAATATA**	**1863932**	**-**	**srn_9340**	**promoteur**
2	**2335425**	**ATAAAATGCTAAGTTTGATATATA**	**2335449**	**-**	**sarV**	**promoteur**
3	**1861587**	**ATAAAATATTTATATTATAATATA**	**1861611**	**-**	**srn_3610_sprC**	**promoteur**
4	2709559	ATAAAATGAGATAACTTAAATATA	2709583	?	citrate transporter	promoteur
5	2691011	ATAAAATTCAGATTTGTAAATATA	2690987	?	srn_4940	promoteur
6	2443049	ATAAAATGAATATATTTAAAAATAA	2443073	?	phage infection protein	promoteur
7	566025	ATAAAAATACTTATTCATTATATAA	566001	?	srn_1490	promoteur
8	585207	ATAAAATTAAAATATTGAAAGATAA	585183	?		région intergénique
9	2206394	ATAAAATGAAGTAATTAATATTTAA	2206418	?		région intergénique
10	2196793	ATAAAAAGATTTATTTTCAATATAT	2196817	?	saouhsc_02373	gène

Parmi ces séquences candidates, 7 correspondent à des promoteurs, 2 à des à une région intergénique et une dernière à une séquence intragénique. Nous retrouvons les trois séquences promotrices, *srn_3610_sprC*, *srn_9340* et *sarV*, que nous avons précédemment utilisées pour la mise en place du premier motif. Nous identifions une protéine membranaire permettant le transport de citrate. De manière intéressante, nous constatons que deux autres ARN régulateurs ont été identifiés : *srn_1490* et *srn_4940*. Nous avons décidé d'étudier l'expression des gènes *saouhsc_02943* (codant pour le transporteur de citrate), *srn_1490* et *srn_4940*.

Nous avons mesuré, par RT-PCRq, la quantité d'ARN Srn_3610_SprC, Srn_9340, *sarV*, *saouhsc_02943*, Srn_1490, Srn_4940 et *sarA* dans les souches HG003 et HG003Δ*sarA*. Comme précédemment, nous confirmons l'absence de transcrits *sarA* dans la souche délétée du gène *sarA*. L'augmentation de la quantité d'ARN Srn_3610_SprC, Srn_9340 et *sarV* confirme le rôle répresseur de SarA sur l'expression de ces gènes.

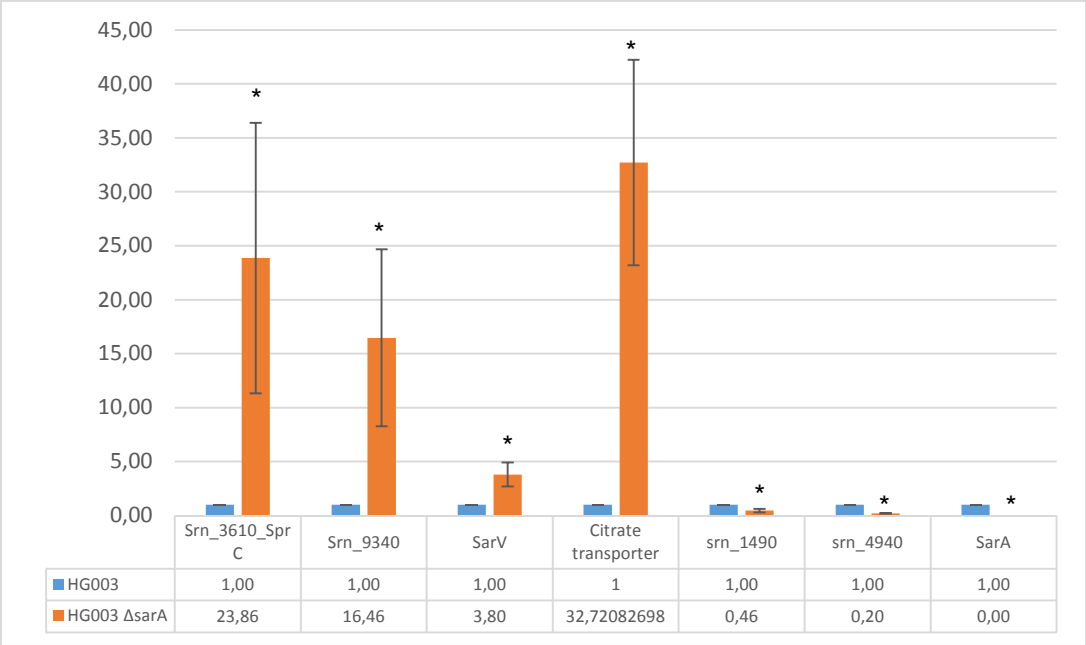


Figure 108 : Expression des cibles identifiées par la recherche de la séquence consensus8 dans le génome de la souche HG003. Les ARN des souches HG003 et HG003Δ*sarA* ont été extraits à 4h30 de croissance. La quantité des ARN a été mesurée par PCR quantitative et normalisée par rapport à la quantité d'ARN *hu*. La moyenne de trois expériences indépendantes (+/- écart-type) a été représenté. (test unilatéral de Mann-Whitney : * : p-value<0,05%)

Pour le transcrit *saouhsc_02943*, nous constatons que la quantité d'ARN est environ 33 fois plus importante dans la souche HG003Δ*sarA* que dans la souche sauvage (Figure 108). L'absence de SarA entraîne une surexpression du gène *saouhsc_02943*. SarA serait un répresseur de l'expression du gène codant pour le transporteur de citrate.

En comparant la quantité des ARN régulateurs Srn_1490 et Srn_4940 dans la souche HG003 et HG003Δ*sarA*, nous constatons une diminution de l'expression de *srn_1490* et *srn_4940* dans la souche HG003Δ*sarA* (respectivement d'environ 2 et 5 fois) (Figure 108). L'absence de SarA impacte donc négativement la production de ces deux ARN régulateurs. SarA serait donc un activateur de l'expression de *srn_1490* et *srn_4940*.

III. Détermination de cibles de SarA par une méthode à haut débit

Pendant l'écriture de ce manuscrit, nous avons décidé de réaliser une analyse à haut débit pour déterminer les cibles de SarA. Je ne présenterai ici que la préparation des échantillons, l'analyse RNA-seq étant en cours de réalisation.

Nous avons choisi de réaliser le séquençage à haut-débit des ARN (RNA-seq) des souches HG003 et HG003Δ*sarA*. Après 2 heures ou 4h30 de croissance dans le milieu TSB, les ARN totaux des souches ont été extraits. Deux traitements à la DNase I permettent

d'éliminer les traces d'ADN résiduelles. Nous avons ensuite déplété les ARN ribosomaux des ARN totaux extraits. Cette étape est nécessaire pour obtenir une plus grande profondeur de séquençage pour les ARN d'intérêt.

Le séquençage des ARN permettra de recenser l'intégralité des cibles directes et indirectes du facteur de transcription SarA.

Ces résultats seront comparés aux données transcriptomiques de Dunman et al., 2001. Cette analyse identifiera les cibles ARNrég qui ne l'ont pas été dans la précédente étude. De plus, nous avons déjà listé un certain nombre de cibles dont nous connaissons la régulation par SarA, ces cibles nous permettront de dresser un seuil de sélectivité.

Enfin, nous isolerons les promoteurs de toutes les cibles déterminées par RNA-seq. Grâce à l'étude précédente permettant la mise en place d'une séquence consensus de fixation de la protéine SarA, nous tenterons de retrouver le motif Consensus8 dans les promoteurs étudiés. Ces données nous permettront de faire une liste des cibles directes de SarA.

Pour valider nos résultats, des expériences de retard sur gel seront réalisées afin de confirmer la fixation de SarA sur le promoteur et des expériences d'empreinte à la DNase I détermineront une zone fixatrice de SarA.

L'intégralité des résultats nous permettra :

(1) de classer les cibles directes dont l'expression est soit activée soit réprimée par SarA

(2) de déterminer les zones de fixation de SarA

(3) d'établir un schéma général de régulation de SarA selon la localisation de son site de fixation.

IV. Site de fixation de SarA et régulation

Le motif consensus8 a permis d'identifier à la fois des cibles dont l'expression est activée ou réprimée par SarA. Néanmoins, l'analyse par RT-PCRq proposée précédemment met en évidence une régulation par SarA mais n'indique pas si SarA agit directement sur le promoteur comme il peut le faire pour srn_3610_sprC, srn_9340 ou sarV, ou indirectement via d'autres régulateurs. Même s'il parait nécessaire de vérifier ce lien direct par notamment des expériences de retards sur gel, nous avons recherché où se situe le site de fixation putatif de SarA sur les séquences promotrices de *atlR*, *scpA, saouhsc_02943, srn_1490* et *srn_4940*.

Nous savons que l'expression de *srn_3610_sprC* et *srn_9340* est réprimée par SarA en se fixant sur les boîtes -10 et -35 (sites de fixation de l'ARN polymérase). Trois nouveaux

gènes ont été identifiés comme cibles régulées négativement par SarA (*scpA, atlR* et *saouhsc_02943*). Nous avons voulu savoir si le motif ayant permis leur identification se situait comme pour *srn_3610_sprC* et *srn_9340* au niveau des sites de fixation de l'ARN polymérase

Pour cela, nous avons récupéré, dans la base de données NCBI, la séquence codante des gènes *scpA, atlR* et *saouhsc_02943*. Nous avons élargi élargi la séquence au 150 nt situés en amont du gène afin d'obtenir la région promotrice. Nous avons ensuite replacé le site de la fixation du ribosome RBS (en violet) sur cette séquence (Figure 109).

Pour la détermination du +1 de transcription, nous avons utilisé les données de Shaw *et al.* qui spécifiaient la localisation des boîtes -10 et -35 pour le gène *scpA* (Shaw et al., 2004). Ces données ont été validées par les travaux de Prados *et al.* (Prados et al., 2016). La localisation du +1 de transcription d'*atlR* et de *saouhsc_02943* a dû être prédite. En effet, aucune étude n'a localisé cette position pour le répresseur de l'autolysine et pour le transporteur de citrate. Néanmoins, nous avons pu identifier les boîtes -10 et -35 du fait de leur très forte homologie avec les séquences consensus de fixation du facteur σ^{70} d'*E. coli*. Ainsi, nous avons pu localiser un +1 de transcription putatif.

>Staphopain thiol protease (scpA : SAOUHSC_02127)
TTGCAATTTAGTAAATAAGCTACTAAATTATTACATAAACTGATAATGTGTAAAAATTAATTATTTTTTAAACAT
AGATTAAAATTTTATATTGAAACTCTTAACTAAATTTAATATAATTATTTTATAATAAAATAAAGGAGTTTTTAT
ATGAAAAGAAACTTTCCAAAATTAATAGCTTTAAGTTTAATATTTTCTTTAAGTGTTACCCCTATTGCAAACGCT
GAGAGCAATTCAAATATCAAAGCAAAAGATAAGAAGCATGTGCAAGTTAATGTCGAGGACAAGAGTGTTCCTACT
GACGTACGTAATTTAGCGCAAAAAGACTATCTTTCTTATGTTACTTCACTTGATAAAATCTATAATAAAGAGAAA
GCCTCGTATACACTTGGTGAACCTTTTAAAATTTATAAATTCAACAAAAAGAGTGATGGTAATTATTATTTTCCA
GTTTTGAACACTGAAGGTAATATCGATTATATAGTTACAATTTCACCAAAAATCACTAAGTATTCTAGTTCATCT
TCTAAATACACAATCAATGTTTCTCCATTTTTATCTAAAGTCTTAAATCAATATAAAGATCAACAAATTACTATT
TTGACAAACTCAAAAGGTTACTATGTGGTTACTCAAAATCATAAAGCTAAACTAGTACTAAAAACACCACGTCTA
GAAGATAAAAAGTTAAAAAAGACTGAGAGTATACCTACTGGCAATAATGTAACGCAACTAAAGCAAAAAGCTTCT
GTCACTATGCCAACTAGTCAATTTAAAAGTAATAATTATACTTACAACGAGCAATATATAAATAAACTTGAGAAT
TTCAAGATAAGAGAAACTCAAGGTAACAATGGTTGGTGTGCAGGCTATACGATGTCTGAATTACTAAATGCGACT
TATAATACTAACAAATATCATGCAGAAGCTGTGATGAGATTTTTACACCCAAACTTACAAGGACAACGATTCCAA
TTTACTGGTTTAACGCCAAGAGAGATGATTTATTTTGGACAAACTCAGGGAAGAAGTCCACAACTTCTTAATA…
>AtlR (SAOUHSC_00992)
TTTATCTCCTTTATGTCCATAAGACTCTTACTAATTTATTATATTTTGTATTAGTTATGCAACAAATTTGTCTAT
TGTGTCGAATAGTATTGTAAAAAATTATAAAAAAAGCATATAATATTATAAATAATATGTAGTGGAGTAGAACAAC
ATGTATAAACAACTTGAAAAACTTATTACACTGACTAACAATGACTTAAACTTAGTGAATAGAAGATTTGGACAA
CGCACGGATATCACATCTGAACAGCTAGAACTTCTCCGTATTTTATTTAATTACGATCGCTTATCACAGTATGAT
TTGACGATGAAGATTAGCAGGGAACAATCTATAGTTTCAAGGTGGATTAAGAAATTAGTTTTGAAAGGGTACATC
ACAAGTCAACAATCTAGCGAAGATTTAAGATGTAAAGAATTAATTTTGACTGATCAAGCACGTACATTAATTTCA
CAAATAAATAATGCACGTTGCGAATTGATTGAAGCAAGATGTCAATGTTTATCGGAAGTCGAATTAGACAATTTA
AATCAATTACTTGATAAGTTAAATCAACGACGCATATCGTTGTAA
>Saouhsc_02943 (Citrate transporter)
AGATCAAAATGAAATACAGGACAGGATATACAAGGATTAGTCATGCCATGTTATCAAGTAGGAAAATCAAACTTC
ACTATTGATAGTTACGCAAAAAAGATTTTTTTGATAAAATGAGATAACTTAAATATAAAAAAATTATATTAATTAT
AATATTTAAGTTAAAGAGGGGGATTATGTAAATTGTATTAAAAGTGGAGGGAGAAAATAATATGAATAGTGATAA
TATGTGGTTAACAGTAATGGGGCTCATTATTATTATTTCAATTGTAGGTTTACTCATTGCCAAAAAGATAAATCC

Figure 109 : Séquences de *scpA, atlR* et *saouhsc_02943* (promoteur et gène). Les boîtes -10 et -35 sont respectivement écrites en rouge et en vert. Le +1 transcription est en gras et souligné. Le site de fixation du ribosome est écrit en violet et la phase codante du gène est en bleu. La séquence ayant permis la révélation de ces gènes est surligné en jaune.

Après avoir replacé l'ensemble des sites caractéristiques pour chaque promoteur (Figure 109), nous avons surligné en jaune les sites de fixation putatifs de SarA obtenus lors de l'analyse bio-informatique (Figure 109).

Nous constatons que, pour les trois gènes candidats, les motifs obtenus se localisent au niveau de la boîte -10 des promoteurs. Les caractéristiques de fixation de SarA sur *scpA, atlR* et *saouhsc_02943* semblent similaires de celles de *srn_3610_sprC* et *srn_9340* avec une fixation de SarA sur la région promotrice. Un mode de répression identique pourrait être envisagé : SarA réprimerait la transcription de *scpA, atlR* et *saouhsc_02943* car sa fixation sur les régions promotrices de ces gènes masquerait les sites de fixation de l'ARN polymérase.

Nous avons localisé tout d'abord les sites putatifs révélés par le logiciel GLAM2 au niveau des promoteurs des gènes de *srn_1490* et *srn_4940*, gènes activés par SarA. Les nucléotides caractéristiques (+1 de transcription, boîtes -10 et -35) ont été positionnés pour *srn_1490* et *srn_4940* grâce aux données obtenues des travaux de Prados et al., 2016 (Figure **110**).

>Srn_1490
GTTTTCAACTGTTCTAAAATTACGTTCATCAATATCCCTCACTCAATGTAAACTTTAAAATACATATTGTTAACA
TTATATAATGAATAAGTATTTTTATCAGGTAAACCTGCCTATTATAAAAATCAAACTTAAGCCCTAGATAAATTT
GATACTGTAGTACCATGTTTTTTATTAGATAAGACGTTACTATATGTATATAGAGAAAAGAAGTTATCGCAATCG
CAATAACTGACACAAATAAGTCACCCAATCTATTTAAGTTTGTGTTCTCTATAATAATATTGAAGGTTAGAGTCT
>Srn_4940
GTGGCCTCTCTTAAATCGATTATAAAATCCGATTTAAAATCCTGCTCAAATACCGTCATGCCTCTATGAATACTA
TCCAATATACTAATAAAATTCAGATTTGTAAATATATTCTGATAATTAAGAATTTTATATCAATCTACTTTGTTT
ATGCGCATAATTTTGTTCGATTAGGGAATTGTTTACTATATTCCATTTGACTCTGAGGTGATACGAATGCGTTTA
AAAGTATTGTTCCATTTCATTGCTGCTATTTTCATATCATTCATGTTGCTATGGATGACCATGTTGTTTGACTTA

Figure 110 : Séquences de *saouhsc_02943, srn_1490* et *srn_4940* (promoteur et gène). Les boîtes -10 et -35 sont respectivement écrites en rouge et en vert. Le +1 transcription est en gras et souligné. Le site de fixation du ribosome est écrit en violet et la phase codante du gène est en bleu. La séquence ayant permis la révélation de ces gènes est surlignée en jaune.

Pour *srn_1490* et *srn_4940,* les sites de fixation putatifs de SarA se localisent en amont de la boîte -35, à 99 pb pour *srn_1490* et à 75 pb pour *srn_4940* (par rapport au +1 de transcription). Or, nous observons que l'expression de ces gènes est activée par SarA. Ces résultats sont cohérents avec une localisation des sites de fixation de SarA non chevauchante avec les sites de fixation de l'ARN polymérase.

L'ensemble de ces résultats permettent de proposer une hypothèse à propos de la régulation de l'expression des gènes par SarA. SarA, quel que soit son rôle régulateur, reconnait et fixe des motifs proches de celui déterminés précédemment (5'-ATAAAAT-12nt-ATATA-3'). Lorsque le site de reconnaissance de SarA est situé au niveau des boîtes -10/-35, alors SarA agira comme un répresseur de la transcription. Lorsque le site de reconnaissance de SarA est situé en amont de la boîte -35, alors SarA pourrait activer l'expression du gène

cible. Si cette hypothèse est correcte, nous pourrions déterminer le rôle activateur ou répresseur de SarA sur l'expression d'un gène en étudiant la position du site de fixation de SarA dans la région promotrice.

V. Discussion et proposition d'un modèle de régulation

Dans cette partie, nous avons essayé de déterminer des séquences consensus pour la fixation de SarA. Nous avons recensé dans le Tableau 9 les différentes séquences consensus présentées. Nous y avons ajouté des commentaires sur les forces et les faiblesses de chaque séquence.

Tableau 9 : Récapitulatif des différentes séquences consensus. Les nucléotides sont représentés par leur initiale ou en suivant la nomenclature suivante : w : A, T ; d : T, A, G ; r : A, G ; k :T,G ; y : T,C.

Nom	Séquence consensus	Forces de la séquence	Faiblesses de la séquence
Consensus1 (Chan)	**aaAtNTaatTtgtattNaAtattTtaNtN**	Définie à partir des promoteurs de 7 protéases (gènes réprimés par SarA)	Ne prend pas en compte les gènes activés par SarA ; non confirmée par des expériences d'empreintes à la DNase I
Consensus2 (Chien)	**ATTTgTATtTAATATTTataTAAtTg**	Définie à partir de gènes activés et réprimés + confirmation par des expériences d'empreintes à la DNase	
Consensus3 (Sterba)	**ATTTTAT**	Motif facile à identifier ; beaucoup de séquences identifiées	Petites tailles, perte du pouvoir discriminant
Consensus4	**ATTTGTATTTAATATTTwwATA**	Combinaison de deux séquences consensus (**1** et **2**) et contient **consensus3**	
Consensus5	**wwwdAwAwTATwwwdAwTTTw**	Définie par séquences protégées par SarA sur gènes réprimés ; séquence **consensus3** (Sterba) identifiée	Beaucoup de variabilité
Consensus6	**TTwwwdkTAAwwddTAwywwww**	Définie par séquences protégées par SarA sur gènes activés	Beaucoup de variabilité ; pas de séquence **consensus3** (Sterba) identifiée
Consensus7	**AAAwTwAArTAwrTATwATAATATrAwywA**	Définie par alignement de séquences de *srn_3610_sprC, srn_9340* et *sarV* et cohérent avec les empreintes à la DNase I	
Consensus8	**ATAAAAT-12nt-ATATA**	Permet la séparation en deux parties pour s'adapter à la fixation en dimère de SarA	L'espace entre les deux séquences est aléatoire et variable, difficile à implémenter dans la recherche de cibles

Ce tableau reflète combien il est difficile d'établir une séquence consensus parfaite pour la fixation de SarA. Néanmoins, les données, obtenues lors de cette étude et celles que

nous acquerrons lors de l'expérience de RNA-seq, seront une mine d'informations qui, nous l'espérons, permettra de définir plus précisément cette séquence consensus.

Grâce aux séquences consensus 7 et 8, nous avons déterminé cinq nouvelles cibles de SarA, trois d'entre elles voient leur expression réprimée par SarA alors que pour les deux autres nous observons une activation de leur expression. En émettant l'hypothèse que ces cibles sont directement régulées, nous pouvons les ajouter aux cibles déjà connues dont leur promoteur est fixé par le dimère SarA. Ainsi, en positionnant les sites de fixation de SarA putatifs ou vérifiés sur les promoteurs, nous pouvons obtenir deux modèles de régulation selon que SarA est répresseur ou activateur de la transcription (Figure 111).

- Lorsque SarA réprime l'expression de *srn_3610_sprC, srn_9340, sarV* (Figure 106), *scpA, atlR* (Figure 107) et *saouhsc_02943* (Figure 108), nous constatons que le site de fixation réel ou putatif (analogie de séquence) de SarA est positionné au niveau des boîtes -10 et/ou -35 du promoteur (Figure 111A). Ce positionnement empêcherait alors l'ARN polymérase de se fixer comme nous avons pu le montrer pour *srn_3610_sprC* et *srn_9340*.
- A l'inverse, lorsque SarA active l'expression de *srn_1490* et *srn_4940,* nous observons que les sites de fixation de SarA, déterminés par homologie de séquence avec le motif (Tableau 8), se placent en amont de la boîte -35 (Figure 111B). Dans ce cas, nous supposons que la fixation de SarA facilite le recrutement de l'ARN polymérase sur le promoteur induisant ainsi une augmentation de la transcription.

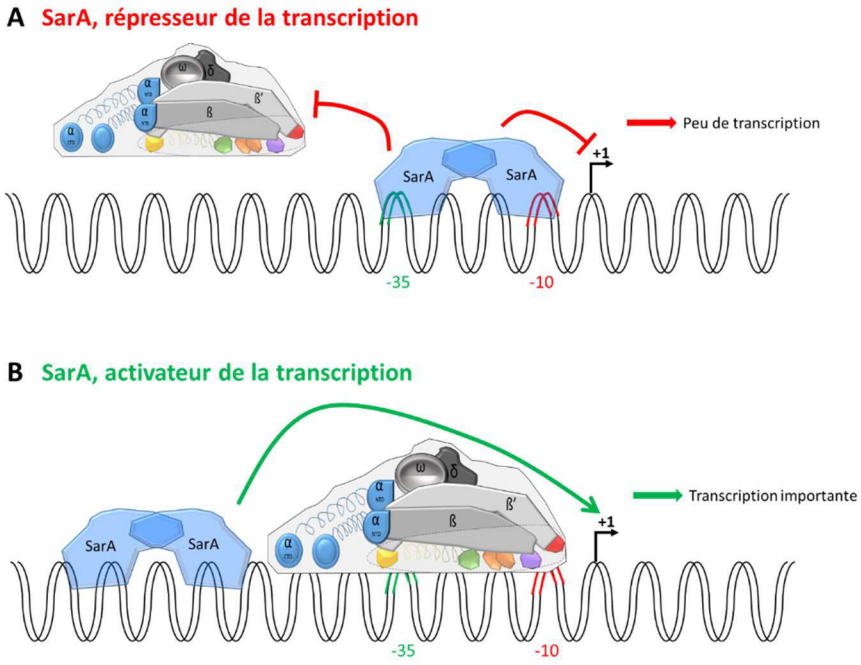


Figure 111 : Modèle de répression (A) et d'activation (B) par SarA de l'expression de gènes selon la position du dimère sur le promoteur.

Les modèles proposés à la Figure 111 évolueront avec les nouvelles données obtenues après analyses des résultats de RNAseq. En effet, seule une analyse sur un plus grand nombre de cibles directes permettra de définir les modes d'action activateur et répresseur de SarA.

CONCLUSION ET PERSPECTIVES

CONCLUSION ET PERSPECTIVES

L'objectif principal de ce travail était d'appréhender la fonction d'un ARN régulateur de *S. aureus*, Srn_3610_SprC, au travers de la régulation transcriptionnelle de son gène. Srn_3610_SprC est impliqué dans l'atténuation de la virulence et la phagocytose par des macrophages humains (Le Pabic et al., 2015). Nous avons montré que le facteur de transcription SarA réprime la transcription de *srn_3610_sprC* par fixation au niveau des sites de liaison de l'ARN polymérase sur son promoteur. L'analyse du site de fixation de SarA a permis de caractériser un second ARN régulateur, Srn_9340, dont l'expression est également réprimée par SarA. Par ailleurs, le mode de répression opéré par SarA sur les promoteurs de ces deux ARNrég semble similaire (Mauro et al., 2016 : ANNEXE 2). Il a été montré que l'expression de *srn_3610_sprC* est fortement augmentée lors des premières minutes de la phagocytose (Le Pabic et al., 2015). Ce résultat semble indiquer qu'à cet instant précis, le promoteur de *srn_3610_sprC* est accessible pour l'ARN polymérase, c'est-à-dire que SarA ne s'y positionne plus. Cette forte expression de *srn_3610_sprC* semble transitoire puisqu'il a été observé une diminution de l'expression peu après la phagocytose (Le Pabic et al., 2015).

Dans cette partie, nous discuterons dans un premier temps des méthodes à utiliser pour identifier d'éventuels autres régulateurs transcriptionnels de *srn_3610_sprC* et *srn_9340*. Ensuite, nous nous intéresserons aux différentes modifications post-traductionnelles de SarA (phosphorylation, oxydation) et de leur impact sur la capacité de fixation de cette dernière sur ses promoteurs cibles. De plus, les signaux extérieurs permettant les modifications de SarA seront discutés. La relation entre Srn_3610_SprC et Srn_9340 sera abordée ; de même qu'il sera proposé des expériences pour déterminer les cibles de ces ARNrég. Enfin, pour intégrer les connaissances obtenues sur Srn_3610_SprC et Srn_9340 au cœur du régulon SarA, je vous présenterai une stratégie expérimentale combinant plusieurs expériences à haut débit afin de déterminer l'ensemble des cibles de SarA (Figure 112).

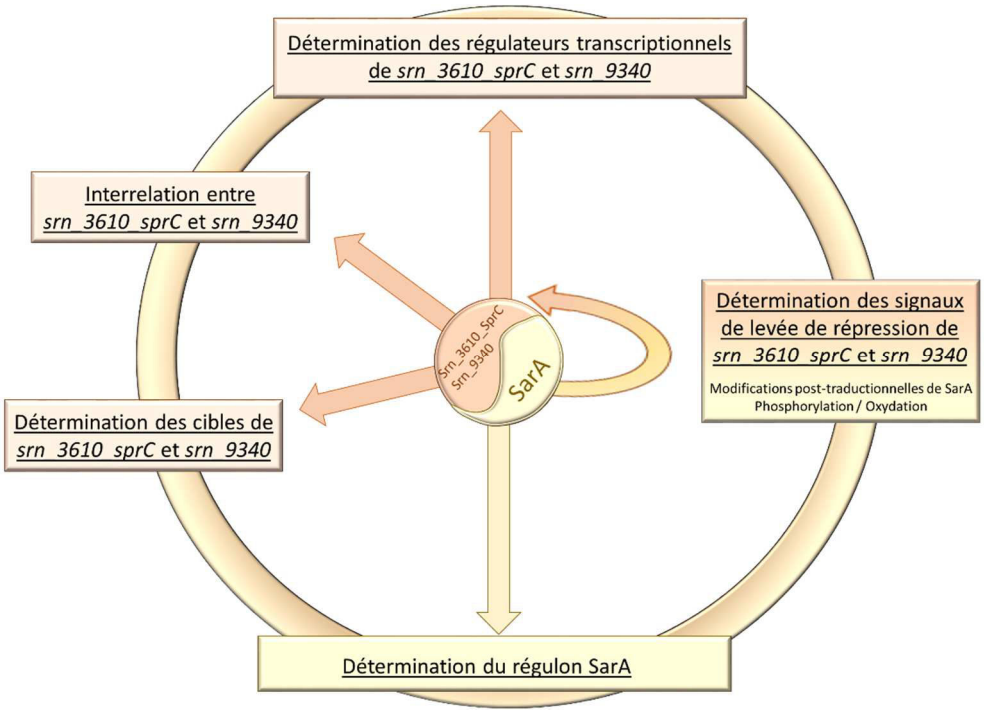


Figure 112 : Représentation des différentes perspectives de ce travail de thèse

Cette étude pourrait se poursuivre avec l'**identification des autres régulateurs de l'expression des gènes *srn_3610_sprC* et *srn_9340***. Si le lien entre SarA et les régions promotrices *srn_3610_sprC* et *srn_9340* et son rôle répresseur dans leur expression ne sont plus à démontrer, il reste encore beaucoup à faire pour déterminer l'ensemble des régulateurs de ces ARNrég. Notre étude a permis d'identifier l'ARN III comme acteur de la régulation de *srn_3610_sprC* et *srn_9340* (cf pages 68 et 86). L'ARN III pourrait agir selon deux modes : 1) l'ARN III agirait sur la stabilité des deux ARN régulateurs, 2) l'ARN III modifierait l'expression d'un gène codant pour un régulateur transcriptionnel de *srn_3610_sprC* et *srn_9340.* Pour tester la première possibilité, il serait intéressant de mesurer et comparer la demi-vie de Srn_3610_SprC et Srn_9340 dans une souche HG003Δ*arnIII* par rapport à une souche sauvage. En ce qui concerne la seconde hypothèse, il a été montré que l'ARN III modulait l'expression de deux facteurs de transcription : il réprime l'expression de *rot* et active celle de *mgrA* (Gupta et al., 2015) (Boisset et al., 2007; Geisinger et al., 2006). Alors que nous n'avons pas observé de différences notables dans l'expression de *srn_3610_sprC* et *srn_9340* entre la souche sauvage et celle délétée du gène *rot*, des études préliminaires indiqueraient que le facteur de transcription MgrA réprimerait l'expression de *srn_3610_sprC* (étude non réalisée pour *srn_9340*). Néanmoins, cette observation a été effectuée dans la souche RN6390. Même si RN6390 est dérivée de la souche NCTC8325, comme l'est HG003, les deux souches sont assez éloignées d'un point de vue génétique. La délétion du gène *mgrA* dans la souche HG003 est en cours. L'étude de l'expression de *srn_3610_sprC* et *srn_9340* dans la souche HG003Δ*mgrA* permettra de confirmer ou d'infirmer le rôle répresseur de ce facteur de

transcription dans l'expression de *srn_3610_sprC* et d'étudier son rôle potentiel dans l'expression de *srn_9340*.

Au vu du grand nombre de régulateurs transcriptionnels identifiés et potentiels (Figure 18), l'étude de l'impact de la délétion d'un gène, codant pour chacun de ces facteurs de transcription, sur l'expression de *srn_3610_sprC* et *srn_9340* parait difficilement réalisable. En effet, pour ce faire, il faudrait des mutants de délétion pour chaque facteur de transcription potentiel et ce dans le contexte génétique HG003. Aussi, je propose deux nouvelles stratégies de recherche pour **l'identification de régulateurs** : le simple hybride ou de la résonance plasmonique de surface (RPS ; technologie Biacore) couplée à la spectrométrie de masse.

a) La première technique consiste à identifier les protéines capables de se lier à des promoteurs d'intérêt dans un système hétérologue (levure). Elle nécessite la création d'une librairie d'expression de protéines fusionnées à un domaine d'activation et la création d'une fusion entre le promoteur du gène d'intérêt et un gène rapporteur. Cette technologie *in vivo* est puissante mais présente quelques limitations. En étant dans un système hétérologue et en ajoutant un domaine d'activation aux protéines testées, nous ne pouvons pas être assurés que la structure tridimensionnelle des protéines sera conservée. De plus, la capacité de fixation des protéines sur l'ADN pourrait être également impactée. Nous imaginons également que cette technologie n'identifiera que les protéines avec une fixation « classique » à l'ADN. Si elles ont besoin de modifications post-traductionnelles pour fixer le promoteur, alors nous serons dans l'impossibilité de caractériser ces protéines.

b) Succinctement, la technique de RPS consiste à fixer, sur une puce, le promoteur d'intérêt et de le mettre, dans notre cas, en présence d'extraits protéiques bruts de *S. aureus*. Les protéines capables de fixer la séquence d'ADN particulière du promoteur vont être retenues sur la puce. En couplant la RPS à une spectrométrie de masse, nous pourrions caractériser l'ensemble des régulateurs potentiels de *srn_3610_sprC* et *srn_9340*. Cette technologie permet également d'établir les affinités des protéines pour ces régions promotrices. Néanmoins elle présente le désavantage d'être une technique *in vitro* et est donc susceptible de ne pas être le reflet des fixations qui ont lieu *in vivo*.

La recherche de signaux de transduction permettant le décrochage de SarA des promoteurs de *srn_3610_sprC* et *srn_9340* permettrait de replacer l'expression de ces ARNrég dans la physiologie de la bactérie. La poursuite de cette étude semble donc importante et pourrait être abordée de deux manières :

a) par une étude des capacités de liaisons de SarA sur ces promoteurs en fonction des modifications post-traductionnelles de la protéine. La fixation de SarA sur les promoteurs est

connue pour être impactée par la phosphorylation de SarA (Sun et al., 2012). Nous avions essayé de phosphoryler *in vitro* la protéine SarA par la kinase Stk1 purifiée. Cette expérience n'était pas concluante du fait de la très grande activité de l'enzyme qui ne permettait pas d'obtenir une gamme de protéine SarA phosphorylée suffisamment étendue pour permettre cette analyse. Cependant, il existe, dans la littérature, des souches de *S. aureus* produisant les versions mutées des kinases Stk1 et Stk2 qui sont incapables de s'auto-phosphoryler et donc de phosphoryler leurs cibles (Didier et al., 2010). Nous pourrions utiliser ces souches pour savoir si la phosphorylation ou la déphosphorylation de SarA est nécessaire pour permettre la fixation ou le décrochage des promoteurs. De plus, une fois le signal identifié, nous pourrions utiliser les souches mutées pour ces kinases pour déterminer ces kinases sont actrices de la transduction du signal de levée de répression de *srn_3610_sprC* et *srn_9340.*

En parallèle, nous souhaiterions étudier l'impact de l'oxydation de SarA sur l'expression de *srn_3610_sprC* et *srn_9340.* Dans des expériences préliminaires utilisant la construction du promoteur d'intérêt fusionnée au gène rapporteur *gfp*, nous avions essayé d'induire un stress oxydant en ajoutant du peroxyde d'hydrogène (H_2O_2). Nous n'avions constaté aucune modification de l'activité des promoteurs de *srn_3610_sprC* et *srn_9340* autant dans la souche HG003 que la souche mutée pour le gène *sarA*. Néanmoins, il serait intéressant d'utiliser d'autres agents oxydants car il a été montré que selon l'agent appliqué, l'expression de gènes habituellement réprimés par SarA est augmentée. Par exemple, les gènes *sodA, sodM* et *trxB*, dont l'expression est faible en présence de SarA, sont fortement exprimés en présence de l'agent oxydant diamide (Ballal and Manna, 2009b). De plus, il a été démontré que la capacité de fixation de SarA sur le promoteur de *trxB* est diminuée en présence de diamide (Ballal and Manna, 2010).

Nous avons identifié que le Triton X-100 entraînait une diminution de l'activité des promoteurs P_{sprC} et P_{srn_9340} dans la souche HG003Δ*sarA*. L'ajout du Triton X-100 entraîne donc soit une diminution de l'expression d'un activateur soit une augmentation de l'expression d'un répresseur (cf page 120). Ce détergent peut également provoquer une modification de l'affinité des régulateurs pour les promoteurs. Le Triton X-100 est décrit pour engendrer un stress d'autolyse, or MgrA et SarV (deux membres de la famille SarA) ainsi que les systèmes à deux composants, LytSR et ArlRS, sont connus pour être impliqués dans les mécanismes d'autolyse (Brunskill and Bayles, 1996; Liang et al., 2005; Manna et al., 2004; Trotonda et al., 2009). Il serait alors intéressant d'étudier l'impact de MgrA, comme proposé précédemment, mais également de LytSR et ArlRS sur l'expression de *snr_3610_sprC* et *srn_9340.* D'un autre côté, il a été surprenant de ne voir aucun effet dans la souche sauvage. En effet, il a été montré que l'expression de *sarA* est augmentée dans une solution contenant du Triton X-100 (Cordwell et al., 2002). Nous aurions attendu une diminution de l'activité des promoteurs P_{sprC}

et P_{srn_9340} par l'augmentation de la quantité de la protéine SarA. Nous supposons alors que soit la technologie de gène rapporteur (*gfp* ou *blaZ*) ne permet pas d'observer les diminutions, sans doute faibles, de l'activité des promoteurs, soit que la surproduction de SarA n'engendre pas une « sur-répression » des promoteurs d'intérêt.

b) je poursuivrais la recherche du signal de dérépression, notamment dans des expériences de phagocytose par des macrophages humains. En effet, Le Pabic *et al.* a montré l'implication de Srn_3610_SprC dans la phagocytose de *S. aureus* par des macrophages humains. Il s'agirait d'étudier l'expression de *srn_9340* lors de la phagocytose. Des études préliminaires, réalisées dans la souche Newman, indiqueraient que l'expression de *srn_9340* est augmentée lors de l'internalisation de la bactérie par des macrophages humains. Si cette observation est confirmée, nous aurions donc, au moins, deux ARN régulateurs dont l'expression est impactée positivement par des macrophages (contact/internalisation). Il a été montré que l'expression de *sarA* est légèrement activée à 30 et 60 minutes après phagocytose par des neutrophiles polynucléaire (Voyich et al., 2005). Ces résultats sont cohérents avec la répression de *srn_3610_sprC,* une heure après que la bactérie est phagocytée (Le Pabic et al., 2015). Néanmoins, Srn_3610_SprC est fortement présent 5 minutes après internalisation. Nous pouvons imaginer que lors du contact cellule hôte/pathogène, un signal est transmis à la bactérie entraînant une modification de la protéine SarA. SarA modifiée serait dans l'impossibilité de fixer le promoteur de ces cibles, les ARN Srn_3610_SprC et Srn_9340 seraient alors produit massivement. Il serait intéressant alors de suivre l'activité des promoteurs de ces deux ARNrég au cours de la phagocytose. Grâce aux plasmides pCN52c-P_{sprC} ou pCN52c-P_{srn_9340}, nous pourrions déterminer à quel moment nous observons une augmentation de la fluorescence synonyme d'une dérépression de ces promoteurs. Par cette analyse, nous déterminerions si le signal de surexpression de *srn_3610_sprC* et *srn_9340* arrive avant le contact avec le macrophage (via des protéines sécrétées), lors du contact cellule/bactérie (par des protéines de surface) ou bien durant l'internalisation du pathogène (variation de l'osmoralité, stress oxydant). Une fois le signal identifié, nous pourrions étudier si ce signal impacte la capacité de fixation de SarA pour les promoteurs d'intérêt.

Nous pourrions également nous intéresser **aux interrelations qui existent entre les ARN Srn 3610 SprC et Srn 9340 ainsi qu'à la recherche des cibles de ces deux ARNrég**. En étudiant Srn_3610_SprC dans une souche délétée du gène *srn_9340*, nous avons constaté que sa quantité est augmentée d'un facteur 2 par rapport à la souche sauvage (données préliminaires). Nous supposons alors que soit il agit sur la production d'un facteur de transcription impliqué dans l'expression de *srn_3610_sprC*, soit Srn_9340 agit sur la stabilité de Srn_3610_SprC. Des études préliminaires ont montré que la légère surexpression de *srn_3610_sprC* est due, entre autres, à une modification de l'activité du promoteur

(données récentes préliminaires). Ces résultats doivent être confirmés mais nous pouvons imaginer une relation étroite entre ces deux ARN régulateurs réprimés par le même facteur de transcription. L'action entre Srn_3610_SprC et Srn_9340 pourrait être double, une action sur la stabilité en plus de l'action sur l'expression du gène. Nous pourrions alors étudier l'éventualité d'une fixation de ces ARN entre eux. Pour cela, des expériences de retard sur gel (ARN-ARN) entre Srn_3610_SprC et Srn_9340 (produits par transcription *in vitro*) peuvent être effectuées. Si la liaison est confirmée, nous pourrions étudier la demi-vie de Srn_3610_SprC en absence de Srn_9340 et inversement. Ainsi, nous pourrions répondre à la question d'une modification de la stabilité de ces ARN lors de leur association. Quel que soit le résultat de ces expériences, je poursuivrais la recherche de leurs cibles.

Le séquençage à haut débit des ARN (RNA-seq) de la souche NewmanΔ*srn_3610_sprC*, comparé à celui de la souche Newman sauvage, a été effectué au laboratoire. Aucune différence significative dans la quantité d'ARN (messagers ou régulateurs) entre les deux souches n'a été trouvée, il n'y a donc pas eu l'émergence de cibles pour Srn_3610_SprC. Dans la souche Newman comme dans la souche HG003, le gène *srn_3610_sprC* est très faiblement exprimé. La quantité de cet ARN est donc faible. La comparaison entre une souche ayant peu de Srn_3610_SprC et une souche qui n'en a pas ne semble pas adéquat. La surexpression des ARNrég mime une levée de répression de SarA sans impacter les autres cibles de ce facteur de transcription. C'est pourquoi une étude transcriptomique par RNA-seq sur des souches surexprimant *srn_3610_sprC* et/ou *srn_9340* identifiera les cibles de ces ARNrég. Il faut tout de même garder à l'esprit que les ARNrég agissent majoritairement sur la traduction d'ARNm. Ainsi, les résultats obtenus regrouperont les cibles directes (par modification de la stabilité) et principalement les cibles indirectes (par variation des niveaux des régulateurs transcriptionnels par exemple. La caractérisation des cibles indirectes est très intéressante car elle servira de base pour définir les rôles physiologiques de Srn_3610_SprC et Srn_9340.

Au laboratoire, Simon Reynaud (étudiant en thèse) a effectué une expérience de MS2-TRAP (MS2-tagged RNA affinity purification) pour identifier les cibles ARN de Srn_3610_SprC. Pour cela, il a construit un ARN Srn_3610_SprC comprenant à son extrémité une étiquette MS2. Cette dernière est reconnue par la protéine de fusion MS2-GST (glutathione S-transferase) qui elle-même sert à la fixation sur une colonne de glutathion. Ainsi, les ARN liés à Srn_3610_SprC sont ensuite élués et séquencés pour identification. L'identification des cibles ARN de Srn_3610_SprC est donc en cours. Nous pourrions effectuer cette même expérience pour Srn_9340 dont aucune cible n'a pour l'instant été identifiée. Cette technologie permet de caractériser des interactions ARN-ARN *in vivo*. Néanmoins, le fait de réaliser cette technique avec un ARN de fusion (ARN-MS2) pourrait engendrer des biais par de possibles

changements de structure de l'ARN. Ces changements entraîneraient alors soit une diminution de l'affinité pour des ARN cibles soit l'apparition de candidats faux positifs (cibles non fixées par l'ARN dans des conditions standards). Les candidats faux positifs seraient facilement discriminés par des expériences de retard sur gel entre les deux ARN.

Comme présenté dans la section Résultats, nous envisageons de privilégier l'impact de la surproduction de Srn_3610_SprC et/ou Srn_9340 sur la production de protéines. Pour cela, nous pourrions effectuer, dans un premier temps, des gels protéique 1D afin d'identifier, par spectrométrie de masse, des protéines différemment produites entre une souche HG003 sauvage et des souches surexprimant *srn_3610_sprC* ou *srn_9340* (utilisation des plasmides de surexpression contenant le promoteur P_{veg}). Nous pourrions également envisager de faire cette analyse à plus large échelle en passant en 2D-DIGE qui permet une séparation des protéines selon leur point isoélectrique et leur masse moléculaire. Ces expériences révéleraient l'ensemble des protéines dont les ARNm seraient les cibles directes ou indirectes de Srn_3610_SprC et Srn_9340.

Comme nous l'avons dit précédemment, SarA réprime l'expression de deux gènes transcrits en ARN régulateurs. Nous savons que SarA est impliquée dans de nombreuses fonctions comme la formation de biofilm, la résistance aux stress oxydants ou l'antibiorésistance. La répression de SarA sur *srn_3610_sprC* et *srn_9340* semble similaire à celle observée pour les gènes codant pour les protéases (*aur, sspA*) ou les protéines impliquées dans la résistance aux agents oxydants (*sodM, sodA, trxB*). Cette observation permet d'avoir des pistes sur le rôle physiologique de Srn_3610_SprC et Srn_9340. Si leur gène est exprimé comme ceux codant les protéases, nous pouvons alors penser que Srn_3610_SprC et Srn_9340 seraient des ARN régulateurs impliqués dans la dispersion du biofilm. Nous pourrions effectuer des expériences de formation de biofilm avec les souches surexprimant *srn_3610_sprC* et *srn_9340* (grâce au promoteur P_{veg}). Si l'hypothèse de départ est correcte, ces souches devraient avoir une capacité diminuée dans la formation du biofilm bactérien.

L'implication de Srn_3610_SprC dans la résistance d'un stress oxydant (H_2O_2) a été observée par Le Pabic *et al.* (Le Pabic et al., 2015). La surproduction de Srn_3610_SprC est délétère à la croissance bactérienne en présence d'un stress oxydant (H_2O_2). Dans un premier temps, nous pourrions étudier l'impact d'une surexpression de *srn_9340* sur la croissance en stress oxydant. Néanmoins, l'action néfaste de la surexpression de Srn_3610_SprC sur la résistance au stress oxydant est en désaccord avec l'action de SodA, SodM et TrxB. En effet, SarA réprime l'expression de *sodA, sodM* et *trxB*, qui produisent des molécules permettant d'agir contre les agents oxydants. Cependant, SarA réprime *srn_3610_sprC* qui lorsque ce

gène est surexprimé entraîne une diminution de la croissance lors d'ajout d'une quantité sub-létale de peroxyde d'oxygène.

Pourquoi SarA réprime-t-il ces gènes alors qu'ils ont des rôles contraires ? Nous pouvons soulever l'hypothèse qu'ils ne répondent pas aux mêmes signaux de dérépression ou que Srn_3610_SprC permet d'établir des nuances dans la résistance aux stress oxydants. Est-ce que SodA, SodM et TrxB sont des cibles de Srn_3610_SprC ? Ces questions montrent la nécessité d'un côté de rechercher les cibles de Srn_3610_SprC et de Srn_9340 mais également de caractériser les signaux permettant la levée de répression de SarA pour l'expression de ces gènes.

L'étude de *srn_3610_sprC* et *srn_9340* proposée englobe la compréhension de leur régulation transcriptionnelle jusqu'à la détermination de leur rôle physiologique. Cette analyse globale est nécessaire pour intégrer ces deux ARNrég dans les réseaux de régulation de *S. aureus*. **La recherche des cibles directes du facteur de transcription SarA pourrait être poursuivie par la détermination du régulon SarA, régulon auquel appartiennent Srn_3610_SprC et Srn_9340.**

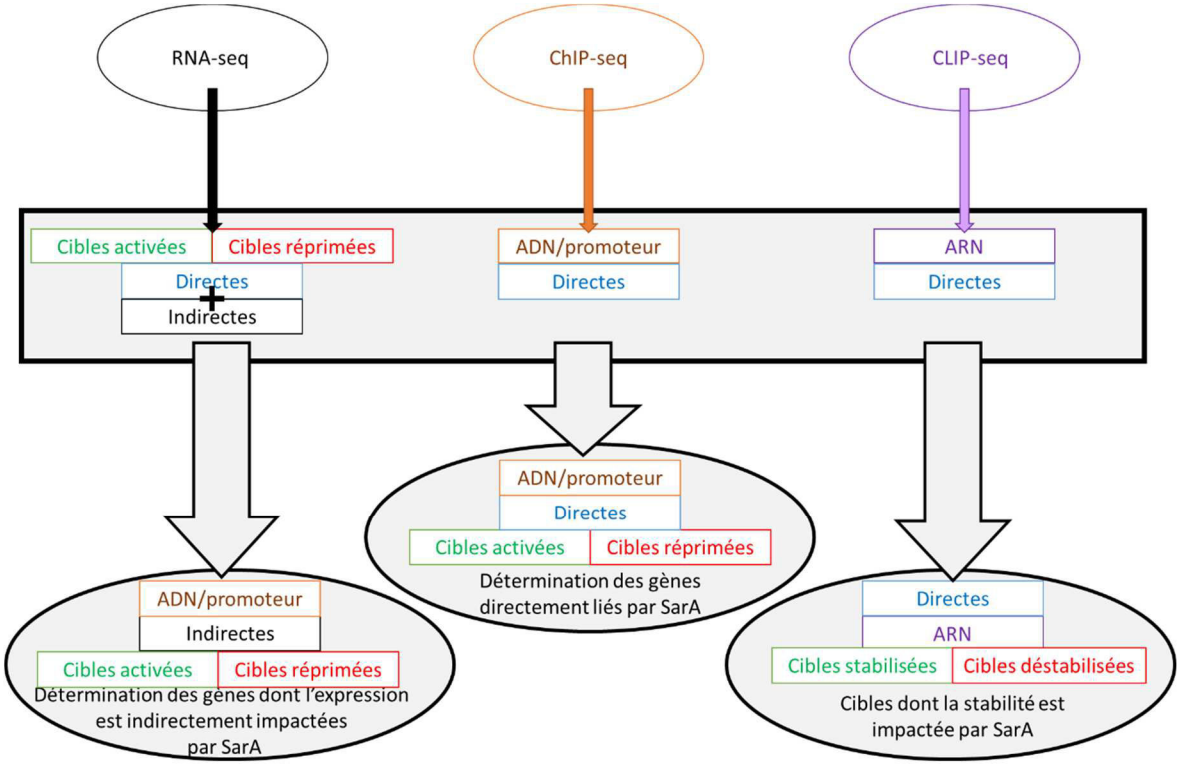


Figure 113 : Stratégie expérimentale pour la détermination du régulon de SarA

Pour réaliser cette étude, nous proposons une approche combinatoire basée sur trois expériences de séquençage à haut débit (RNA-seq, ChIP-seq et CLIP-seq) (Figure 113). Dans ce travail de thèse, nous avons initié la recherche bio-informatique des nouvelles cibles de

SarA en nous basant sur les sites de fixation connus à ce facteur de transcription. Ainsi, nous avons identifié cinq nouveaux gènes dont l'expression est impactée par l'absence de *sarA* (cf pages 144 et 147). SarA réprime l'expression d'*atlR* (répresseur de l'autolysine), *scpA* (staphopaïne thiol réductase) et *saouhsc_02943* (transporteur de citrate). De plus, SarA active l'expression de deux gènes transcrits en ARN régulateur (*srn_1490* et *srn_4940*). Nous n'avons pas encore caractérisé les mécanismes de régulation de SarA pour ces gènes, mais le fait de les avoir identifiés grâce à des homologies de séquences avec le site de fixation de SarA suggère que SarA est capable de lier le promoteur de ces gènes. Afin de tester cette hypothèse, nous envisageons d'effectuer des expériences de retard sur gel entre ces promoteurs et la protéine SarA purifiée.

Cependant, il paraît peu aisé de déterminer l'ensemble des cibles de SarA en utilisant ce même procédé. C'est pourquoi, nous avons débuté la mise en place d'un séquençage à haut-débit des ARN (RNA-seq) issu de la souche HG003Δ*sarA* (en comparaison avec la souche HG003). Les résultats obtenus engloberont les cibles directes et indirectes de SarA dans des conditions de culture conventionnelle de laboratoire après 2h et 4h30 de croissance. L'analyse des promoteurs des gènes cibles identifiés permettra de différencier, grâce aux données du site de fixation que nous possédons, ceux qui seraient directement liés par SarA et ceux dont l'expression seraient modifiée de manière indirecte. En parallèle, nous pensons réaliser des expériences d'immunoprécipitation à haut débit (ChIP-seq). Cette technologie est la même que celle effectuée au cours de cette thèse et décrite précédemment mais elle se termine avec le séquençage de tous les fragments d'ADN fixés par la protéine d'intérêt qui est, dans notre cas, SarA. En gardant les mêmes conditions expérimentales que celles du RNA-seq, nous allons pouvoir ainsi déterminer l'ensemble des promoteurs liés par le facteur de transcription SarA. Enfin, nous pourrions dériver la technique du ChIP-seq pour rechercher les ARN liés à SarA (CLIP-seq). Pour cela, au lieu de lier de manière covalente, par ajout de formaldéhyde, les protéines avec l'ADN, nous pouvons appliquer un rayonnement ultraviolet afin de lier les protéines aux ARN. Cette expérience se justifie d'autant plus qu'il a été montré que SarA est capable de fixer des ARN afin de modifier leur stabilité (Morrison et al., 2012).

Les résultats de ces trois expériences pourront être compilés et fournir une liste des cibles de SarA. Une analyse plus fine des séquences promotrices cibles, couplée à des comparaisons d'expression, détermineront le type de régulation impliqué (positive ou négative ; directe/indirecte) (Figure 113). De plus, nous pourrions affiner avec encore plus de précision le(s) séquence(s) consensus de fixation de SarA par rapport à ce que nous avons essayé de faire durant cette thèse : Y a t'il un consensus activateur ou represseur ou tout simplement un consensus de fixation de SarA ? Ces résultats permettront de déterminer le

modèle de régulation effectué par SarA : un modèle global de fixation ou spécifique des modes d'activation ou de répression.

Enfin, les résultats obtenus par les expériences proposées amèneront d'autres expériences. En effet, après avoir identifié les signaux de dérepression de SarA pour *srn_3610_sprC* et *srn_9340*, nous pourrions imaginer renouveler ces expériences afin de déterminer les cibles de SarA impactées par ces signaux et donc d'identifier les cibles répondant à un même signal. Nous pourrons ainsi établir un régulon de SarA impliquant l'ensemble des cibles directes et indirectes de ce facteur de transcription selon l'arrivée d'un signal ou non.

Ce travail laisse entrevoir la possibilité d'identifier le régulon SarA, lequel évolue avec les conditions environnementales. Les liens entre SarA et l'antibiorésistance ou la formation du biofilm pourront être révélés par des études systématiques de recherches de cibles dans les conditions favorisant l'antibiorésistance ou le biofilm.

De manière plus globale, ce travail laisse entrevoir que la régulation des ARNrég pourrait être une régulation en cascade. Un signal extérieur activerait ou réprimerait un facteur de transcription lequel agirait sur l'expression d'ARN régulateurs qui eux-mêmes auraient plusieurs cibles. Parmi ces cibles, nous pourrions retrouver des facteurs de transcription entrainant à la fois une réponse étendue du signal et la possibilité de l'apparition de boucles de régulation. Ces phénomènes de régulation ne sont pas nouveaux mais ils montrent la complexité de s'intéresser à un seul effecteur pour étudier l'expression de gènes (ici SarA pour *srn_3610_sprC* et *srn_9340*). De plus, notre travail montre combien il est difficile d'identifier le signal déclencheur de l'activation ou la répression d'un régulateur. Pour cette étude, l'activation de *srn_3610_sprC* et *srn_9340* passe par le non recrutement de SarA sur le promoteur de ces deux gènes. Ce non recrutement pourrait être dû à une modification post-traductionnelle de la protéine qui ferait suite à la réception de signaux environnementaux. Dans notre étude les signaux et les modifications agissant sur SarA restent à identifier. Cependant, nous avons, au travers de ce travail de thèse, réussi à faire le lien, chez *S. aureus,* entre ARNrég du génome accessoire et facteur de transcription du core génome, et apporté de nouvelles connaissances sur la régulation transcriptionnelle de *srn_3610_sprC* et *srn_9340*. Ces données permettent d'intégrer les ARN régulateurs Srn_3610_SprC et Srn_9340 dans le régulon SarA et plus globalement dans les mécanismes généraux de la bactérie.

MATERIELS ET METHODES

MATERIELS ET METHODES

I. Souches bactériennes et conditions de cultures

A. Conditions de cultures

Les souches nécessaires aux expériences ont été listées dans le Tableau 10. Les souches d'*Escherichia coli* ont été cultivées dans le milieu Luria-Bertoni (LB, Mo-Bio) et les milieux BHI (Brain Heart Infusion), TSB (Trypton Soy Broth) et MH (Mueller Hinton) ont été utilisés pour la croissance des souches de *Staphylococcus aureus*. Des antibiotiques ont été ajoutés si nécessaire (Ampicilline : 50 µg/mL ; Erythromycine, Chloramphénicol, Tétracycline : 10 µg/mL ; Kanamycine : 50 µg/mL (*E. coli*), 250 µg/mL (*S. aureus*)).

Les densités bactériennes des cultures sont égalisées le matin à une densité optique ($DO_{600\ nm}$) de 0,1 à partir d'une pré-culture de nuit. La température de croissance est principalement de 37°C.

B. Purification d'ADN plasmidique et Transformations

Les minipréparations sont réalisées grâce au kit QIAprep Spin Miniprep Kit (QIAGEN), suivant les protocoles du fabriquant. Dans le cas des souches de *S. aureus*, une étape préalable est effectuée : les bactéries sont mises en suspension dans le tampon P1 avec 50µg/mL de lysostaphine (Sigma) et incubée à 37°C pendant 10 minutes. Les plasmides obtenus sont listés dans le Tableau 12.

Les souches de *S. aureus* sont mises en culture une nuit en milieu BHI à 37°C. Au matin, elles sont ensuite diluées au centième puis cultivées jusqu'à DO_{600nm} de 1 à 1,5 avant d'être placées sur glace pendant 15 minutes. Les bactéries sont ensuite centrifugées à 4000 g à 4°C pendant 10 mn puis les culots sont successivement lavés par différents volumes de glycérol 10 % puis repris par 500 µL de cette solution. Pour les transformations, 65µL de bactéries compétentes sont mis en présence de 2 µL de plasmide pendant 15 mn à 4°C et électroporées à 2500V. Les bactéries sont ensuite incubées en milieu BHI à 37°C pendant une heure et trente minutes puis étalées sur milieu BHI supplémenté par l'antibiotique adapté.

C. Transduction phagique

A partir d'une pré-culture d'une souche mutée pour un gène (remplacé par un gène de résistance à un antibiotique), 100 µL de bactéries sont mis en contact avec 10 ou 20 µL de phage 80 en présence de CaCl2 (10 mM final) pendant 15 minutes à température ambiante. Ensuite, 3,5 mL de TSB TOP agar (10 mM $CaCl_2$; 10 mM $MgSO_4$; 0,6% Agar) sont mélangés

à la suspension puis étalés sur boîte TSB. Les boîtes sont incubées sur la nuit à 30°C. La gélose correspondant au TOP agar est ajoutée à 4 mL de tampon (10 mM Tris pH 7,5 ; 10 mM MgSO4) et de 100 µL de chloroforme et placée sur la nuit à 4°C. Après centrifugation, le surnageant, correspondant à la solution de phages, est récolté et filtré. Cette solution de phages va être utilisée pour la transduction phagique de la délétion du gène dans la souche receveuse (HG003).

500 µL d'une pré-culture de la souche HG003 sont supplémentés de CaCl2 (5 mM final) et mis en contact avec 50 ou 150 µL de la solution de phages. Après une incubation de 20 mn à 37°C sans agitation, 600µL de milieu BHI/citrate de sodium (100 mM) sont ajoutés puis mis à 37°C pendant 2 heures. Le culot bactérien, récupéré après centrifugation, est étalé sur boîte TSB supplémenté de 10 mM de citrate de sodium et de l'antibiotique utilisé et incubé la nuit à 37°C. Les clones résistants sont ensuite vérifiés pour la délétion par PCR et Northern blot.

II. Etudes des ARN

A. Extraction d'ARN

Les ARN totaux de souches de *S. aureus* ont été extraits à différents point de croissance. Pour cela, le culot bactérien est repris dans du tampon de lyse (0,5% SDS ; 20mM Acétate de Sodium ; 1mM EDTA ; pH 5,5) puis ajouté sur un mélange de billes de verre (Sigma-Aldrich) et de phénol pH 4 (Eurobio). Les cellules sont lysées mécaniquement au broyeur Fast-Prep puissance 6,5 pendant 30 secondes. Après centrifugation à 13000 rpm à 4°C, le surnageant est placé sur un mélange phénol pH 4/Chloroforme/Alcool isoamylique (ratio 25/24/1) avant une nouvelle centrifugation. La phase aqueuse est vortexée avec du Chloroforme/Alcool isoamylique (24/1) puis centrifugée. Les ARN présents dans la phase aqueuse sont précipités sur la nuit avec 0,7 volume d'isopropanol, 0,1 volume d'acétate de sodium. Au matin, après une centrifugation de 30 minutes à 13000 rpm à 4°C, le culot contenant les ARN est lavé avec de l'éthanol 70% puis séché au SpeedVac. Les culots secs sont repris dans 30 µL d'eau dépourvue de RNase.

Pour les expériences de PCRq et de RNA-seq, les ARN totaux sont traités deux fois à la DNase Amp Grade (Invitrogen).

B. Détermination des extrémités : RACE (Rapid Amplification cDNA ends)

Les extrémités des ARN Srn_3610_SprC et Srn_9340 ont été déterminées par RACE à partir d'ARN totaux issu de la souche HG003Δ*sarA*. Après circularisation des ARN par

l'enzyme T4 RNA ligase (Promega), les deux ARN sont sélectionnés par un oligonucléotide spécifique (**Tableau 11**) et la transcription inverse est réalisée par l'enzyme M-MLV (Promega). La séquence d'ADN comprenant les extrémités 5' et 3' est amplifiée par deux réactions consécutives de PCR. Les fragments d'ADN sont insérés dans le plasmide pGemT-Easy (Promega) et séquencés.

C. Northern Blot

Pour analyser les ARN de petites tailles (inférieur à 500 nt), des Northern Blot sur gel de polyacrylamide (acrylamide/bisacrylamide 19/1) ont été réalisés. 10 µg d'ARN totaux sont dénaturés 5 minutes à 65°C en présence du tampon de charge Ambion II puis mis sur un gel de 8% préalablement migré 30 minutes à 150V. Les échantillons migrent pendant 1H30 à 150V dans du tampon TBE 1X. Le transfert est effectué sur une membrane de nitrocellulose (Zeta-Probe GT, BioRad) pendant 2 heures à 25 V dans du TBE 0,5X. Les ARN sont réticulés aux UV (120mJ/cm^2) sur la membrane.

Pour les ARN de grandes tailles, un gel d'agarose est nécessaire pour la réalisation du Northern Blot. Le gel est composé de 1% d'agarose, du tampon MOPS 1X et de formaldéhyde. Le gel est pré-migré 10 minutes à 100 V dans du tampon MOPS 1X. Après dépôt des ARN dénaturés dans le tampon de dénaturation (Formamide 50%, Formaldéhyde 2,2M et tampon MOPS 1X), l'électrophorèse est réalisée à 25V sur la nuit. Le transfert est effectué pendant 1H30 sous une pression de 7mm Hg dans du tampon SSC 10X sur une membrane de Nytran précédemment lavée au SSC 10X.

La révélation des ARN d'intérêt est faîte par des oligonucléotides préalablement marqués à la digoxigénine (Roche) (sondes listées dans le **Tableau 11**). Après dénaturation des sondes à 65°C, l'hybridation est effectuée sur la nuit à 37°C dans du DIG Express Hyb (Roche). En respectant les différents lavages préconisés par la fabriquant, un anticorps anti-digoxigénine couplé à la phosphatase alcaline est ajouté. L'ajout du substrat (CSPD) entraîne la réaction enzymatique et la membrane est placée à 37°C pendant 15 minutes. L'acquisition du signal est faite grâce à l'ImageQuant LAS 4000 (GE Healthcare).

D. RT-PCRq

Les ARN totaux sont déplétés de l'ADN contaminant par deux traitements à la DNase Amp Grade (Invitrogen). La reverse-transcription est réalisée par le kit High-Capacity cDNA Reverse Transcription (Life Technologies) en suivant le protocole du fournisseur.

L'expérience de PCR quantitative est effectuée par ajout, aux ARN, de RealMasterMix SYBR Kit (Life Technologies) et des amorces spécifiques aux gènes étudiés. Les échantillons

sont analysés dans le StepOnePlus Real-Time PCR system (GE Healthcare). Nous avons utilisé la méthode de comparaison ΔΔCt en utilisant comme référence la quantité d'ARN mesuré pour le gène de ménage HU.

E. Transcription *in vitro*

Les séquences *sprC*$_{456}$ et *srn_9340*$_{337}$ sont amplifiées par PCR à partir d'ADN génomique de la souche HG003 en utilisant des amorces spécifiques (**Tableau 11**). Ces séquences correspondent aux régions promotrice, génique et terminatrice de *srn_3610_sprC* et *srn_9340.* La transcription *in vitro* est réalisée avec 100 fmol d'holoenzyme ARN polymérase d'*E. coli* (New England Biolabs), 10 fmol de sonde, 0,75mM de nucléosides triphosphates (ATP, CTP et GTP), 0,5 µCi de [α-^{32}P]UTP, le tout placé dans tampon 5X associé à l'enzyme. Quand nécessaire, 5 pmol de SarA purifiée sont ajoutés à la réaction. Les échantillons sont ensuite déposés sur gel dénaturant de polyacrylamide. La détection est faite avec le Typhoon FLA 9500 (GE Healthcare).

III. Etudes des Protéines

A. Extraction de protéines

Les protéines totales de cultures de *S. aureus* sont extraites à partir des cellules centrifugées et reprises dans le tampon de lyse (50mM Tris-HCl, $pH_{7,5}$; 3mM $MgCl_2$; 0,1mg/mL lysostaphine). Les cellules sont incubées 20 minutes à 37°C puis mises sur glace. Les cellules sont broyées, au FastPrep (MP Bio) puissance 6,5 pendant 30 secondes, par des billes de verres préalablement traitées à l'acide chlorhydrique. Le surnageant est prélevé et dosé par le système fluorimétrique Qubit.

B. Western Blot

Des échantillons protéiques, en même quantité, sont déposés et migrés dans un gel dénaturant de polyacrylamide (SDS-PAGE). Après migration à 120V dans du tampon Tris-Glycine-SDS, les protéines sont transférées sur membranes Hybond P PVDF (Amersham). Les membranes sont ensuite lavées puis un anticorps primaire polyclonal anti-SarA de lapin est ajouté. L'anticorps secondaire anti-lapin marqué à la peroxidase de raifort (HRP) est placé sur la membrane. La détection des protéines est réalisée par le kit de détection ECL Prime Western Blotting (Amersham) et observée par le logiciel ImageQuant LAS 4000. Pour s'assurer du dépôt d'une même quantité de protéines totales pour chaque échantillon, un duplicata des gels est réalisé et les protéines sont révélées à l'ajout de SYPRO Ruby (Bio-Rad).

C. Expérience de gènes rapporteurs : ß-lactamase

Les cultures de nuit des souches de *S. aureus* contenant le plasmide pCN41c (+/- les promoteurs d'intérêt) sont ré-ensemmencées dans le milieu de culture à une densité optique (600 nm) de 0,1. Pour chaque temps de prélèvement, les cellules sont centrifugées et remises en suspension dans un volume de tampon PBS 1X (NaCl 137mM ; KCl 2,7mM ; Na_2HPO_4 8,1mM ; KH_2PO_4 1,8mM, pH 7,4) variable ajusté par rapport à la DO de la culture et ce afin d'obtenir une quantité de cellules équivalentes pour chaque échantillon. La lyse des cellules est effectuée à 37°C pendant 20 minutes par ajout de 0,7mg/mL de lysostaphine, 0,2U/µL de benzonase et 0,1mM de $MgCl_2$. Après ajout de nitrocéfine (0,25mg/mL), l'activité de la ß-lactamase est quantifiée par mesure colorimétrique à une longueur d'onde de 492 nm à 37°C toutes les 10 minutes pendant 40 minutes (Biotek, logiciel Synergy2). L'activité est normalisée à la quantité de protéines totales dosées par dosage de Bradford.

D. Expérience de gènes rapporteurs : Green Fluorescent Protein (GFP)

Le protocole des expériences de gène rapporteur *gfp* se base sur le protocole suivant. Les modifications faites pour les stress appliqués sont présentées directement dans la partie résultats. Des cultures de nuit de souches de *S. aureus* ayant reçu le plasmide pCN52c sont égalisées à une $DO_{600\ nm}$ de 0,1. Les cultures sont ensuite réparties en plaque 96 puits puis mises dans un spectromètre thermostaté à 37°C (Biotek, logiciel Synergy2). Quand nécessaire, après 3 heures de croissance, l'ajout d'un stress (par différents produits) est appliqué. Aux différents points de mesure, la densité optique à 600 nm permet d'estimer la quantité de cellules présentes d'échantillons. La fluorescence est mesurée à la longueur d'onde de 509 nm après excitation à 488 nm.

IV. Etudes de l'interaction Protéines-ADN

A. Production de SarA purifiée

La séquence codante de *sarA* a été placée dans le plasmide pET42a afin de placer, dans le cadre de lecture, une étiquette de 6 acides aminés histidines à l'extrémité 5' de la phase codante. Le plasmide est introduit dans la souche *E. coli* BL21. Après 3 heures de culture, l'expression de *sarA* est induite pendant 3 heures par ajout d'1 mM IPTG (isopropyl ß-D-1-thiogalactopyranoside). Les cellules sont centrifugées, lavées et resuspendues dans le tampon de lyse (10mM HEPES, $pH_{7,5}$; 500mM NaCl). La purification de la protéine SarA étiquetée par 6 histidines est réalisée grâce à des colonnes de nickel (HisTrap HP ; GE Healthcare) en utilisant le système de chromatographie liquide haute performance ÄKTA

(Amersham Biosciences). ®). Après équilibration de la colonne par les tampons A (HEPES 10mM $pH_{7,5}$, NaCl 1M, Glycérol 5%, Imidazole 15mM) et B (HEPES 10mM $pH_{7,5}$, NaCl 100mM, Glycérol 5%, Imidazole 500mM), le lysat injecté au niveau de la colonne est lavé avec le tampon A. Pour effectuer l'élution, un gradient d'imidazole est réalisé en ajustant le débit du tampon B au sein du tampon A. L'élution est suivie par ordinateur pour identifier les tubes correspondant au pic d'élution. L'imidazole est éliminé de l'éluât (par action d'un solution d'HEPES 10mM pH 7,5, NaCl 200mM, Glycérol 10%, $MgCl_2$ 5mM, ß-mercaptoéthanol 2mM) et est concentré grâce à un centricon de 5kDa. La pureté et la masse moléculaire de la protéine a été vérifiée par SDS-PAGE.

B. Retards sur gel

1. Préparation des sondes

Les sondes ADN ont été amplifiées à partir du génome de la souche HG003 par des amorces spécifiques afin de créer les sondes *sprC*$_{P267}$, 16S$_{267}$, *srn_9340*$_{P225}$ et 16S$_{225}$. Pour les sondes de petites tailles, une hybridation de deux oligonucléotides complémentaires permet la production des sondes *sprC*$_{47}$, *srn_9340*$_{51}$, SprC$_{30Gauche}$, SprC$_{30Milieu}$, SprC$_{30Droite}$, Srn_9340$_{30Gauche}$, Srn_9340$_{30Milieu}$ et Srn_9340$_{30Droite}$. Enfin, pour les sondes délétées d'une séquence nucléotidique, trois étapes de PCR sont nécessaires comme décrit en Figure 114. Ainsi, nous avons créé les sondes *sprC*$_{P267\Delta47}$, *sprC*$_{P267\Delta22}$, *srn_9340*$_{P225\Delta51}$ et *srn_9340*$_{P225\Delta23}$.

2. Réaction de fixation

Le protocole utilisé pour effectuer les retards sur gel est dérivé de celui décrit par Carey *et al.* (Carey et al., 2013).

Les sondes ADN sont marquées avec le nucléotide radioactif [γ-^{32}P] adénosine triphosphate (ATP) en utilisant l'enzyme T4 polynucléotide kinase (T4PNK). 10 fmol de sondes radiomarquées sont placés dans du tampon de réaction (20mM HEPES-KOH, $pH_{7,9}$; 20% glycérol ; 0,2mM EDTA ; 0,1M KCl ; 0,1 $MgCl_2$; 5mM DTT ; 0,2µg poly(dIdC)). Quand il est nécessaire, des quantités de la protéine SarA purifiée ou de l'holoenzyme ARN polymérase d'*E. coli* (New England Biolabs) sont ajoutées. La réaction de fixation est placée 30 minutes à 37°C avant d'être déposée sur gel natif de polyacrylamide. La détection s'effectue par analyse au Typhoon FLA 9500 (GE Healthcare).

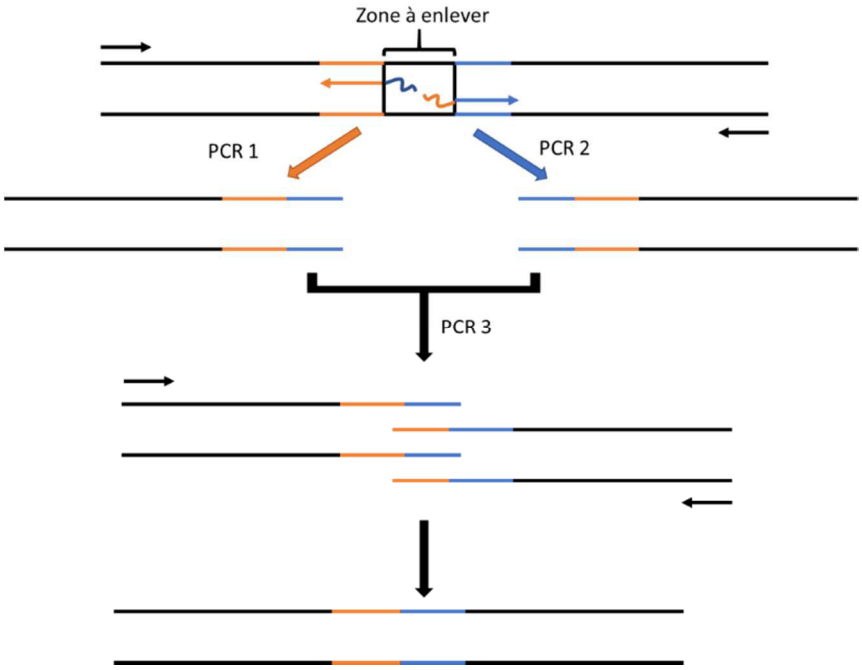


Figure 114 : Description des trois étapes de PCR pour la délétion au sein des sondes

C. Empreinte à la DNase

1. Préparation des sondes

Les sondes *sprC*$_{P267}$ et *srn_9340*$_{P225}$ utilisées pour les expériences d'empreintes à la DNase sont radiomarquées à une seule extrémité. Pour cela, nous digérons ces deux sondes avec l'enzyme HindIII donc un site de restriction est présent pour *sprC*$_{P267}$ ou préalablement ajouté lors de la PCR pour *srn_9340*$_{P225}$. L'extrémité cohésive, ainsi produite, est comblée par les nucléotides GTP, CTP et TTP et par le nucléotide radioactif [α-^{32}P]ATP en utilisant le fragment Klenow (New England Biolabs).

2. Réaction de fixation

La réaction de fixation est exactement la même que celle utilisée pour les expériences de retard sur gel. Néanmoins avant le dépôt sur gel, les réactions subissent une digestion par la DNase I ($2x10^{-2}$ U pour *sprC*$_{P267}$ et $7,5x10^{-2}$U pour *srn_9340*$_{P225}$) pendant 1 minute à 37°C avec la tampon correspondant (40mM $CaCl_2$; HEPES 20mM pH 7,6 ; Glycérol 20% ; EDTA 0,2 mM ; KCl 0,1M ; BSA 10µg/µL ; ß-mercaptoéthanol 28mM). La réaction est arrêtée par ajout du tampon Stop (10µg/mL protéinase K ; 400mM acétate de sodium ; 0,2% SDS ; 10mM EDTA et 50µg/mL ARNt de levure). Après 15 minutes à 55°C (action de la protéinase K), les fragments d'ADN sont purifiés par extraction au phénol/chloroforme et précipités à l'éthanol. Les culots récupérés sont suspendus dans une solution de bleu de charge (Formamide Dye Mix) et dénaturés 5 minutes à 95°C. Les échantillons sont déposés sur gel dénaturant de polyacrylamide.

D. Immunoprécipitation de la chromatine : ChIP

Les souches bactériennes HG003Δ*sarA* contenant soit le plasmide pCN38 vide soit pCN38-SarA6His sont mises en culture pendant 2H30 sous agitation (160rpm) à 37°C. Un traitement de 30 minutes au formaldéhyde (1% final) est effectué pour fixer de manière covalente les protéines sur le chromosome bactérien. Les bactéries sont mises sous agitation dans un bain glacé pendant 30 minutes en présence de glycine (30mM). Le culot bactérien obtenu après centrifugation est lavé 3 fois au PBS 1X puis stocké à -80°C. La lyse de la paroi bactérienne est réalisée à 37°C dans 300 µL de tampon de lyse (200 mM Tris pH 8, 600mM NaCl, 4% Triton X-100, 1 mM PMSF et 50µg de lysostaphine). 5 cycles de sonication sont appliqués dans un bain glacé. Un premier traitement à la RNase A (0,5 µg) est effectué pendant une heure à 4°C. En parallèle, des billes de sépharose couplées à la protéine G sont préalablement lavées sur la nuit à 4°C avec du tampon de dilution (0,01 % SDS, 1,1% Triton X-100, 1,2mM EDTA, 16,7mM Tris pH 8, 167mM NaCl) ainsi qu'avec 1% de BSA et 0,1mg/mL de sperme de saumon soniqué. Le surnageant obtenu après traitement à la RNase est mélangé pendant 30 minutes à 4°C aux billes lavées pour éliminer les interactions non-spécifiques. Le surnageant clarifié est divisé en différentes fractions : les fractions « ChIP », « DIG » et « Pol » auxquelles sont ajoutés respectivement 8µg d'anticorps anti-6Histidine, anti-Digoxigénine et anti-σ^{70}. Une fraction « INPUT » est prélevée pour le contrôle positif interne. Les fractions et les anticorps sont mélangés pendant 5 heures à 4°C. Puis 16µL de billes magnétiques couplées à la protéine G sont ajoutées et le tout est remis en rotation sur la nuit à 4°C.

Les billes sont lavées à 4°C par différents tampons (2x Dilution Buffer, 1x Low Salt Immune Complexe Wash Buffer, 2x High Salt Immune Complexe Wash Buffer, LiCl Immune Complexe Wash Buffer et 2x TE Buffer) et l'élution est effectuée avec le tampon d'élution pendant 15 minutes à température ambiante puis 0,2M de NaCl sont ajouté avant de mettre à 65°C sur la nuit. La fraction « INPUT » est directement mise à 65°C avec le tampon d'élution et le NaCl.

Les différentes fractions sont traitées une seconde fois à la RNase A (20µg) 20 minutes à 37°C puis à la Protéinase K (20µg) 1 heure à 45°C.

Afin de normaliser les pertes d'ADN, 2pg de plasmide (pYAR144) sont ajoutés à toutes les fractions puis l'ADN est extrait au phénol-chloroforme et précipité sur la nuit à -20°C avec 2 volumes d'éthanol 95%, 0,1 volume de NaCl (5M) et 20µg de glycogène. Le culot d'ADN est lavé avec de l'éthanol 70% puis séché sous vide. L'ADN est resuspendu dans 15 µL d'eau dépourvue de RNase et DNase. La quantité des fragments d'ADN est mesurée par PCRq (voir section RT-PCRq).

Tableau 10 : Liste des souches utilisées pour l'étude

	Souches	Commentaires	Références
Souche d' ***E. coli***	XL1-Blue	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)	
Souches de ***S. aureus***	8325-4	Souche dérivée de 8325 (ΔΦ11, Φ12 et Φ13)	Novick, 1967
	RN4220	Souche dérivée de 8325-4 avec le système de restriction défectueux	Kreiswirth et *al.*, 1983
	HG003	Souche 8325 avec les gènes *rsbU* et *tcaR* restaurés	Herbert et *al.*, 2010
	HG003 ΔsarA	Souche HG003 mutée pour *sarA* (résistance à l'érythromycine)	Laboratoire
	RN6390 ΔsarR	Souche RN6390 mutée pour *sarR* (résistance à l'érythromycine)	Manna *et al*, 2001
	HG003 ΔsarR	Souche HG003 mutée pour *sarR* (résistance à l'érythromycine)	Cette étude
	8325-4 ΔsarS	Souche 8325-4 mutée pour *sarS* (résistance à l'érythromycine)	Oscarsson *et al.*, 2005
	HG003 ΔsarS	Souche HG003 mutée pour *sarS* (résistance à l'érythromycine)	Cette étude
	RN6390 ΔsarT	Souche RN6390 mutée pour *sarT* (résistance à l'érythromycine)	Schmidt *et al.*, 2001
	HG003 ΔsarT	Souche HG003 mutée pour *sarT* (résistance à l'érythromycine)	Cette étude
	RN6390 ΔsarV	Souche RN6390 mutée pour *sarV* (résistance à l'érythromycine)	Manna *et al.*, 2004
	HG003 ΔsarV	Souche HG003 mutée pour *sarV* (résistance à l'érythromycine)	Cette étude
	RN6390 ΔsarX	Souche RN6390 mutée pour *sarX* (résistance à l'érythromycine)	Manna *et al.*, 2006
	HG003 ΔsarX	Souche HG003 mutée pour *sarX* (résistance à l'érythromycine)	Cette étude
	HG003 ΔarnIII	Souche HG003 mutée pour *arnIII* (résistance à la kanamycine)	Laboratoire
	HG003 ΔsarAΔsarR	Souche HG003 mutée pour *sarA* et *sarR* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude

	HG003 ΔsarAΔsarS	Souche HG003 mutée pour *sarA* et *sarS* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude
	HG003 ΔsarAΔsarT	Souche HG003 mutée pour *sarA* et *sarT* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude
	HG003 ΔsarAΔsarV	Souche HG003 mutée pour *sarA* et *sarV* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude
	HG003 ΔsarAΔsarX	Souche HG003 mutée pour *sarA* et *sarX* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude
	HG003 ΔsarAΔarnIII	Souche HG003 mutée pour *sarA* et *arnIII* (résistances respectives à l'érythromycine et la kanamycine)	Cette étude
	HG003 ΔsarAΔrot	Souche HG003 mutée pour *sarA* et *rot* (résistances respectives à la kanamycine et l'érythromycine)	Cette étude
	HG003 ΔsigB	Souche HG003 mutée pour *sigB* (résistance à l'érythromycine)	Cette étude
	8325-4 Δrot	Souche 8325-4 mutée pour *rot* (résistance à l'érythromycine)	Cette étude

Tableau 11 : Liste des oligonucléotides utilisés pour l'étude

Oligonucléotides	Séquences (5'→ 3')	Expériences
P_SprC_EcoRI_3'	GGCCGAATTCATTTATATTATAATATAAATATT	Clonage
P_SprC_144_BamHI_5'	GGCCGGATCCAAGTGTATAATATCAATTTACTAC	Clonage
SprC_EcoRI_3'	GGGGGAATTCGTATACTGTTATAACTGAAATG	Clonage
SprC_KpnI_5'	GGGGGGTACCAGTCAACGACCATGCGTGCAC	Clonage
Gene-sprC_EcoRI-3'	GGCCGAATTCCGTTTTACTGTCCACGCATG	Clonage
SprC_3'_FP	ATCGCTTACTTCATCTAAAA	Empreinte à la DNase
SprC_Sequençage_FP	AGCTTCTACTCTCATGGCAATTTATAC	Empreinte à la DNase
Footprint_Star_3_seq_v2	AGCTTACAGTAGTCACCTGTTCGCA	Empreinte à la DNase
FP_5_Star225_Hind	GGCCAAGCTTAAATTCTTTAATCAAATGTG	Empreinte à la DNase
FP_5_Star225_seq	AGCTTAAATTCTTTAATCAAATGTG	Empreinte à la DNase
Srn_3610_SprC_NB	CGGCTACTACATTCGCATGT	Northern Blot
Srn_9340_NB	CGACTAACTCCTGGTTATGC	Northern Blot
SarA_NB	AGCTAATTGAAAATTCCTTTTTAATTAAACTTTTTAATTTGTCAGCATAAGTGACCATTGATAACAACTCAAAGCAATCATTGATTTTTGTAATTGCCAT	Northern Blot
ARNtm_NB	ACACGCTTAATGAGCTCGGG	Northern Blot
ARN_16S_NB	TTCATCATATAACAAATGTAATGAGATATGATGTTGGTAGTAAAATTATTTTAACTTTTAATGATCAAACAGTTTATTTATGTGAAATTGCTGACATAAG	Northern Blot
SprC_R1_RACE	CTACATTCGCATGTACTTCG	RACE
SprC_R2_RACE	ACTTCGTTTTACTGTCCACGC	RACE
SprC_F1_RACE	GTATAAATTGCCATGAGAGTAG	RACE
SprC_F2_RACE	GCTTCAGTTTATGAAGACTTC	RACE
SprC_267Δ47_for_FP	ATTCGCATGTACTTCGTTTTACTGTCCACGAATACTTAATTTTAAAAAATCC	Retard sur gel

SprC_267Δ47_sens_FP	TTTCATAGGGGATTTTAAAATTAAGTATTCGTGGACAGTAAAACGAAGTAC	Retard sur gel
SprC_267Δ22_rev	GTCCACGCATGGTCGTAAATATTTTATTTTTCAATAC	Retard sur gel
SprC_267Δ22_for	GAAAAATAAAATATTTACGACCATGCGTGGACAGTAAAACG	Retard sur gel
SprC_30_milieu_for	TTGACTTATTTATATTATAATATAAATATT	Retard sur gel
SprC_30_milieu_rev	AATATTTATATTATAATATAAATAAGTCAA	Retard sur gel
SprC_30_gauche_for	GAAAAATAAAATATTTATATTATAATATAA	Retard sur gel
SprC_30_gauche_rev	TTATATTATAATATAAATATTTTATTTTTC	Retard sur gel
SprC_30_droite_for	TATTATAATATAAATAAGTCAACGACCATG	Retard sur gel
SprC_30_droite_rev	CATGGTCGTTGACTTATTTATATTATAATA	Retard sur gel
Star_30_gauche_sens	AATATTATATATACGCTTGAAAAAATAAA	Retard sur gel
Star_30_gauche_antisens	TTTTATTTTTTCAAGCGTATATATAATATT	Retard sur gel
Star_30_droite_sens	AAAATAAAATAATAATATTATAATATAATC	Retard sur gel
Star_30_droite_antisens	GATTATATTATAATATTATTATTTTATTTT	Retard sur gel
Star_30_milieu_sens	ATACGCTTGAAAAAATAAAATAATAATATT	Retard sur gel
Star_30_milieu_antisens	AATATTATTATTTTATTTTTTCAAGCGTAT	Retard sur gel
Star225Δ23_Sens_droite	CAAATATTATATATACGCTATAATATAATCAATCGGCATCTACG	Retard sur gel
Star225Δ23_antiSens_gauche	GCCGATTGATTATATTATAGCGTATATATAATATTTGATTTTC	Retard sur gel
qPCR_Srn_9340_sens	GCATCTACGTGCGGATTATAAAGTAAAG	RT-PCRq
qPCR_Srn_9340_rev	TTACTTTAGCCTTTATTAATAATTTAGTCG	RT-PCRq
qPCR_SarA_sens	GACATACATCAGCGAAAACAAAGAG	RT-PCRq
qPCR_SarA_rev	CGTTGTTTGCTTCAGTGATTCGT	RT-PCRq
qPCR_sarV_sens	ACCGCTAATCTTTTCAGCTTCA	RT-PCRq
qPCR_sarV_antisens	GCAGAAAGGGAGTTAAGTCAGT	RT-PCRq
qPCR_citratetransporter_sens	TAAGGCTGGTGAAAATGGGC	RT-PCRq

qPCR_citratetransporter_antisens	AAACAGCAGGGCAATTTGGT	RT-PCRq
qPCR_AtlR_sens1	GGACAACGCACGGATATCAC	RT-PCRq
qPCR_AtlR_antisens1	TCCCTGCTAATCTTCATCGTCA	RT-PCRq
qPCR_Staphopain_Sens	GTTGGTGTGCAGGCTATACG	RT-PCRq
qPCR_Staphopain_Reverse	CGTTGTCCTTGTAAGTTTGGGT	RT-PCRq
qPCR_srn_4940_sens	GACTCTGAGGTGATACGAATGC	RT-PCRq
qPCR_srn_4940_reverse	AAGTCAAACAACATGGTCATCC	RT-PCRq
qPCR_srn_1490_sens	TCTTAGAGTATTTTGTCCAACGCA	RT-PCRq
qPCR_srn_1490_reverse	TCGCAATCGCAATAACTGACA	RT-PCRq
qPCR_SprC_sens	AAGCTTCTACTCTCATGGCAATTT	RT-PCRq
qPCR_SprC_rev	TGCGTGGACAGTAAAACGAA	RT-PCRq
pCN_for_seq	GGCCTTTTGCTCACATGTTC	Séquençage
pCN_rev_seq	GTGCTGCAAGGCGATTAAGT	Séquençage
MCS_KpnI_Pveg	TCTAGAGGATCCCCGGTACCAGAGTTTGACAAAAATGGCTCGTGTTGTACAATAAATGTA	Surexpression
Gibson_3'Star	TAAAACGACGGCCAGTGAATTCTCAAATATCATTGAAACTTT	Surexpression
Gibson_3'SprC	TAAAACGACGGCCAGTGAATTCCGTATACTGTTATAACTGAA	Surexpression
Pveg_5'Star	GTTGTACAATAAATGTAAATCGGCATCTACGTGCGG	Surexpression
Pveg_5'SprC	GTTGTACAATAAATGTAAAGTCAACGACCATGCGTGGAC	Surexpression
qPCR_Hu_rev	CCTCAAAGTTACCGAAACCAA	RT-PCRq
qPCR_HU_sens	AGCTGGTTCAGCAGTAGATGC	RT-PCRq
SprC_Delta47_rev	CATGGTCGTTGACTTATTTATATTATAATATAAATATTTTATTTTTC	Retard sur gel
SprC_Delta47_sens	GAAAAATAAAATATTTATATTATAATATAAATAAGTCAACGACCATG	Retard sur gel

Tableau 12 : Liste des plasmides utilisés pour l'étude

Plasmides	**Commentaires**	**Références**
pCN36	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et à la tétracycline dans *S. aureus*)	Charpentier et al., 2004
pCN36-SarA	pCN36 avec le gène *sarA* sous le contrôle de son promoteur endogène	This study
pCN36-PvegSrn_9340	pCN36 avec le gène *srn_9340* sous le contrôle du promoteur Pveg	This study
pCN38	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et à l'érythromycine dans *S. aureus*)	Charpentier et al., 2004
pCN38-SarA6His	pCN38 avec le gène *sarA* étiqueté avec 6 histidines sous le contrôle de son promoteur endogène	This study
pCN38-PvegSprC	pCN38 avec le gène *srn_3610_sprC* sous le contrôle du promoteur Pveg	This study
pCN41	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et à l'érythromycine dans *S. aureus*) avec le gène *blaZ*	Charpentier et al., 2004
pCN41c	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *blaZ*	This study
pCN41c-PsprC	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *blaZ* sous contrôle du promoteur de *srn_3610_sprC*	This study

pCN41c-Psrn_9340	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *blaZ* sous contrôle du promoteur de *srn_9340*	This study
pCN52	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et à l'érythromycine dans *S. aureus*) avec le gène *gfp*	Charpentier et al., 2004
pCN52c	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *gfp*	This study
pCN52c-PsprC	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *gfp* sous contrôle du promoteur de *srn_3610_sprC*	This study
pCN52-Psrn_9340	Plasmide faible copie (résistance à l'ampicilline dans *E. coli* et au chlormaphénicol dans *S. aureus*) avec le gène *gfp* sous contrôle du promoteur de *srn_9340*	This study
pET42a	Vecteur d'expression d'E*. coli*	Novagen
pET42a-SarA	pET42a avec le gène *sarA*	Hallier M, laboratory

ANNEXES

ANNEXE 1 :

SRD: a Staphylococcus regulatory RNA database

Sassi, M., Augagneur, Y., Mauro, T., Ivain, L., Chabelskaya, S., Hallier, M., Sallou, O. & Felden, B.

RNA, (2015), 21(5), 1005-1017

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

SRD: a *Staphylococcus* regulatory RNA database

MOHAMED SASSI,[1,3] YOANN AUGAGNEUR,[1,3] TONY MAURO,[1] LORRAINE IVAIN,[1] SVETLANA CHABELSKAYA,[1] MARC HALLIER,[1] OLIVIER SALLOU,[2] and BRICE FELDEN[1]

[1]Inserm U835 Biochimie Pharmaceutique, Rennes University, 35043 Rennes, France
[2]Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes University, 35043 Rennes, France

ABSTRACT

An overflow of regulatory RNAs (sRNAs) was identified in a wide range of bacteria. We designed and implemented a new resource for the hundreds of sRNAs identified in Staphylococci, with primary focus on the human pathogen *Staphylococcus aureus*. The "Staphylococcal Regulatory RNA Database" (SRD, http://srd.genouest.org/) compiled all published data in a single interface including genetic locations, sequences and other features. SRD proposes novel and simplified identifiers for Staphylococcal regulatory RNAs (srn) based on the sRNA's genetic location in *S. aureus* strain N315 which served as a reference. From a set of 894 sequences and after an in-depth cleaning, SRD provides a list of 575 *srn* exempt of redundant sequences. For each sRNA, their experimental support(s) is provided, allowing the user to individually assess their validity and significance. RNA-seq analysis performed on strains N315, NCTC8325, and Newman allowed us to provide further details, upgrade the initial annotation, and identified 159 RNA-seq independent transcribed sRNAs. The lists of 575 and 159 sRNAs sequences were used to predict the number and location of *srns* in 18 *S. aureus* strains and 10 other Staphylococci. A comparison of the *srn* contents within 32 Staphylococcal genomes revealed a poor conservation between species. In addition, sRNA structure predictions obtained with MFold are accessible. A BLAST server and the intaRNA program, which is dedicated to target prediction, were implemented. SRD is the first sRNA database centered on a genus; it is a user-friendly and scalable device with the possibility to submit new sequences that should spread in the literature.

Keywords: small regulatory RNAs; bacteria; *Staphylococcus aureus*; database; RNA-seq; sRNA identification; sRNA targets

INTRODUCTION

In the recent years, a plethora of regulatory RNAs (sRNAs) were identified in diverse bacterial genomes, including several human pathogens. sRNAs enable bacteria to induce efficient and prompt physiological feedbacks to adjust on their environments, and also to establish infection (Caldelari et al. 2013). Mechanistically, sRNAs intervene on transcription, mRNA turnover, and/or translation of target genes. sRNAs proceed in gene expression regulations, from transcription initiation to translation control and protein activity (Storz et al. 2011). The majority of sRNAs characterized up to now act by pairings with target mRNAs, either encoded on the opposite strand (*cis*-encoded) or transcribed apart from their targets (*trans*-encoded). Some have been shown to encode small peptides. While most of the bacterial sRNAs were originally studied in *E. coli* and other Gram-negative bacteria (Mizuno et al. 1984), a recent outburst of sRNAs was identified in Gram-positive bacteria (Brantl and Brückner 2014), including the major human pathogen *Staphylococcus aureus* (Fechter et al. 2014).

Staphylococcus aureus is an opportunistic pathogen that has sophisticated regulatory tracks to rapidly and efficiently adapt its growth in response to its disparate habitats and hosts. Several groups have shown experimentally that *S. aureus* express many sRNAs, delivered from the core genome, mobile and accessory elements (Guillet et al. 2013; Tomasini et al. 2014). They include several predicted riboswitches (*cis*-acting regulatory mRNA leader sequences), many *cis*-encoded antisense RNAs, several *trans*-encoded sRNAs (Romilly et al. 2012) with some containing small open reading frames that were shown to be expressed (Sayed et al. 2012; Pinel-Marie et al. 2014). However, and for the most part, their functions and mechanisms are unexplored yet. For the few sRNAs with associated functions, some detect bacterial density, modify cell surface properties for host immune escape, adjust central metabolism for optimal growth, regulate the expression of virulence factors, influence antibiotic resistance (Lalaouna

[3]**These authors contributed equally to this work.**
Corresponding authors: brice.felden@univ-rennes1.fr, yoann.augagneur@univ-rennes1.fr
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.049346.114.

© 2015 Sassi et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

et al. 2014), trigger cell death, and encodes toxins (Felden et al. 2011; Guillet et al. 2013).

The systematic and recent use of high-throughput RNA-sequencing technologies substantially raised the number of sRNAs sequences identified per bacterial species. To cope with that plethora of sRNAs, several databases have emerged from generalists to more specialize. Generalist databases such as fRNAdb (Kin et al. 2007) or NONCODE (Xie et al. 2014) focus exclusively on eukaryotes, while Rfam 11.0 (Burge et al. 2013) partitions RNAs data into families for both the eukaryotes and prokaryotes. RNAdb was originally designed for mammalian noncoding RNAs but was officially retired in June 2012 (Pang et al. 2007). On the other hand, there are databases specifically devoted to the bacterial kingdom. sRNAMap is a repository for the microbial genomes but Gram-positive bacteria, and therefore Staphylococci, are absent from this browser (Huang et al. 2009). sRNATarbase (Cao et al. 2010) was implemented to provide a list of sRNA targets, but is out of date for Staphylococci. Recently, two bacterial sRNA databases were developed: (i) sRNAdb, which focused exclusively on Gram-positive bacteria (Pischimarov et al. 2012) but provides only 39 sRNA sequences for *S. aureus*; (ii) BSRD is a generalist bacterial sRNA database with >700 species included (Li et al. 2013). Finally, RNAspace.org is a platform devoted to the prediction, annotation, and analysis of noncoding RNA but does not provide data set (Cros et al. 2011). In *S. aureus* and more generally in the Staphylococcal genus, a unified sRNA nomenclature is lacking, while many redundancies, as single sequence described under several IDs, and potential misannotated sRNAs (e.g., repeated sequences, mRNA leader or trailer sequences) would require an in-depth manual cleaning.

Therefore, there is an urgent need for additional sRNA databases focusing on a bacterial genus to provide an accurate and simple list of sRNAs. Here, we report a *Staphylococcus* Regulatory RNA Database (SRD, http://srd.genouest.org/) which compiles, after an in-depth scrubbing all the sRNA sequences identified so far, with a primary focus on the human pathogen *S. aureus* as a reference. Starting from a large set of sRNA sequences, SRD proposes a new and simple nomenclature together with individual functional, structural, and phylogenetic information and predictions. It provides a unified repository based on additional RNA-seq data analysis.

RESULTS

Construction of a database encompassing the Staphylococcal regulatory RNAs

Staphylococcal sRNAs were identified and studied principally in several strains of *S. aureus* (Tomasini et al. 2014). The chronological discovery of the Staphylococcal sRNAs expressed in *S. aureus* is listed in Table 1. Those RNAs were identified by combining diverse experimental and bioinformatics approaches (Novick et al. 1989, 1993; Pichon and Felden 2005; Anderson et al. 2006; Roberts et al. 2006; Marchais et al. 2009; Nielsen et al. 20011; Morrison et al. 2012; Xue et al. 2014) including the use of Next-Generation RNA-Sequencing technologies (Geissmann et al. 2009; Abu-Qatouseh et al. 2010; Beaume et al. 2010; Bohn et al. 2010; Howden et al. 2013). A total of 894 sequences transcribed as sRNA were compiled from the literature (Fig. 1; Supplemental Data S1). We then focused on the following extensively studied and completed *S. aureus* genomes: N315, Newman, NCTC8325, and JKD6008 (Table 2). The BLAST program was used to locate the coordinates of each sRNA gene in any of the four genomes. Some sequences appeared, as previously suggested (Beaume et al. 2010; Howden et al. 2013), to be repeated onto the genomes, that led to an increase in the total number of sRNA sequences collected. Therefore, sequences identified as DNA repeated sequences by these authors were removed after confirming the initial statements using Blast (Supplemental Data S2). In addition, sequences located in CDSs, rRNAs, tRNAs, or spacers within the four genomes as well as the RNA sequences flanking the genes transcribed as ribosomes (reads overlapping with the ribosomes or within the intergenic regions of ribosomes) were discarded (Liu et al. 2009) to generate a first data set of 773 sequences. A significant number of redundant sequences annotated as a single sRNA could be retrieved under other names. This data set included, among others, the *sau*, *rsa*, *jkdsRNA*, *teg*, and *spr* genes. As an example, up to five other different gene IDs were identified for *rsaE* (*RsaON_Sau20_Teg92_IGR6_sRNA183*). Therefore, we manually cured this data set to provide a list of 575 sRNA genes exempted of redundancy.

TABLE 1. Sequential identification of regulatory RNAs expressed in *Staphylococcus aureus*

sRNA genes	Year	References	Main strains
RNAI	1989	Novick et al. (1989)	NCTC8325
RNAIII	1993	Novick et al. (1993)	NCTC8325
spr genes	2005	Pichon and Felden. (2005)	N315
WAN genes	2006	Roberts et al. (2006)	N315
ssr genes	2006	Anderson et al. (2006)	N315
rsaO genes	2009	Marchais et al. (2009)	N315
rsa genes	2009	Geissmann et al. (2009)	N315, COL
sau genes	2010	Abu-Qatouseh et al. (2010)	N315
rsa genes	2010	Bohn et al. (2010)	N315
teg genes	2010	Beaume et al. (2010)	N315
sbr genes	2011	Nielsen et al. (2011)	Newman
SSR42	2012	Morrison et al. (2012)	UAMS-1
artR	2014	Xue et al. (2014)	NCTC8325
jkd sRNA genes	2013	Howden et al. (2013)	JKD6008

Need and proposal for a novel and simplified identifier

The recent outburst in sRNAs led to spreading a large confusion in the actual number of sRNA genes and for

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

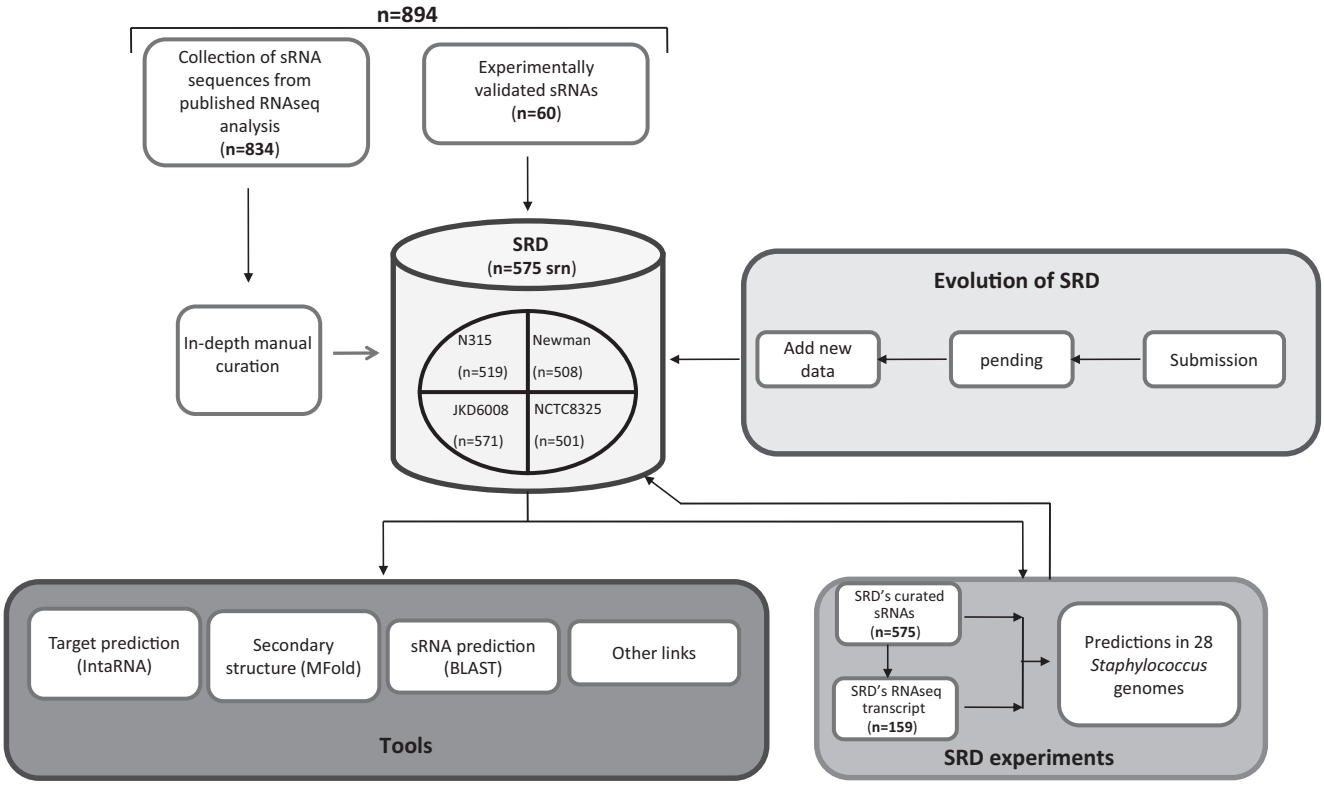


FIGURE 1. Overview of the SRD's inputs and outputs.

communications as a single sRNA sequence can harbor multiple IDs. To cope with that, we assigned novel and simple identifier that clarifies the actual repertoire of *S. aureus* sRNAs. The genome of N315 (Kuroda et al. 2001) was used as a reference as many sRNA sequences were identified in this strain, with some experimentally validated (Pichon and Felden 2005; Beaume et al. 2010; Chabelskaya et al. 2010). Each *srna* gene was assigned with a *srn* (***s**taphylococcus **r**ibo**n**ucleotide*) gene identifier. The *srn* gene identifiers were numbered based on their genetic location onto the genome of *S. aureus* N315 strain and starting from the origin of replication and therefore do not reflect their transcriptional level. *srn* numbers were assigned by increments of 10 to anticipate the identification of new sRNAs in the upcoming future which would be also numbered based on their genomic locations. When a *srn* gene is present in other strains it keeps the *srn* number assigned onto the N315 genome. When a *srn* identified in N315 is present in multiple copies in N315 or in another strain, the number of copies is provided (*srn_1930.1, srn_1930.2*, and *srn 1930.3*), unless experimental evidence indicated that they should be considered as distinct sRNAs (Supplemental Data S3, S4). *srn* identified in other strains and absent in N315 strain will be assigned with numbers starting from *srn_9000* (*srn_9480* formerly known as *sRNA334*). There were only four sRNAs (4.5 S, 6S, tmRNA, and RNase P RNA) for which a srn identifier was not generated, based on their extensive nomenclature, sequence, and functional conservation beyond Staphylococcal genomes (Supplemental Data S5). Therefore, regulatory RNAs such as riboswitches, RNAIII, or RsaE that are pioneered and/or conserved within the genus were also assigned with a srn identifier similar to what was done with the JKD6008 strain (Howden et al. 2013). However, a column entitled "most common name" is present on the website to avoid confusion when dealing with already well-described sRNAs. Additional information, including all other previously published names is provided with the list of *srn* on the SRD website.

Description of sRNAs in SRD

Among the 575 genomic sequences described in SRD, there are only 60 transcripts identified by multiple experimental approaches (including Northern Blot, 5′ and 3′ RACE, RT-qPCR, or RNA-seq) and a few for which their functions were characterized (Supplemental Data S3 and SRD website). Among these 60 sRNAs, 49 were described as transcripts whereas 11 were annotated as antisense sRNAs. The majority of the 575 sRNAs were identified or validated by Next-Generation RNA Sequencing (Beaume et al. 2010; Bohn et al. 2010; Lasa et al. 2011; Lioliou et al. 2012; Howden et al. 2013). Although powerful, RNA-seq, as any global approach, can lead to a substantial amount of false positive transcripts due to genomic DNA contaminations, reads mapping onto repeated sequences or/and the inaccurate detection of

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

TABLE 2. *Staphylococcus* strains used for implementing the SRD database

Staphylococcus strains	GenBank accession number	Genome length	Number of srn genes ([a]/[b])
Reference strains			
S. aureus subsp. *aureus* str. N315	BA000018.3	2,814,816	153/519
S. aureus subsp. *aureus* str. Newman	AP009351.1	2,878,897	156/508
S. aureus subsp. *aureus* str. NCTC8325	CP000253.1	2,821,361	155/501
S. aureus subsp. *aureus* str. JKD6008	CP002120.1	2,924,344	155/571
srn genes predictions in *Staphylococcus aureus*			
S. aureus 502A	CP007454.1	2,764,699	151/485
S. aureus subsp. *aureus* str. COL	CP000046.1	2,809,422	147/490
S. aureus M1	HF937103.1	2,864,125	149/499
S. aureus subsp. *aureus* str. M013	CP003166.1	2,788,636	138/461
S. aureus subsp. *aureus* str. MRSA252	BX571856.1	2,902,619	143/468
S. aureus subsp. *aureus* str. MSHR1132	FR821777.2	2,762,785	112/281
S. aureus subsp. *aureus* str. Mu50	BA000017.4	2,878,529	153/485
S. aureus subsp. *aureus* str. MW2	BA000033.2	2,820,462	147/485
S. aureus NRS 100	CP007539.1	2,823,087	147/490
S. aureus subsp. *aureus* str. USA300 FPR3757	CP000255.1	2,872,769	151/498
S. aureus XN108	CP007447.1	3,052,055	156/499
S. aureus subsp. *aureus* str. ST398	AM990992.1	2,872,582	140/448
S. aureus subsp. *aureus* JH1	CP000736.1	2,906,507	156/498
S. aureus subsp. *aureus* JH9	CP000703.1	2,906,700	156/498
S. aureus subsp. *aureus* MSSA476	BX571857.1	2,799,802	147/481
S. aureus subsp. *aureus* Mu3	AP009324.1	2,880,168	153/498
S. aureus subsp. *aureus* USA300_TCH1516	CP000730.1	2,872,915	152/499
S. aureus RF122	AJ938182.1	2,742,531	138/445
srn genes predictions in other *Staphylococcus*			
S. carnosus subsp. *carnosus* TM300	AM295250.1	2,566,424	9/22
S. epidermidis ATCC 12228	AE015929.1	2,499,279	19/53
S. epidermidis RP62A	CP000029	2,616,530	21/65
S. haemolyticus JCSC1435	AP006716.1	2,685,015	19/56
S. lugdunensis HKU09-01	CP001837.1	2,658,366	17/48
Staphylococcus pasteuri SP1	CP004014.1	2,559,946	26/55
Staphylococcus pseudintermedius ED99	CP002478.1	2,572,216	13/31
S. saprophyticus subsp. *saprophyticus* ATCC 15305	AP008934.1	2,516,575	15/41
S. warneri SG1	CP003668.1	2,486,042	25/54
S. xylosus HKUOPL8	CP007208.1	2,836,901	15/31

[a]Number of SRD's RNA-seq transcribed *srna* genes.
[b]Numbers of *srn* were determined after RNA-seq analysis and predictions performed with a set of 575 nonredundant gene sequences.

transcripts by bioinformatics. Some of the sRNAs previously identified by RNA-seq were of short lengths (<50 nt), mostly detected as *cis*-encoded antisense RNAs, including a wide antisense transcription that is not understood yet (Lasa et al. 2011). Also, some of the previously published sRNAs were already suspected as potential 5′ UTRs or 3′ UTRs (Beaume et al. 2010; Howden et al. 2013). Therefore, we searched for the presence of srn transcripts in strains N315 and Newman (data submitted to GEO with the accession number GSE64026), cultivated until the exponential phase of growth in a rich medium (BHI), and for strain NCTC8325 cultivated under 16 different growth conditions (growth phase, temperature, O_2 limitations, etc.; data kindly provided by Drs. P. Bouloc and T. Rochat). Those three strains are representative as their genomes were completed and annotated (Kuroda et al. 2001; Gillaspy et al. 2006; Baba et al. 2008). Based on our initial compilation and curation, N315, Newman, and NCTC8325 are predicted to contain 518, 508, and 501 independent *srn* genes, respectively. Altogether, they represent a pool of 535 *srn* genes. RNA-seq reads were mapped at unique locations onto their respective genomes, counted using HTSeq (Anders et al. 2014) and the presence of transcripts analyzed using Artemis (Carver et al. 2012). Based on the results from HTSeq (Supplemental Data S4), an FPKM normalization (Fragment per kilobase per millions of fragments mapped) and from the visualization of reads onto the annotated genomes (Fig. 2), the *srn* were either described as transcripts (Fig. 2A), *cis*-antisense RNAs (Fig. 2B), 5′ UTR (Fig. 2C), 3′ UTR (Fig. 2D), CDS (Fig. 2E, coordinates inside an annotated gene) or not detected (Fig. 2F, ND) in the "SRD's RNA-seq evidence" column of the SRD website. A srn was considered as a transcript or *cis*-antisense (reads mapping onto the opposite strand of a CDS) when the HTSeq count was equal or >15 (Howden et al. 2013), the FPKM normalization >2, the mapping quality (MAPQ score in a SAM file) >30 (probability of a correct match equal to 0.999), and when the

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

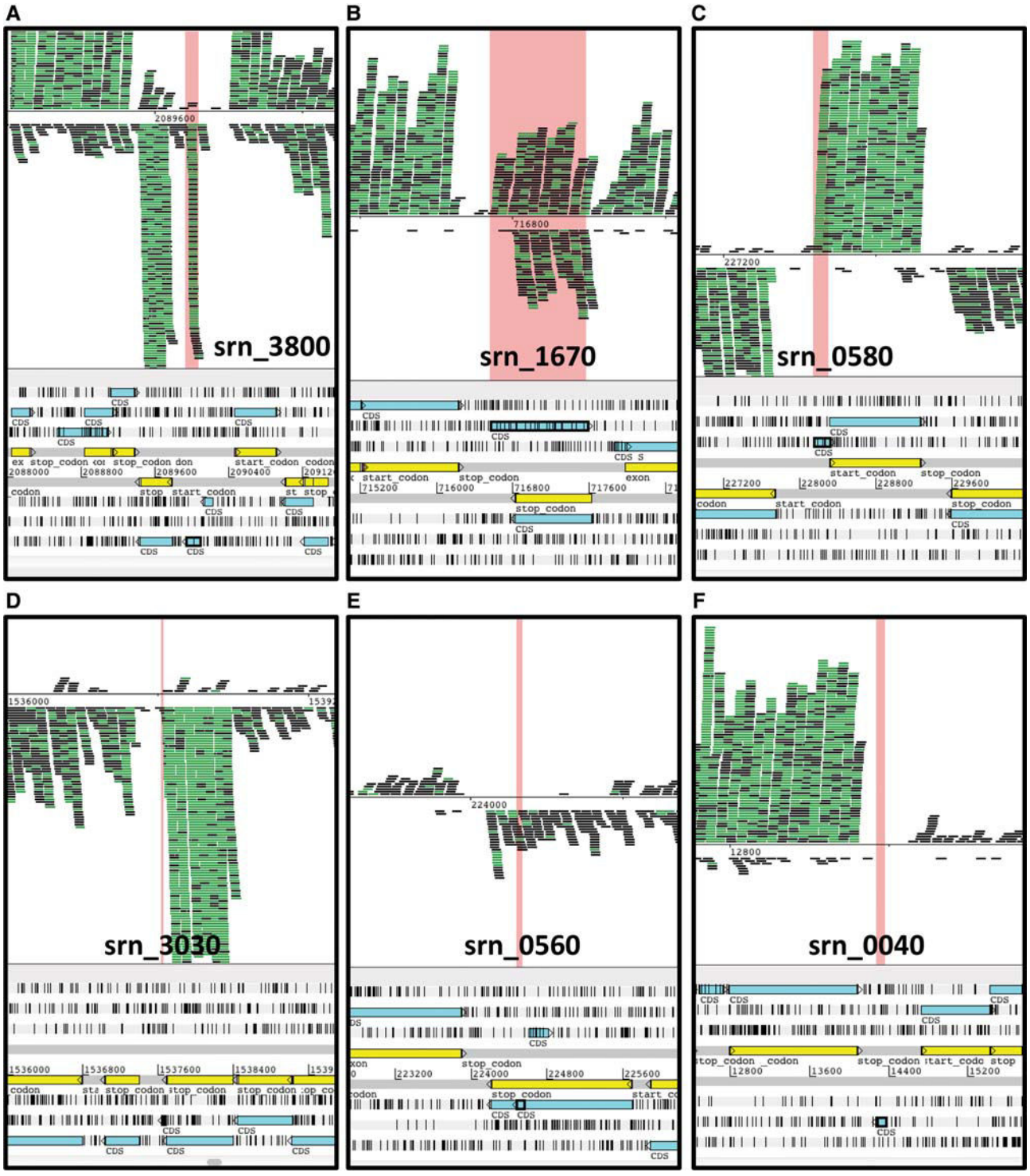


FIGURE 2. Examples of the visualization of read mapping from strain Newman using Artemis. The srns are highlighted in pink. (*A*) Typical visualization for a srn described as transcript. (*B*) Example of an antisense srn. (*C*) Reads overlapping with a CDS and considered as a 5′ UTR. (*D*) srn described as a 3′ UTR. (*E*) srn identified within a CDS. (*F*) srn not detected.

reads did not overlap with annotated genes. The UTRs were described as 5′ UTRs or 3′ UTRs when the reads are assembled with CDS and when the expression was similar to the flanking genes, as described (Yoder-Himes et al. 2009). The results of our analysis are summarized in Table 3. In N315, we identified 94 *srn* as transcripts, 24 as *cis*-antisense RNAs, 14 as CDS, 58 as

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

TABLE 3. Description of *srns* based on SRD's RNA-seq analysis

	N315	Newman	NCTC8325
CDS	14	13	24
5′ UTR	58	54	59
3′ UTR	31	32	28
Not detected (ND)	297	278	262
Transcript	94	100	104
Antisense	24	30	24
Sum	518	507	501

The *srns* were described as "CDS," "5′ UTRs," "3′ UTRs," "ND," "Transcripts," or "Antisense" based on the following criteria: HTSeq count ≥15, FPKM ≥2, MAPQ ≥30 and no clusterization with flanking CDS.

5′ UTR, 31 as 3′ UTR, and 297 were not detected. Similar results were obtained for Newman and NCTC8325 strains, respectively (see the SRD website). *srn* annotated as "ND" were either not detected by HTSeq count or did not meet some of the criteria retained for describing a sequence as transcript or a *cis*-antisense. Current Illumina sequencing kits are not perfectly adapted to the search of small bacterial transcripts and therefore it is difficult to know whether there are limits in the technology or that transcripts are absent. Forty-five of the *srn* that were not detected were predicted to have a length <50 nt. Ninenty-seven *srn* were predicted to be between 50 and 100 nt and the others were longer than 100 nt. Overall, only 159 srn were detected as transcripts or as RNA antisense under our experimental conditions (Supplemental Data S5 and http://srd.genouest.org/browsevalidated). Also, there were 24 ambiguous *srn* (Supplemental Data S6). These ambiguous *srn* presented criterion that did not allow us to add them in the list of 159 sRNA (see comments in Supplemental Data). All these results suggest that some sRNAs are expressed or detected under specific conditions that await further experimental assessment, while other *srn* may not be considered as independent sRNA transcriptional units or may be false positive and arise from transcriptional noise.

sRNA predictions in other Staphylococcal species and strains

Four *S. aureus* genomes were used to generate SRD. It is however important to include in a database a large number of strains and other species. Therefore, we performed *srn* predictions on a set of 28 strains which included 18 *S. aureus* subspecies and 10 other Staphylococci (Table 2). BLAST (see Materials and Methods) was used to predict the presence of *srn* genes using either the list of 159 *srn* confirmed by our RNA-seq analysis or the curated list containing 575 *srn*. Using the of 159 *srn*, this resulted in a number of predicted nucleotide sequences that ranged from 112 *srn* genes for *Staphylococcus aureus* subsp. *aureus* str. MSHR1132 to 156 *srn* genes for *Staphylococcus aureus* subsp. *aureus* str. JH1 and *Staphylococcus aureus* subsp. *aureus* str. JH9, respectively. On the opposite a reduced number of genes was predicted for the 10 other *Staphylococcus* species, with only 9 *srn* genes found for *Staphylococcus carnosus* subsp. *carnosus* str. TM300 and a maximum of 26 *srn* genes for *Staphylococcus pasteuri* SP1. Taken together these results suggest a low level of nucleotide sequence conservation for the *srn* genes within the genus.

Comparative analysis of the *srna* gene sequences

To investigate further the 159 *srn* gene sequences detected in the various strains a phylogenetic analysis was performed. A "phylogenomic" tree based on *Staphylococcus* whole-genome content (*Staphylococcus* tree) (Fig. 3A) and a tree based on the *srn* gene content were constructed (*srn* genes tree) (Fig. 3B). Interestingly, the tree based on the *srn* content showed an outstanding similar topology compared wth the tree based on the genome content and clearly differentiates *Staphylococcus aureus* from the other *Staphylococcus* species (Fig. 3). For both trees, *S. aureus* subsp. *aureus* str. MRSH 132 appeared more distant from other *S. aureus* subspecies. This could explain why only 112 *srn* genes were predicted using BLAST for that strain. A "heatmap" representation constructed using a matrix of presence/absence of *srn* sequences in *Staphylococcus* genomes showed a species clusterization similar to that of the *Staphylococcus* tree (Fig. 4). These results confirm a weak conservation of the 159 *srn* outside the *aureus* species. In addition, *rnaseP* RNA gene, *4.5S* RNA gene, *srn_3910* (*RNAIII*), and *srn_2130* (*rsaE*) were identified in all *Staphylococcus* strains while the *tmRNA* sequences appears to be substantially degenerated only in *Staphylococcus xylosus* str. HKUOPL8. Regarding the *S. aureus* subspecies, we identified 96 *srn* conserved which we defined as a core sRNA set in this species.

Database overview and usage

Users can access the list of *srn* genes through the Web interface (Fig. 5A). From the SRD home page they can (i) access the data set corresponding to the four genomes used to construct SRD, (ii) retrieve a short list of *srn* genes with a unified annotation, (iii) browse for predictions in other Staphylococci (described elsewhere in the text), (iv) BLAST (Altschul et al. 1990) their own sRNA sequences against the entire SRD database, or (v) search for RNA–RNA interactions using the intaRNA program (Busch et al. 2008). After entering in the webpage of one of the four curated genomes, the users will have access to the full list of *srn* reported so far. For each table, the new nomenclature, the genomic coordinates, the orientation, previous names, and experimental support are provided. The column "experimental evidence" was added for the community in order to have a quick overview of the *srn* that were identified as transcripts through RNA-seq analysis or other experimental approaches. The molecular targets and

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

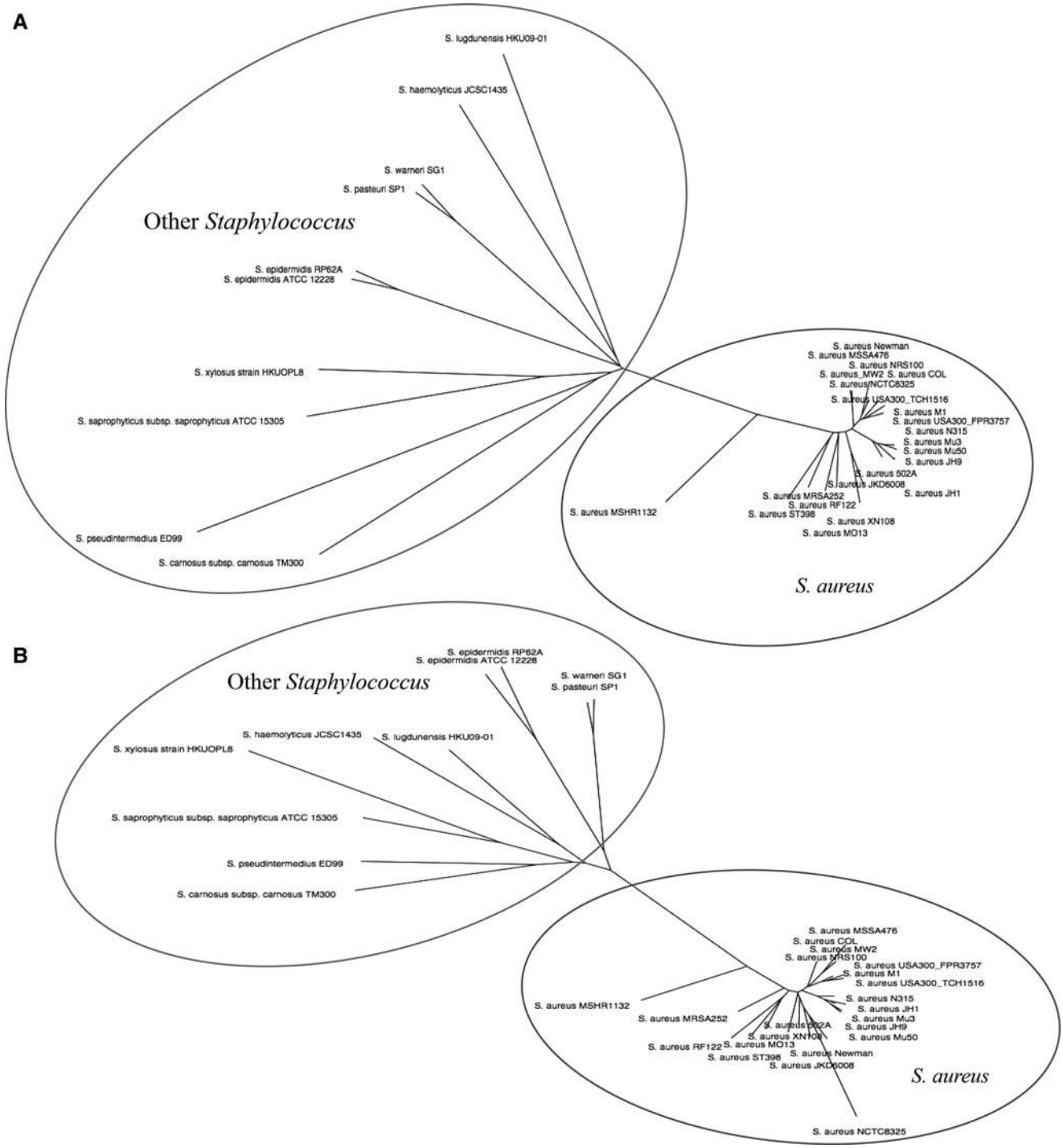


FIGURE 3. Phylogenetic analysis on the genome and *srn* content of 32 strains of the *Staphylococcus* genus using the Neighbor-joining algorithm. (*A*) *Staphylococcus* tree-based on genome content. (*B*) *Staphylococcus* tree-based on *srn* content.

references to the experimental secondary structure information are provided within the "SRD's RNA-seq transcribed sRNA" tab of the website. In addition, the users can download the Staphylococcal genomes and the *srn* genes in diverse formats ("FASTA" or "gff") suitable for subsequent RNA-seq analysis (Fig. 5B).

SRD specific features

Several functions have been included in the SRD website to provide an efficient device for the community working on sRNAs. BLAST and intaRNA (Altschul et al. 1990; Busch et al. 2008) were implemented in the database while external

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

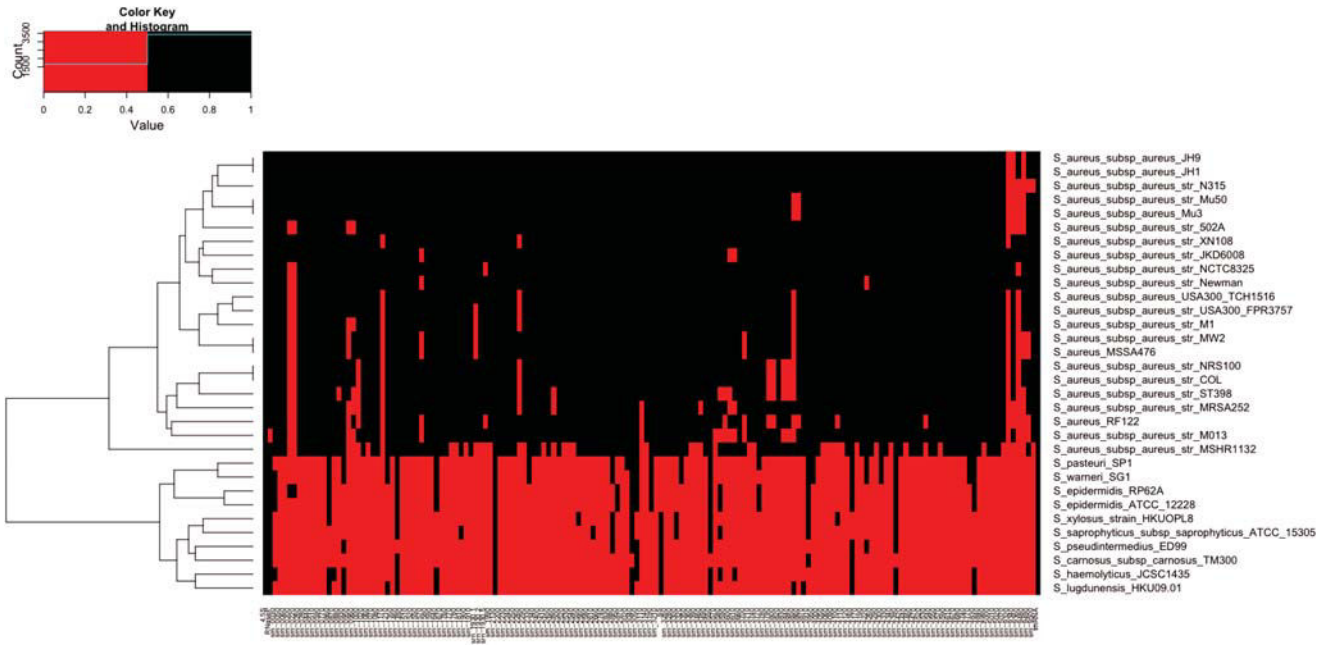


FIGURE 4. Comparative analysis through a "heatmap" cluster based on a matrix of presence (black) and absence (red) of *srn* sequences.

links to RNAtarget2 (Kery et al. 2014) and to RNA predator (Eggenhofer et al. 2011) are easily accessible from the webpage describing each sRNA. For BLAST comparisons, our databank includes the four initial reference genomes, the lists of validated sRNAs, and all the *srn* curated in the four strains. IntaRNA can be activated from the intaRNA page or directly from any *srn* sequences by clicking on the dedicated symbol. The users will only have to paste their mRNA sequences to see whether they may interact with each sRNA from SRD. MFold structure predictions (Zuker 2003), defined for the 575 srn RNA sequences, can be downloaded also directly from the SRD website.

Evolution of the database and future directions

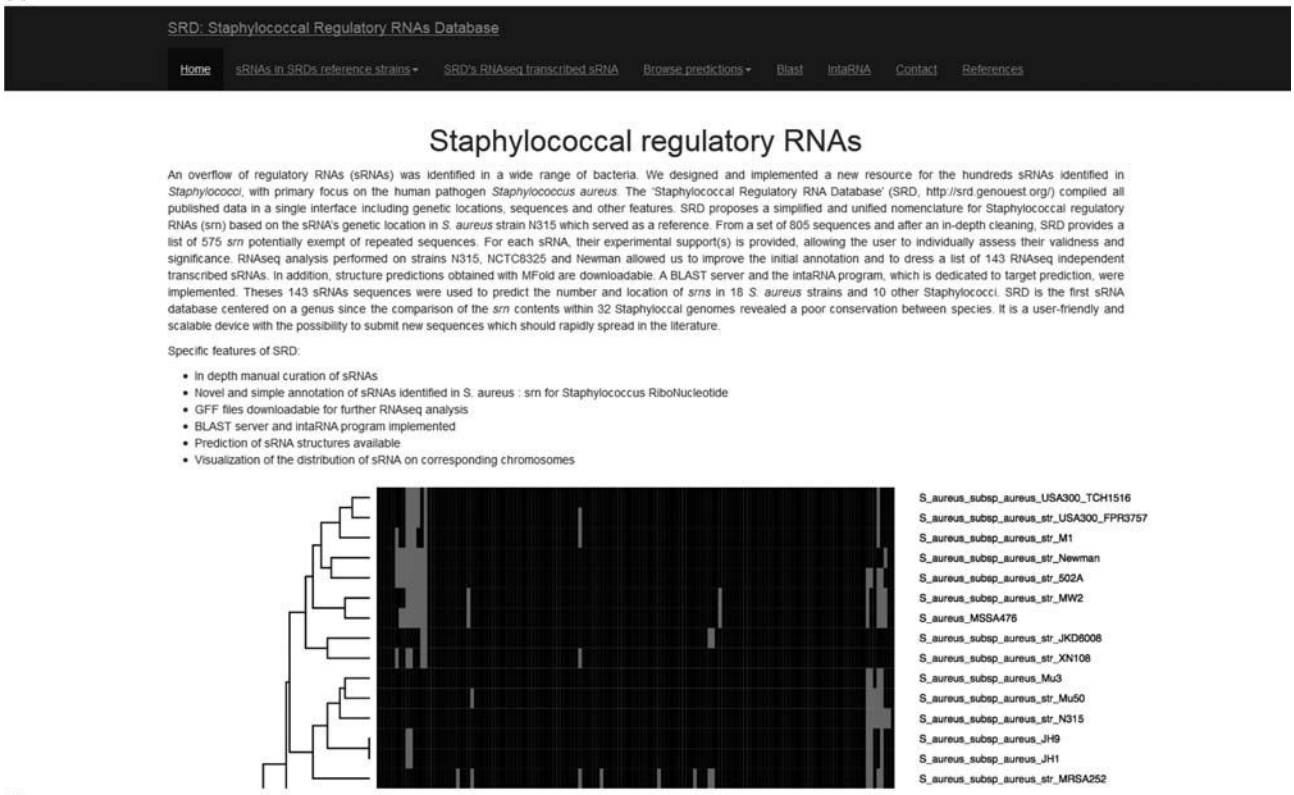
SRD is a repository for the Staphylococcal sRNAs with primary focus on *S. aureus*. By combining an in-depth cleaning and novel RNA-seq analysis, it clarifies the status regarding the absence of a consensus nomenclature and also to the actual expanding number of individual sRNA genes. With the rising interest of the community in the field of sRNAs, the number of sRNAs detected or characterized in bacteria is predicted to constantly increase. SRD will therefore evolve, based on the identification of novel sRNAs published in the literature. In addition, researchers who would be willing to unify new discoveries under the srn nomenclature will have the possibility to submit their sRNA sequence(s) to SRD under the contact information box. SRD would allow both already published and nonpublished submission. For data that would not be already published, temporary srn numbers will be assigned and disclosed upon acceptance of the publications in peer-reviewed journals. Therefore, researchers will have the possibility to disclose their results directly with the *srn* nomenclature. To avoid the diffusion of a large number of genes that may not be confirmed later, researchers will be invited to describe how a new sRNA was identified. For the novel sRNAs detected by RNA-seq analysis, that could lead to false-positive sRNAs, a secure ftp link will be provided to solicit researchers to deposit their "fastq" files to check whether they meet the criteria described in the text for being annotated as sRNA transcripts.

DISCUSSION

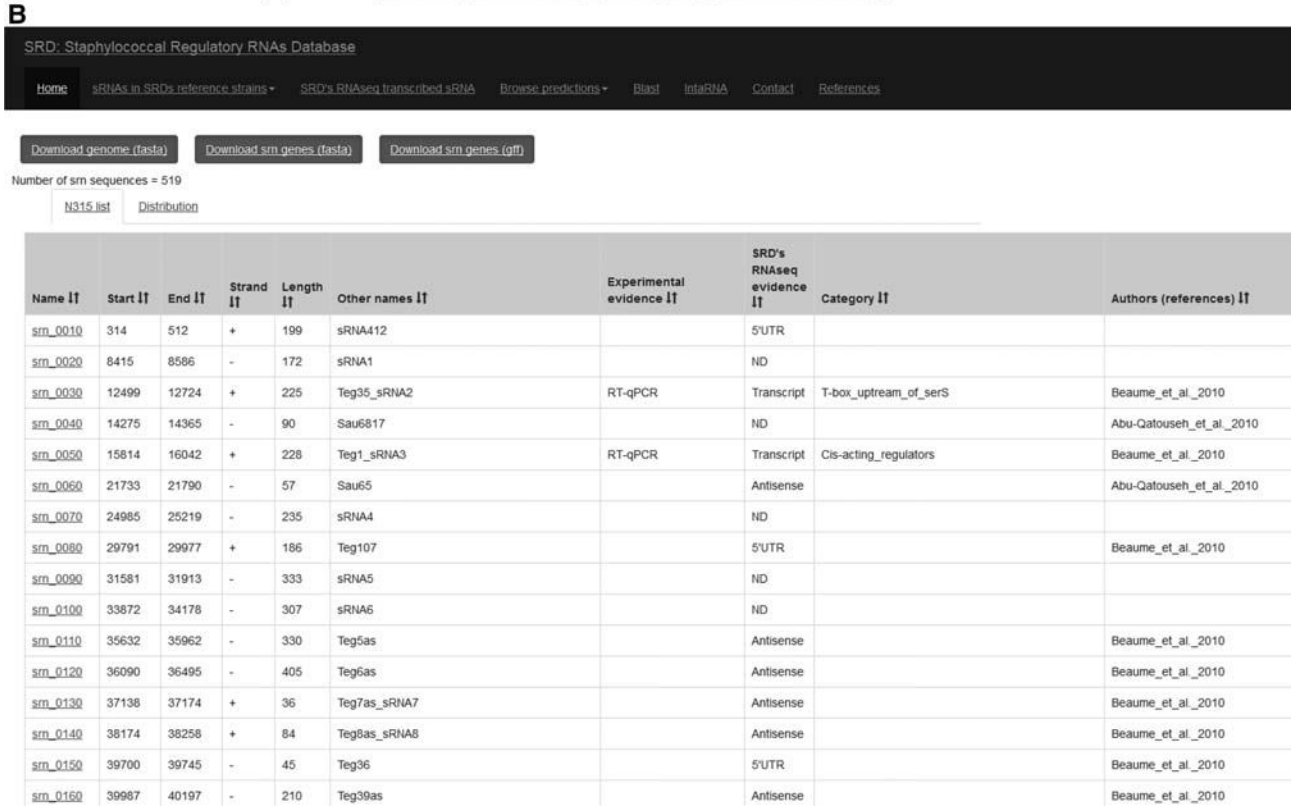
Over the last 20 yr, the number of sequences identified as sRNAs dramatically increased in *Staphylococcus aureus*. However, the democratization of Next-Generation Sequencing and the absence of a consensus in the community for annotating newly identified sRNAs led to a growing confusion that become detrimental to the field. SRD compiled the existing data in a single interface, removed the repeated and or redundant sequences and proposed a novel, simplified, nomenclature. Compared with the databases on other prokaryotes sRNAs, SRD is the first sRNA database dedicated to the *Staphylococcus* genus. SRD assigned a single identifier for the whole genus while other sRNA databases, such as BSRD (Li et al. 2013), provide an index per strain. A unique identifier offers a substantial advantage when comparing different strains, as it should avoid the dissemination of multiple redundancies. Therefore, the data provided by SRD would have been difficult to fuse with other databases. SRD hosts a

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

A



B



Name	Start	End	Strand	Length	Other names	Experimental evidence	SRD's RNAseq evidence	Category	Authors (references)
srn_0010	314	512	+	199	sRNA412		5'UTR		
srn_0020	8415	8586	-	172	sRNA1		ND		
srn_0030	12499	12724	+	225	Teg35_sRNA2	RT-qPCR	Transcript	T-box_uptream_of_serS	Beaume_et_al._2010
srn_0040	14275	14365	-	90	Sau6817		ND		Abu-Qatouseh_et_al._2010
srn_0050	15814	16042	+	228	Teg1_sRNA3	RT-qPCR	Transcript	Cis-acting_regulators	Beaume_et_al._2010
srn_0060	21733	21790	-	57	Sau65		Antisense		Abu-Qatouseh_et_al._2010
srn_0070	24985	25219	-	235	sRNA4		ND		
srn_0080	29791	29977	+	186	Teg107		5'UTR		Beaume_et_al._2010
srn_0090	31581	31913	-	333	sRNA5		ND		
srn_0100	33872	34178	-	307	sRNA6		ND		
srn_0110	35632	35962	-	330	Teg5as		Antisense		Beaume_et_al._2010
srn_0120	36090	36495	-	405	Teg6as		Antisense		Beaume_et_al._2010
srn_0130	37138	37174	+	36	Teg7as_sRNA7		Antisense		Beaume_et_al._2010
srn_0140	38174	38258	+	84	Teg8as_sRNA8		Antisense		Beaume_et_al._2010
srn_0150	39700	39745	-	45	Teg36		5'UTR		Beaume_et_al._2010
srn_0160	39987	40197	-	210	Teg39as		Antisense		Beaume_et_al._2010

FIGURE 5. Screenshots of SRD. (*A*) Main webpage interface to navigate and access specific features within the database. (*B*) Example of the presentation of *srn* data determined after curation of repeated and redundant sequences.

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

large collection of manually curated Staphylococcal sRNAs (575 *srn* genes) mostly exempted of repeated sequences and of redundancies. Furthermore, coupled with an SRD's RNA-seq analysis, a list of 159 RNA-seq transcribed srn is provided that includes all the 60 previously experimentally supported sRNAs. This suggests that the SRD's criteria used to describe the srn transcripts that combined previously published cut offs (Yoder-Himes et al. 2009; Howden et al. 2013) with high mapping quality scores and FPKM normalization (to prevent the detection of incorrectly mapped reads or of long transcripts that do not arise from transcriptional noise) were relevant. However, among the nearly 300 *srn* genes not transcribed under the physiological conditions tested, some may be later annotated as sRNA transcripts once additional data and work will provide transcriptional evidence. From this set of 159 *srn* genes, predictions were issued by sequence similarity to identify homologous genes in other Staphylococcal species. A comparison of the SRD's predictions with the BSRD entries for *Staphylococcus* strains shows that a genus-specific database is relevant. Indeed, in BSRD there is a huge discrepancy in the number of sRNAs sequences being available within the *S. aureus* subps. *aureus* species. While 154 sRNAs were listed for strain N315, mostly based on the work of Beaume et al. (2010), there were <60 sequences inventoried by similarity for other Staphylococcal strains (Li et al. 2013). All BSRD *S. aureus* entries are present in our database, and the predictions performed within the genus from a set of 575 *srn* genes (*versus* 8248 genes for BSRD), led to a larger set of predicted genes in SRD (Table 4). The comparative analysis performed for the *srn* content confirmed a low level of conservation within the genus and therefore in the bacterial kingdom. The weak sequence conservation at the DNA level in comparison with the protein level is therefore a serious limitation for retrieving genes, and therefore sRNAs, between species (Konstantinidis and Tiedje 2007; Sentausa and Fournier 2013). To our knowledge, there is no recognized standard to assign identifiers or nomenclature to sRNAs. Therefore, to not favor anyone name (spr, rsa, sau, teg, and others) we created the srn identifier. In addition, a most common name was assigned based on the chronological discovery of experimental confirmation to avoid the community to name some of the already very well-described sRNAs under a srn identifier while using the database. However, for the sRNAs that were only described by RNA-seq, the community is encouraged to use the srn identifier in an effort for unifying the work done on the sRNAs so far. We believe that SRD will help the scientific community working on Staphylococcal sRNA identification, function, and biology. SRD provides a simple and unified sRNA resource, a detailed annotation for each sRNA, a direct access to various RNA and genomic analysis tools. Finally, it shall encourage the community to participate in an effort to submitting new *Staphylococcus* sRNAs in SRD and to develop other genus specific sRNA databases that should be an essential extension to the generalist sRNA databases.

TABLE 4. Comparison between BSRD and SRD number of entries

Staphylococcus strains	BSRD sRNAs	SRD's RNA-seq and predictions
S. aureus subsp. *aureus* N315	154	519
S. aureus subsp. *aureus* Newman	58	508
S. aureus subsp. *aureus* NCTC 8325	55	571
S. aureus subsp. *aureus* JKD6008	NA	501
S. aureus subsp. *aureus* M013	NA	461
S. aureus subsp. *aureus* MSHR1132	NA	281
S. aureus subsp. *aureus* ST398	NA	448
S. aureus subsp. *aureus* COL	53	490
S. aureus subsp. *aureus* JH1	55	498
S. aureus subsp. *aureus* JH9	55	498
S. aureus subsp. aureus MRSA252	57	468
S. aureus subsp. *aureus* MSSA476	55	481
S. aureus subsp. *aureus* Mu3	55	498
S. aureus subsp. *aureus* Mu50	55	485
S. aureus subsp. *aureus* MW2	56	485
S. aureus subsp. *aureus* USA300_FPR3757	58	498
S. aureus subsp. *aureus* USA300_TCH1516	59	499
S. aureus 502A	NA	485
S. aureus XN108	NA	499
S. aureus M1	NA	499
S. aureus NRS 100	NA	490
S. aureus RF122	54	445
S. carnosus subsp. *carnosus* TM300	52	22
S. epidermidis ATCC 12228	53	53
S. epidermidis RP62A	57	65
S. haemolyticus JCSC143	56	56
S. lugdunensis HKU09-01	NA	48
S. pasteuri SP1	NA	55
S. pseudintermedius ED99	NA	31
S. saprophyticus subsp. *saprophyticus* ATCC 15305	53	41
S. warneri SG1	NA	54
S. xylosus HKUOPL8	NA	31

BSRD numbers of sRNAs were determined from data collected in the literature and predictions performed from a set of 8248 nonredundant gene sequences. SRD numbers of *srn* were determined after RNA-seq analysis and predictions performed with a set of 575 nonredundant gene sequences. (NA) Not assayed.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Staphylococcus aureus strains Newman and N315 were grown in liquid Brain Heart Infusion broth (BHI, Oxoid) and Tryptone Soya Broth (TSB, Oxoid) at 37°C, under agitation.

RNA extraction

Overnight cultures of *S. aureus* were diluted to an $OD_{600\ nm}$ of 0.1 into fresh BHI broth and cultured for 5 h at 37°C at 160 rpm. Cells were harvested by centrifugation at 15,000*g* for 30 sec and pellets washed with 500 µL of cold lysis buffer (20 mM sodium acetate, 1 mM EDTA, 0.5% SDS at pH 5.5). Cells were broken out using acid

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

treated glass beads (Sigma) in the presence of phenol (pH 4) in an FP120 FastPrep cell disruptor (MP Biomedicals) for 30 sec at power 6.5. Lysates were centrifuged for 5 min at 16,000*g* at 4°C. Total RNAs were extracted with phenol/chloroform and precipitated overnight. The RNA samples were treated with DNase I, Amplification Grade (Invitrogen). The absence of DNA contamination was checked by qPCR in an Applied Biosystems instrument. The integrity of each RNA preparation was verified on a "Bioanalyzer" (Agilent).

cDNA Library construction and Illumina RNA-seq

Ribosomal RNAs were depleted using the Ribo-Zero Magnetic Kit (Epicentre) and following manufacturer's recommendations. Stranded cDNAs libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). The concentration, quality, and purity of the libraries were determined on a BioAnalyzer (Agilent), a Qubit fluorometer (Invitrogen), and a Nanodrop (Thermo Scientific). Libraries were pooled and sequenced on an Illumina Hiseq 1500 instrument following the manufacturer's recommendations and using the rapid run mode for 200 cycles in paired-end.

Read mapping and visualization

The genome sequences and annotation files were obtained from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). A "fastqc" report was performed prior mapping the data onto the appropriate genome. RNA-seq reads were mapped using Tophat2 and BWA (Li and Durbin 2010; Kim et al. 2013) with initial settings modified to allow the mapping of stranded library of an average mean distance between mates of 250 (for Tophat2) and to allow the alignment of a read to a single location with no mismatch (Tophat2 and BWA). BAM files were converted into SAM files and filtered on bitwise flag values (Li et al. 2009) to select properly paired reads. SAM files were then converted to BAM files, sorted by query, and counted by HTSeq count (Anders et al. 2014) for stranded library with the mode union. BAM files were visualized using the Artemis program (Carver et al. 2012).

Data set building and comparative analysis

The 32 genomes of *Staphylococcus* used in this study were downloaded from GenBank (Table 2). The published sRNA sequences were extracted from their respective genomes to construct a pool of sRNA. This pool was used to predict the *srn* genes in reference genomes. Then a pool of SRD's RNA-seq transcribed sRNA was used to predict *srn* genes in 28 *Staphylococcus* genomes. The predictions were done using "BLASTN" with a cutoff *E*-value $<1 \times 10^{-20}$, percentage similarity >80% and an alignment length >60 nt of the query length. The 32 *Staphylococcus* genomes and the predicted *srn* genes sequences were aligned using Muscle aligner implemented in Mauve software (Darling et al. 2010). Mauve alignment generated a genome content matrix for which the identity scores range between 0 and 1, where 0 indicates that no identical homologous nucleotides were found, and 1 indicates that every homologous nucleotide was identical. A matrix based on the *srn* gene content was generated (the similarity between two species is defined as the number of genes that they have in common divided by the total number of *srn* genes) (Snel et al. 1999, 2005; Huson and Bryant 2006). The genome content based matrix and the *srn* gene content based matrix were then used to construct, respectively, a *Staphylococcus* "phylogenomic" tree-based on genome content and a *Staphylococcus* tree-based on srn content, using Neighbor-joining algorithm in the package SplitsTree4 (Huson and Bryant 2006). A "heatmap" clusterization was constructed using a matrix based on presence and absence of *srn* genes using the R package (http://www.r-project.org/).

Database design

The web server has been designed in PHP with the "Symfony" framework (http://symfony.com). It includes a set of scripts that automatically parse the raw input files (*srn*, genomes), fill in a MySQL database for each set of genome/*srn*, and build some templates for all input predictions. Those scripts allow the addition of other genomes easily by simply adding the new files to the directory structure. They will update the existing information and insert the new ones. The website menus are automatically adapted to the list of analyzed genomes, removing the need to modify the website when new information is added (new genomes, new predictions, etc.). The scripts also (i) extract the FASTA sequence of each *srn* (to prefill IntaRNA form for example), (ii) execute MFold to create the pdf files containing the structure, and (iii) prepare a blast index for all genomes and set of *srn*.

A *srn* distribution is proposed in the web interface with pan and zoom options. This distribution is displayed using the D3js library (http://d3js.org/).

DATA DEPOSITION

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE64026 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64026).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This research was supported by a Marie Curie International Incoming Fellowship (project 621959—SarHyb) within the 7th European Community Framework Programme and by the region Bretagne grant SAD 2013—SARS 8254. T.M. is a recipient of a fellowship from the Direction Générale pour l'Armement. We thank Dr. Philippe Bouloc and Dr. Tatiana (Orsay—Université Paris Sud) Rochat for providing a BAM file combining 16 growth conditions for strain NCTC8325. We are most grateful to the "plate-forme Génomique Santé" Biogenouest Genomics, Biosit core facility for its technical support and to the CNRS-UPMC ABiMS bioinformatics platform (http://abims.sb-roscoff.fr) for providing computational resources and support. This work has benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGIF (Centre de Recherche de Gif—www.imagif.cnrs.fr).

Received December 17, 2014; accepted February 16, 2015.

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

REFERENCES

Abu-Qatouseh LF, Chinni SV, Seggewiss J, Proctor RA, Brosius J, Rozhdestvensky TS, Peters G, von Eiff C, Becker K. 2010. Identification of differentially expressed small non-protein-coding RNAs in *Staphylococcus aureus* displaying both the normal and the small-colony variant phenotype. *J Mol Med (Berl)* **88:** 565–575.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166–169.

Anderson KL, Roberts C, Disz T, Vonstein V, Hwang K, Overbeek R, Olson PD, Projan SJ, Dunman PM. 2006. Characterization of the *Staphylococcus aureus* heat shock, cold shock, stringent, and SOS responses and their effects on log-phase mRNA turnover. *J Bacteriol* **188:** 6739–6756.

Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K. 2008. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol* **190:** 300–310.

Beaume M, Hernandez D, Farinelli L, Deluen C, Linder P, Gaspin C, Romby P, Schrenzel J, Francois P. 2010. Cartography of methicillin-resistant *S. aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *PLoS One* **5:** e10725.

Bohn C, Rigoulay C, Chabelskaya S, Sharma CM, Marchais A, Skorski P, Borezée-Durant E, Barbet R, Jacquet E, Jacq A, et al. 2010. Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic Acids Res* **38:** 6620–6636.

Brantl S, Brückner R. 2014. Small regulatory RNAs from low-GC Gram-positive bacteria. *RNA Biol* **11:** 443–456.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41:** D226–D232.

Busch A, Richter AS, Backofen R. 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24:** 2849–2856.

Caldelari I, Chao Y, Romby P, Vogel J. 2013. RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harb Perspect Med* **3:** a010298.

Cao Y, Wu J, Liu Q, Zhao Y, Ying X, Cha L, Wang L, Li W. 2010. sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments. *RNA* **16:** 2051–2057.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28:** 464–469.

Chabelskaya S, Gaillot O, Felden B. 2010. A *Staphylococcus aureus* small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. *PLoS Pathog* **6:** e1000927.

Cros MJ, de Monte A, Mariette J, Bardou P, Grenier-Boley B, Gautheret D, Touzet H, Gaspin C. 2011. RNAspace.org: an integrated environment for the prediction, annotation, and analysis of ncRNA. *RNA* **17:** 1947–1956.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5:** e11147.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210.

Eggenhofer F, Tafer H, Stadler PF, Hofacker IL. 2011. RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res* **39:** W149–W154.

Fechter P, Caldelari I, Lioliou E, Romby P. 2014. Novel aspects of RNA regulation in *Staphylococcus aureus*. *FEBS Lett* **588:** 2523–2529.

Felden B, Vandenesch F, Bouloc P, Romby P. 2011. The *Staphylococcus aureus* RNome and its commitment to virulence. *PLoS Pathog* **7:** e1002006.

Geissmann T, Chevalier C, Cros MJ, Boisset S, Fechter P, Noirot C, Schrenzel J, Francois P, Vandenesch F, Gaspin C, et al. 2009. A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res* **37:** 7239–7257.

Gillaspy AF, Worrell V, Orvis J, Roe BA, Dyer DW, Landollo JJ. 2006. *Staphylococcus aureus* NCTC8325 genome. In *Gram-positive pathogens* (ed. Fischetti V, et al.), pp. 381–412. ASM Press, Washington, DC.

Guillet J, Hallier M, Felden B. 2013. Emerging functions for the *Staphylococcus aureus* RNome. *PLoS Pathog* **9:** e1003767.

Howden BP, Beaume M, Harrison PF, Hernandez D, Schrenzel J, Seemann T, Francois P, Stinear TP. 2013. Analysis of the small RNA transcriptional response in multidrug-resistant *Staphylococcus aureus* after antimicrobial exposure. *Antimicrob Agents Chemother* **57:** 3864–3874.

Huang HY, Chang HY, Chou CH, Tseng CP, Ho SY, Yang CD, Ju YW, Huang HD. 2009. sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Res* **37:** D150–D154.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23:** 254–267.

Kery MB, Feldman M, Livny J, Tjaden B. 2014. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* **42:** W124–W129.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, Kojima A, Kimura Y, Komori T, Asai K. 2007. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res* **35:** D145–D148.

Konstantinidis KT, Tiedje JM. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10:** 504–509.

Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, et al. 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357:** 1225–1240.

Lalaouna D, Eyraud A, Chabelskaya S, Felden B, Massé E. 2014. Regulatory RNAs involved in bacterial antibiotic resistance. *PLoS Pathog* **10:** e1004299.

Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de los Mozos IR, Vergara-Irigaray M, Segura V, Fagegaltier D, Penades JR, Valle J, et al. 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci* **108:** 20172–20177.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Li L, Huang D, Cheung MK, Nong W, Huang Q, Kwan HS. 2013. BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res* **41:** D233–D238.

Lioliou E, Sharma CM, Caldelari I, Helfer AC, Fechter P, Vandenesch F, Vogel J, Romby P. 2012. Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet* **8:** e1002782.

Liu JM, Livny J, Lawrence MS, Kimball MD, Waldor MK, Camilli A. 2009. Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37:** e46.

Marchais A, Naville M, Bohn C, Bouloc P, Gautheret D. 2009. Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. *Genome Res* **19:** 1084–1092.

Mizuno T, Chou MY, Inouye M. 1984. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci* **81:** 1966–1970.

Downloaded from rnajournal.cshlp.org on March 27, 2015 - Published by Cold Spring Harbor Laboratory Press

Morrison JM, Miller EW, Benson MA, Alonzo F III, Yoong P, Torres VJ, Hinrichs SH, Dunman PM. 2012. Characterization of SSR42, a novel virulence factor regulatory RNA that contributes to the pathogenesis of a *Staphylococcus aureus* USA300 representative. *J Bacteriol* **194:** 2924–2938.

Nielsen JS, Christiansen MH, Bonde M, Gottschalk S, Frees D, Thomsen LE, Kallipolitis BH. 2011. Searching for small σ^B-regulated genes in *Staphylococcus aureus*. *Arch Microbiol* **193:** 23–24.

Novick RP, Iordanescu S, Projan SJ, Kornblum J, Edelman I. 1989. pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **59:** 395–404.

Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J* **12:** 3967–3975.

Pang KC, Stephen S, Dinger ME, Engström PG, Lenhard B, Mattick JS. 2007. RNAdb 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res* **35:** D178–D182.

Pichon C, Felden B. 2005. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci* **102:** 14249–14254.

Pinel-Marie ML, Brielle R, Felden B. 2014. Dual toxic peptide-coding *Staphylococcus aureus* RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep* **7:** 424–435.

Pischimarov J, Kuenne C, Billion A, Hemberger J, Cemič F, Chakraborty T, Hain T. 2012. sRNAdb: a small non-coding RNA database for gram-positive bacteria. *BMC Genomics* **13:** 384.

Roberts C, Anderson KL, Murphy E, Projan SJ, Mounts W, Hurlburt B, Smeltzer M, Overbeek R, Disz T, Dunman PM. 2006. Characterizing the effect of the *Staphylococcus aureus* virulence factor regulator, SarA, on log-phase mRNA half-lives. *J Bacteriol* **188:** 2593–2603.

Romilly C, Caldelari I, Parmentier D, Lioliou E, Romby P, Fechter P. 2012. Current knowledge on regulatory RNAs and their machineries in *Staphylococcus aureus*. *RNA Biol* **9:** 402–413.

Sayed N, Jousselin A, Felden B. 2012. A *cis*-antisense RNA acts in *trans* in *Staphylococcus aureus* to control translation of a human cytolytic peptide. *Nat Struct Mol Biol* **19:** 105–112.

Sentausa E, Fournier PE. 2013. Advantages and limitations of genomics in prokaryotic taxonomy. *Clin Microbiol Infect* **19:** 790–795.

Snel B, Bork P, Huynen MA. 1999. Genome phylogeny based on gene content. *Nat Genet* **21:** 108–110.

Snel B, Huynen MA, Dutilh BE. 2005. Genome trees and the nature of genome evolution. *Annu Rev Microbiol* **59:** 191–209.

Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* **43:** 880–891.

Tomasini A, Francois P, Howden BP, Fechter P, Romby P, Caldelari I. 2014. The importance of regulatory RNAs in *Staphylococcus aureus*. *Infect Genet Evol* **21:** 616–626.

Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y. 2014. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* **42:** D98–D103.

Xue T, Zhang X, Sun H, Sun B. 2014. ArtR, a novel sRNA of *Staphylococcus aureus*, regulates α-toxin expression by targeting the 5′ UTR of sarT mRNA. *Med Microbiol Immunol* **203:** 1–12.

Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R. 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci* **106:** 3976–3981.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415.

ANNEXE 2 :

Insights into the regulation of small RNA expression: SarA represses the expression of two sRNAs in *Staphylococcus aureus*.

Mauro, T., Rouillon, A. & Felden, B.

Nucleic Acids Research, (2016), 44(21), 10186-10200

10186–10200 *Nucleic Acids Research, 2016, Vol. 44, No. 21* *Published online 4 September 2016*
doi: 10.1093/nar/gkw777

Insights into the regulation of small RNA expression: SarA represses the expression of two sRNAs in *Staphylococcus aureus*

Tony Mauro[†], Astrid Rouillon[*,†] and Brice Felden[*]

Inserm U835, Biochimie Pharmaceutique, University of Rennes 1, 35000 Rennes, France

Received February 5, 2016; Revised August 9, 2016; Accepted August 24, 2016

ABSTRACT

The opportunistic pathogen *Staphylococcus aureus* expresses transcription factors (TFs) and regulatory small RNAs (sRNAs) which are essential for bacterial adaptation and infectivity. Until recently, the study of *S. aureus* sRNA gene expression regulation was under investigated, but it is now an expanding field. Here we address the regulation of Srn_3610_SprC sRNA, an attenuator of *S. aureus* virulence. We demonstrate that SarA TF represses *srn_3610_sprC* transcription. DNase I footprinting and deletion analyses show that the SarA binding site on *srn_3610_sprC* belongs to an essential 22 bp DNA region. Comparative analysis also revealed another possible site, this time in the *srn_9340* promoter. SarA specifically binds these two sRNA promoters with high affinity *in vitro* and also represses their transcription *in vivo*. Chromatin immunoprecipitation (ChIP) assays confirmed SarA attachment to both promoters. ChIP and electrophoretic mobility shift assays targeting σ^A RNA polymerase subunit or using bacterial RNA polymerase holoenzyme suggested that SarA and the σ^A bind *srn_3610_sprC* and *srn_9340* promoters in a mutually exclusive way. Beyond the mechanistic study of SarA repression of these two sRNAs, this work also suggests that some *S. aureus* sRNAs belong to the same regulon and act jointly in responding to environmental changes.

INTRODUCTION

Staphylococcus aureus is an opportunistic pathogen involved in a wide spectrum of diseases, including but not limited to skin infections, pneumonia, endocarditis, osteomyelitis, indwelling device infections, food poisoning and meningitis (1). Approximately 20% of the human population has the bacteria on their skin and nostrils, creating a reservoir of asymptomatic carriers (2). At the same time, infections caused by antibiotic-resistant *S. aureus* strains have severely increased, impacting civil and military healthcare systems worldwide (3). *S. aureus* is remarkable in its aggressiveness and in its resistance potential to many antibiotics. Its versatility depends on its ability to sense and respond to environmental changes by modulating gene expression, using transcription factors (TFs) and regulatory small RNAs (sRNAs). The expression of *S. aureus* virulence factors is tightly controlled by multiple regulators, including TFs, two-component systems and sRNAs (4–6). Many of these regulators are therefore essential for infection (7).

To start bacterial transcription, direct interaction is required between the RNA polymerase's sigma factors and selected gene promoters. Initiation of RNA synthesis is a highly regulated process involving TFs that bind these gene promoters (8). One of the well-studied TFs in *S. aureus* is SarA, a 14.7-kDa DNA-binding protein acting as a dimer. It is one of 11 members of the SarA protein family, with SarR, SarS, SarT, SarU, SarV, SarX, SarY, SarZ, Rot and MgrA. They all have a winged helix motif (9–12) that is required for binding AT-rich double-stranded DNA sequences such as promoters (11,13–15). Directly or indirectly, SarA influences the transcription of at least 120 genes in *S. aureus* (16), and it can either stimulate (with *cna*, *fnbA*, *agr*, *hla*, *fnta*, *fnb* and *sec*) or repress (with *sarV*, *aur*, *sspa*, *spa*, *rot* and *SarS*) target gene expression (9,17–20).

In addition to TFs, *S. aureus* expresses about 160 regulatory RNAs (4,21), all recently compiled into a staphylococcal regulatory RNA database (SRD) (22). sRNAs are expressed from both the core and variable accessory genomes, the latter including pathogenicity islands (PIs) and transposons (23). Among these *S. aureus* sRNAs, only a handful has identified functions including certain ones that influence staphylococcal virulence in animal models of infection (24,25). In *S. aureus*, sRNAs usually pair with target mRNAs for regulation, influencing mRNA expression by modulating its degradation and/or its translation (26). Not only is the *S. aureus* sRNome poorly described, but little is

[*]To whom correspondence should be addressed. Tel: +33 2 2323 4851; Fax: +33 2 2323 4456; Email: brice.felden@univ-rennes1.fr
Correspondence may also be addressed to Astrid Rouillon. Tel: +33 2 2323 4394; Fax: +33 2 2323 4456; Email: astrid.rouillon@univ-rennes1.fr
[†]These authors contributed equally to the paper as first authors.

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

known about the regulation of its expression. An exception to this is RNAIII, which is directly regulated by AgrA (27) and indirectly by SarA (28). AgrA also directly represses the expression of ArtR sRNA (29), termed 'Srn_4050' in SRD. Conversely, RNAIII and ArtR negatively regulate the Rot and SarT TFs, respectively (4,30,31), and RNAIII stabilizes *mgrA* mRNA thereby increasing MgrA TF production (32). Since TFs have numerous targets, sRNA regulation of TF expression allows sRNAs to play an important role in adaptation. These few examples highlight the need for investigation into the regulation of *S. aureus* sRNAs.

The small PI rna C ('Srn_3610' in the SRD, and referred to herein as 'Srn_3610_SprC') is located in the νSAβ PI (23) which contains several virulence factors. We recently showed that Srn_3610_SprC attenuates *S. aureus* virulence and spread in an animal model of infection (33). Those observations imply that its expression must decrease during infection. Interestingly, *srn_3610_sprC* expression drastically decreases after host cell internalization. This suggests that the gene's expression is tightly regulated during bacterial infection. Here we therefore investigated the molecular basis of this regulation *in vivo* and *in vitro,* and demonstrated that SarA restrains *srn_3610_sprC* expression through its direct interaction with the sRNA promoter. The DNA sequence required for SarA interaction with the *srn_3610_sprC* promoter was uncovered by DNase I footprinting assays and sequential deletions. The sequence was then used *in silico* to search for similarities in the *S. aureus* genome. This led to the identification of *srn_9340,* a second sRNA gene whose expression is repressed by SarA. Chromatin immunoprecipitation assays (ChIP) demonstrated SarA binding onto *srn_3610_sprC* and *srn_9340* promoters in living bacteria. ChIP experiments targeting σ^A further revealed that SarA prevents σ^A binding onto the *srn_3610_sprC* promoter. Analysis of the *srn_3610_sprC* and *srn_9340*/SarA binding sequences and the electrophoretic mobility shift assay (EMSA) data suggested that SarA transcriptional repression is similar for both, with SarA binding the promoter thus preventing σ^A binding. In this report, we demonstrate that the SarA TF inhibits transcription of two sRNAs located in the same PI. Such results suggest that these sRNAs have similar functions and belong to a common regulatory network. If applied to other sRNAs, our approach could reveal sRNA clusters regulated by dedicated TFs.

While sRNAs are key adaptation modulators, to our knowledge, none of them have been described as essential genes in *S. aureus*. This suggests some redundancy in their actions, or a lack of knowledge about when they might be absolutely essential for the bacteria. Therefore, studying sRNA transcriptional regulation could give clues to these key moments, while allowing for the identification of clusters of sRNAs with similar functions.

MATERIALS AND METHODS

Bacterial strains and culture conditions

Strains, primers and plasmids are listed in Supplementary Tables S1, 2 and 3, respectively. All mutants from the SarA family were obtained through phage transduction from 6390 *S. aureus* deleted strains to strain HG003 (34). *Escherichia coli* strains were grown at 37°C in LB (MoBio), and 50 μg/ml ampicillin or kanamycin antibiotics were added when necessary. *S. aureus* strains were grown at 37°C in either brain-heart infusion (BHI) medium or tryptic soy broth (TSB; both from Oxoid), with antibiotics added when needed (10 μg/ml erythromycin, chloramphenicol, or tetracycline, or else 250 μg/ml kanamycin). All experiments were done on the HG003 *S. aureus* strain (35), using the RN4220 *S. aureus* strain as an intermediate (36).

Genetic manipulation

Bioinformatic analysis. The *S. aureus* NCTC 8325 genome (taxid: 93061) was used for all genetic locations.

Reporter gene experiments. To create pCN41c, the *ermR* erythromycin resistance gene from plasmid pCN41 (37) was replaced by the *cat194* chloramphenicol resistance gene. In pCN41c-P_{sprC}, the *srn_3610_sprC* promoter sequence was amplified from *S. aureus* HG003 genomic DNA using 5′_BamHI_P*sprC* and P*sprC*_EcoRI_3′ primers (Supplementary Table S2) digested and inserted into pCN41c.

Complementation studies. A 1624-bp fragment made up of the SarA endogenous promoter gene and terminator was amplified by polymerase chain reaction (PCR) using 5′_BamHI_P*sarA* and *sarA*_KpnI_3′ primers, then inserted into pCN36 (37).

ChIP experiments. Three PCR steps were required to construct pCN38-SarA6His. The following primers were used: 5′_PstI_PsarA; PsarA_BamHI_3′; 5′_BamHI_sarA; sarA6his_KpnI_3′; 5′_KpnI_TsarA; and TsarA_KpnI_3′. These PCR products were combined to form a 1936-bp fragment containing *sarA* with an in-frame tag made of 6× Histidine (6His) at the 3′ end of the open reading frame (ORF). The final SarA6His is controlled by its endogenous promoter and terminator.

RNA isolation

The protocol for total RNA extraction was adapted from Le Pabic et *al.* (33). Overnight cultures of *S. aureus* were diluted to an $OD_{600\,nm}$ of 0.1 in fresh BHI or TSB broth, then incubated at 37°C and 160 rpm. Hourly, cells were harvested and resuspended in lysis buffer (20 mM sodium acetate, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.5% sodium dodecyl sulphate (SDS), pH 5.5). Cells were broken out using glass beads (Sigma-Aldrich) in a FastPrep 120 cell disrupter (MP Biomedicals). Total RNA were extracted using phenol/chloroform and precipitated in 100% ethanol supplemented with 0.3 M sodium acetate.

Northern blotting

Total RNA were separated on either denaturing polyacrylamide gels or on agarose gels, then transferred onto Zeta-Probe GT (Bio-Rad) or Nytran membranes (Schleicher & Schuell), respectively. All northern blots were revealed using the Digoxigenin method per the manufacturer's instructions (Roche). We used 20 nt probes previously marked with 3′-end DIG labeling. Signal acquisition was done using an

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

ImageQuant LAS 4000 imager (GE Healthcare). Approximate quantifications were calculated using the ImageQuant software and normalized against transfer-messenger RNA (tmRNA) or 16S ribosomal RNA.

RNA end mapping by RACE

Rapid amplification of cDNA end (RACE) experiments were performed as previously described (38). R1 _srn_9340 primers was used to reverse transcribe or Srn_9340 RNA. Two PCR reactions were performed with Taq polymerase (Invitrogen) using R2-F1 and R2-F2 primer pairs, respectively. PCR products were cloned using a pGEM vector system (Promega), transformed in XL1-Blue *E. coli* and sequenced with M13 reverse primer (Invitrogen, ThermoFisher Scientific) using MegaBACE DNA sequencers (Amersham).

Protein extraction, cell fractionation and western blotting

For protein extraction throughout bacterial growth, cells were resuspended in lysis buffer (50 mM Tris–HCl, pH 7.5, 3 mM $MgCl_2$, 0.1 mg/ml lysostaphin), incubated for 20 min at 37°C, then kept on ice. As measured with Qubit fluorometric quantification (Invitrogen), equal amounts of total proteins were loaded on 15% polyacrylamide gels. Samples were separated on Tris-Glycine sodium dodecyl sulphate-polyacrylamide gel electrophoresis gels and transferred onto Hybond P PVDF membranes (Amersham). SarA proteins were revealed using rabbit anti-SarA antibodies. An ECL Prime Western Blotting Detection kit (Amersham) and an ImageQuant LAS 4000 imager were used to reveal proteins. To ensure that equal protein amounts were loaded on each lane, duplicates of these gels were stained with SYPRO Ruby (Bio-Rad) as per the manufacturer's instructions.

Functional complementation and reporter gene experiments

Overnight cultures were adjusted to an OD_{600nm} of 0.1. For each time point, cells were centrifuged and resuspended in 1× phosphate-buffered saline to obtain an equal cell density throughout the assay. Cell lysis was performed using 0.7 mg/ml lysostaphin, 0.2 U/μl benzonase and 0.1 mM $MgCl_2$ at 37°C for 20 min. After adding 0.25 mg/ml nitrocefin (a ß-lactamase substrate), ß-lactamase activity was quantified on a BioTek instrument every 10 min for 40 min at a wavelength of 492 nm. This activity was normalized against protein quantities as determined by a Bradford assay.

Purification of SarA from *E. coli*

The *sarA* coding sequence was inserted into pET42a in-frame with the 6His N-terminal tag (provided by Marc Hallier) and transformed in BL21 *E. coli*. SarA expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested, washed and resuspended in lysis buffer (10 mM HEPES, pH 7.5, 500 mM NaCl). Purification was done as previously described (39).

Electrophoretic mobility shift assays (EMSA)

Probes were amplified from the *S. aureus* HG003 genome by PCR using specific primers ($sprC_{P267}$, $16S_{P267}$, srn_9340_{P225} and $16S_{225}$), or obtained by hybridization of complementary oligonucleotides ($sprC_{47}$, srn_9340_{51}, $sprC_{30left}$, $sprC_{30middle}$, $sprC_{30right}$, srn_9340_{30left}, $srn\_9340_{30middle}$ and $srn\_9340_{30right}$). To construct $sprC_{P267\Delta47}$, $srn\_9340_{P225\Delta51}$, $sprC_{P267\Delta22}$ and $srn\_9340_{P225\Delta23}$, a three-step PCR was necessary. This used the following combinations of four primers: 5′_P267Δ47*sprc,* P267Δ47*sprC*_3′, 5′_P*sprC* and *sprC*_seq_3′ for $sprC_{P267\Delta47}$; 5′_P225Δ51*srn_9340,* P225Δ51*srn_9340*_3′, 5′_Psrn_9340 and *srn_9340*_seq_3′ for $srn\_9340_{P225\Delta51}$; 5′_P267Δ22*sprc,* P267Δ22*sprC*_3′, 5′_P*sprC* and *sprC*_seq_3′ for $sprC_{P267\Delta22}$; and 5′_P225Δ23*srn_9340,* P225Δ23*srn_9340*_3′, 5′_Psrn_9340 and *srn_9340*_seq_3′ for $srn\_9340_{P225\Delta23}$. DNA probes were labeled with [γ-^{32}P] adenosine triphosphate (ATP) using T4 polynucleotide kinase (New England Biolabs). Binding reactions were carried out as previously described (40). Briefly, binding reaction medium (10 fmol DNA template, 20 mM HEPES-KOH, pH 7.9, 20% glycerol, 0.2 mM EDTA, 0.1 M KCl, 0.1 M $MgCl_2$, 5 mM DTT and 0.2 μg poly(dI-dC)), was incubated when necessary with purified SarA or the *E. coli* RNA polymerase holoenzyme (New England Biolabs) for 30 min at 30°C. Samples were loaded on native polyacrylamide gels. Detection was done with a Typhoon FLA 9500 (GE Healthcare).

The dissociation constant (K_d) was determined by comparing the free versus the shifted DNA–protein bands (ImageQuant software), and apparent K_d was derived from the computed concentration at half-maximal binding.

DNase I footprinting assay

DNA probes were radiolabeled at one end. We used the HindIII restriction site, which is present in the *srn_3610_sprC* gene or in the 3′ primer for *srn_9340*. Klenow Fragment (New England Biolabs) filled the restriction site with [α-^{32}P] ATP and other unlabeled nucleotides. Binding reactions were done as previously described, then subjected to 2 or $7.5.10^{-2}$ unit of DNase I (Invitrogen) for 1 min at 30°C after supplementing the protein buffer with 40 mM $CaCl_2$. Reactions were halted by adding a stop buffer (10 μg/ml Invitrogen Proteinase K, 400 mM acetate sodium, 0.2% SDS, 10 Mm EDTA and 50 μg/ml yeast tRNA). After 15 min at 55°C of proteinase K action, the digested DNA samples were extracted by phenol/chloroform and ethanol precipitation. Pellets were resuspended in formamide dye mix and after denaturation were loaded on 8% denaturing polyacrylamide gels.

DNA sequencing

DNA probes used in the footprinting assays were sequenced according to the standard protocol (41). Specific primers were end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase (New England Biolabs). The ratio of G, A, T and C nucleotides was adjusted in the sequencing mixture.

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

Chromatin immunoprecipitation assays

ChIP experiments were adapted from Faith et *al.* (42). HG003Δ*sarA* carrying either pCN38 or pCN38-SarA6His was cultured for 2.5 h at 37°C and 160 rpm, then treated with formaldehyde. Washed cells were thawed, incubated for 30 min at 37°C with 50 μg lysostaphin in a buffer (200 mM Tris, pH 8.0, 600 mM NaCl, 4% Triton X-100 and 1 mM PMSF), then sonicated. Dilution buffer was added to supernatant (see recipe in Supplementary Table S4) and precleared overnight at 4°C with end-over-end rotations with Protein G sepharose beads previously coated with 1% BSA and 0.1 μg salmon sperm. Input, ChIP, DIG, and Pol fractions were separated. The ChIP, DIG, and Pol fractions had 8 μg of antibodies to 6His, digoxigenin, and *E. coli* σ^{70} added, respectively, then they were incubated for 8 h at 4°C with 'end-over-end rotation. Immunoprecipitations were carried out overnight on Protein G magnetic beads previously coated with 1% bovine serum albumin (BSA) and 0.1 μg salmon sperm. Bead washes and DNA elution were also done as described by Faith (42)**.** ChIP analysis was performed by qPCR with comparative enrichment of the P_{sprC}, P_{srn_9340} and P_{hla} promoters. For normalization, 25 ng yeast plasmid DNA was added to all samples before phenol–chloroform extraction.

In vitro transcription assays

$SprC_{456}$ and srn_9340_{337} templates (corresponding to the sequences from promoter to terminator of genes of interest) were amplified from *S. aureus* genome by PCR using specific primers (5′_BamHI_PsprC, 3′_EcoRI_TsprC and 5′_BamHI_Psrn_9340, 3′_EcoRI_Tsrn_9340). *In vitro* transcription was realized adding 100 fmol of *E. coli* RNA polymerase (New England Biolabs) to 10 fmol of DNA template ($sprC_{456}$ or srn_9340_{337}), in 5× *E. coli* RNA polymerase reaction buffer (New England Biolabs) supplementing with nucleoside triphosphate ATP, CTP, GTP (0.75 mM each) and 0.5 μCi of [α-^{32}P] UTP. A total of 5 pmol of purified SarA protein were added to reactions, when needed. Samples were loaded on 8% denaturing polyacrylamide gels. Detection was done with a Typhoon FLA 9500 (GE Healthcare).

Statistical analysis

For statistical analysis, the one-tailed Mann–Whitney test was performed on three independent experiments and used to evaluate the significance of the ChIP assays. The two-tailed Mann–Whitney test was used for reporter gene experiments. Data were expressed as means ± standard deviations.

RESULTS

Srn_3610_SprC attenuates *S. aureus* virulence and host cell uptake. Its expression sharply decreases after *S. aureus* phagocytosis by human macrophages (33). We therefore investigated whether a TF might be involved in controlling *srn_3610_sprC* expression.

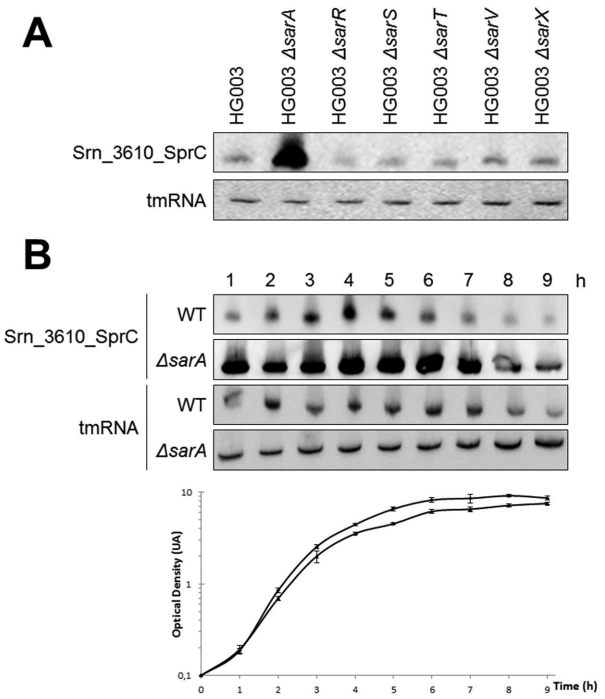


Figure 1. *srn_3610_sprC* expression increases in the absence of SarA. (**A**) *Srn_3610_sprC* expression in different *Staphylococcus aureus* strains lacking transcription factors (TFs). Northern blot analysis of the Srn_3610_sprC and transfer-messenger RNA (tmRNA) transcripts in the *S. aureus* HG003 strain and in isogenic mutants for the SarA protein family. All samples were harvested after 5 h of growth. A total of 10 μg of total RNA were analyzed, and the Digoxigenin method (Roche) was used to reveal Srn_3610_sprC and tmRNA (loading control). (**B**) Northern blot analysis of Srn_3610_SprC and tmRNA transcript expression in *S. aureus* HG003 and isogenic HG003Δ*sarA*. Samples were harvested at the specified time point, and tmRNA was used as a control. In a graph showing optical density as a function of time, the growths of HG003 (diamonds) and HG003Δ*sarA* (triangles) are represented in a semi-logarithmic scale.

The SarA transcription factor reduces levels of Srn_3610_SprC RNA during *S. aureus* growth

To initiate the study of *srn_3610_sprC* transcriptional regulation, we used Northern blotting to monitor RNA levels in six *S. aureus* HG003-derived strains. Each strain was deleted for a member of the sarA family: sarA, sarR, sarS, sarT, sarV or sarX (Supplementary Table S1) (12,19,43,44). The strains were grown in BHI medium, and total RNA extracted after the exponential phase (5 h of growth, OD_{600} ~6). A DNA probe specific to Srn_3610_SprC was used to compare Srn_3610_SprC RNA levels between the mutants and the parental strain (Figure 1A).

A significant 10-fold increase was detected in the strain deleted for *sarA,* suggesting that SarA is the main repressor of *srn_3610_sprC* transcription. In contrast, a slight (about 30%) decrease was detected in the HG003Δ*sarR* strain, with no significant change observed in the other mutants (Figure 1A). The contradictory effects of SarA and SarR deletion make sense, since SarR represses SarA and because these regulators have antagonistic roles in target gene transcription (28).

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

To verify the hypothesis that SarA represses *srn_3610_sprC*, we monitored sRNA gene expression all along growth in HG003 and HG003Δ*sarA* strains. The growth of the HG003Δ*sarA* strain was slightly weaker than that of the parental one (Figure 1B), which is consistent with SarA's physiological role in regulating the expression of at least 120 genes in *S. aureus* (16). In the HG003 strain, Srn_3610_SprC levels fluctuate during growth, and after 4–5 h there was a peak corresponding to about a 3-fold increase (Figure 1B). In the *sarA* mutant, *srn_3610_sprC* expression was ∼10× higher at all collected time points, with maximum RNA levels detected after 4–5 h of growth (Figure 1B). Together, these results confirm that SarA has an essential role in *in vivo* regulation of *srn_3610_sprC*. However, these results also imply that other protein(s) than SarA participate(s) to the fluctuation of *srn_3610_sprC* expression during growth.

SarA negatively regulates srn_3610_sprC transcription

To provide further experimental support for a functional link between Srn_3610_SprC and SarA, we (i) did a *trans*-complementation assay and (ii) verified whether SarA acts on *srn_3610_sprC* at the transcriptional level. To these aims, in each of the following experiments, HG003 and HG003Δ*sarA* strains were co-transformed with a combination of two plasmids. The first plasmid carried *sarA* under control of its own promoter (pCN36-SarA), or was left empty (pCN36). The second carried *blaZ,* encoding the ß-lactamase reporter, either promoterless (pCN41c) or under the control of the *srn_3610_sprC* promoter (pCN41c-P_{sprC}). P_{sprC} was defined as a 144 nt-long DNA fragment (SAOUHSC_01956:18611729_186268) ending at the *srn_3610_sprC* 5′ end, as determined by RACE mapping (33). With pCN41c-P_{sprC}, a measure of the ß-lactamase activity in the presence or absence of SarA reflect regulation of this latter on the *srn_3610_sprC* promoter, thus on the transcription level of *srn_3610_sprC*. Total RNAs and proteins were extracted and expression of both endogenous Srn_3610_SprC and endogenous/exogenous SarA revealed by northern and western blots. As shown in Figure 2A, expression of pCN41c-P_{sprC} did not affect Srn_3610_SprC nor SarA endogenous levels (compare lanes 1 and 2 or 3 and 4). Conversely, exogenous expression of *sarA* through pCN36-sarA led to a significant increase in both SarA mRNA and protein levels in HG003 and HG003Δ*sarA* strains (lanes 3 and 4 versus 1 and 2, and 7 and 8 versus 5 and 6). From this result, we infer that both SarA mRNAs and the encoded protein are overexpressed in cells containing 'low copy' plasmid pCN36-SarA.

More importantly, northern blot analyses showed an ∼9-fold increase in Srn_3610_SprC level in the HG003Δ*sarA* strain as compared to HG003 (Figure 2A, lanes 1 and 2 versus 5 and 6), whereas reintroducing *sarA* restored endogenous Srn_3610_SprC level (Figure 2A, lanes 1 and 2 versus 7 and 8), validating that complementation occurred. Interestingly, when *sarA* was overexpressed in HG003 strain, Srn_3610_SprC level slightly decreased (lanes 1 and 2 versus 3 and 4), indicating an inverse correlation between the evolution of SarA and Srn_3610_SprC RNA levels.

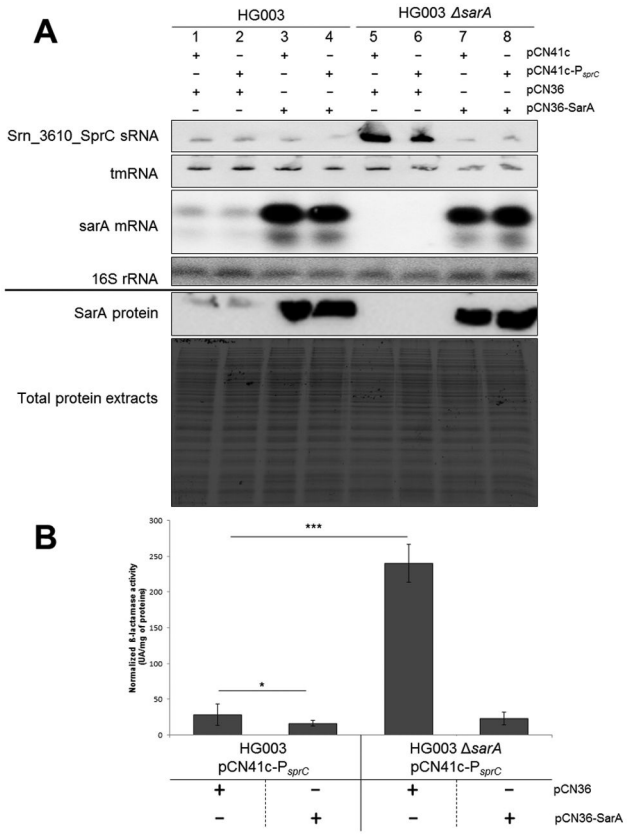


Figure 2. SarA represses *srn_3610_sprC* expression. (**A**) *Staphylococcus aureus* HG003 and HG003Δ*sarA* strains were co-transformed with: pCN36/pCN41c empty vectors (lanes 1 and 5); pCN36/pCN41c-P_{sprC} (lanes 2 and 6); pCN36-SarA/pCN41c (lanes 3 and 7); and pCN36-SarA/pCN41- P_{sprC} (lanes 4 and 8). Srn_3610_SprC and SarA transcripts levels were assessed by northern blot. A 16S rRNA probe was used to reflect total RNA loading for each lane. In parallel, SarA protein levels were checked by western blot (middle panel) and SYPRO Ruby staining of total protein extracts was done to compare loaded protein levels (bottom). (**B**) Effects of SarA on transcriptional activity of the *srn_3610_sprC* promoter (P_{sprC}). *S. aureus* HG003 wild-type strain (HG003) and a strain lacking *sarA* (HG003Δ*sarA*) were co-transformed with pCN41c or pCN41c-P_{sprC} and either pCN36 or pCN36-SarA. P_{sprC} activity was estimated by measuring ß-lactamase substrate hydrolysis. For each lane, indicated ß-lactamase activity was normalized by subtracting the background signal from the same strain where pCN41c-P_{sprC} was replaced by pCN41c empty vector (not shown). Three independent experiments were done, error bars show ± standard deviation. (Mann–Whitney test; $*P < 0.05$; $***P < 0.001$.)

We then measured ß-lactamase activity in all strains. The results for the strains expressing pCN41c-P_{sprC} are depicted in Figure 2B. First, confirming the Northern blot experiment, the ß-lactamase activity underwent a 40% drop when *sarA* was overexpressed (Figure 2B, second bar versus first bar; $P < 0.017$). Second, a 9-fold increase in ß-lactamase activity was measured in HG003Δ*sarA*_pCN36 compared to HG003_pCN36 (third bar versus first bar; $P < 0.0006$), while ß-lactamase activity in HG003Δ*sarA*_pCN36-sarA strain remained at level similar to that of the parental strain (fourth bar versus first bar). All together, these results confirmed that SarA negatively controls Srn_3610_SprC levels, and implied that SarA acts on the *srn_3610_sprC* promoter to reduce RNA expression. The correlation in both

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

ß-lactamase and Northern blot experiments indicates that SarA's repressive action is restricted to the promoter of the sRNA gene.

SarA interacts with the *srn_3610_sprC* promoter *in vitro*

The above results indicate that SarA probably represses *srn_3610_sprC* expression by acting on the *srn_3610_sprC* promoter. To further test this hypothesis, EMSA studies were done with recombinant SarA and the *srn_3610_sprC* promoter region. For these assays, we used a 6His-tagged full-length SarA with a 267-bp *srn_3610_sprC* DNA region ($sprC_{P267}$) that contains the 144 bp *srn_3610_sprC* promoter region along with 123 bp downstream of the +1 *srn_3610_sprC* transcription start. $sprC_{P267}$forms a complex with purified SarA and has an apparent K_d of 86 ± 3 nM (Figure 3A). A 10-fold excess of unlabeled $sprC_{P267}$ efficiently disrupt SarA from a preformed complex of labeled SarA/$sprC_{P267}$. However, a 100-fold excess of a 267 bp unspecific fragment made up of a section of 16S DNA ($16S_{267}$), was unable to do so (Figure 3B). This demonstrates the specificity of the complex formation between SarA and $sprC_{P267}$. Together, these results indicate that purified SarA directly binds the $sprC_{P267}$ promoter *in vitro* and its binding is specific.

The SarA binding site on the *srn_3610_sprC* promoter extends through the transcribed sequence

To further explore the SarA binding site on the *srn_3610_sprC* promoter, we used 3′-end labeled $sprC_{P267}$ probes with increasing concentrations of purified SarA to perform DNase I footprint assays. As expected, increasing amounts of SarA revealed a region protected from DNAse I cleavage within the *srn_3610_sprC* DNA sequence (Figure 3C).This 47 bp-long sequence is located between positions −32 and +15 in *srn_3610_sprC* (nts 1861616-1861570) and we named it '$sprC_{47}$' (Figure 3D).

To see whether this protected sequence was necessary for SarA binding, we created $sprC_{P267\Delta47}$, a 220 bp DNA fragment made from nucleotides −144 to +123 but with the footprint-identified $sprC_{47}$ sequence deleted (Figure 3D and E). EMSA testing of $sprC_{P267\Delta47}$ showed that SarA could not form a complex with the deleted promoter, even though it could with the native one (Figure 3E). These results confirm the key role of $sprC_{47}$ in SarA/P_{sprC} complex formation.

sprC $_{47}$ is sufficient for SarA binding onto *srn_3610_sprC* promoter

To test whether the 47-bp protected DNA region was sufficient for SarA-specific binding, EMSA was performed using $sprC_{47}$ or $Random_{47}$, a scrambled 47-bp DNA sequence. A 50-fold excess of unlabeled $sprC_{47}$ disrupts the preformed complex, whereas the same amount of $Random_{47}$ does not (Figure 3F), so SarA specifically binds $sprC_{47}$. Therefore, $sprC_{47}$ is necessary and sufficient for SarA binding.

srn_3610_sprC promoter dissection

We divided the $sprC_{47}$ sequence into three overlapping 30 bp sequences and challenged them using EMSA. Two of the overlapping DNA sequences which share 22 bp allowed SarA binding and the third did not (Supplementary Figure S1). We thus created a 245 bp DNA fragment from nts −144 to +123 without the shared $sprC_{22}$ section, resulting in the mutant $sprC_{P267\Delta22}$ (Figure 3D). EMSA was performed between $sprC_{P267}$ and purified SarA protein, and $sprC_{P267\Delta22}$ was used as a competitor to assess the SarA/$sprC_{P267}$ complex formation. The 22 bp sequence is necessary for SarA binding, since a 25-fold excess of unlabeled $sprC_{P267\Delta22}$ did not disrupt the $sprC_{P267}$/SarA complex, whereas the same amount of unlabeled $sprC_{P267}$ did (Figure 3G).

srn_9340 is another sRNA regulated by SarA

Using $sprC_{47}$ as a query, we did comparative sequence analysis on the HG003 strain's genome. This identified a 29 bp sequence sharing 93% identity with $sprC_{47}$ (Figure 4). Interestingly, the sequence is located upstream of *srn_9340* (22), an sRNA gene identified by RNAseq (45) and recently validated by northern blot under the name S774 (46). As for *srn_3610_sprC*, the *srn_9340* (NCTC 8325: 1863905–1863790) is located in the SaPIn3 (vSaß) PI. It is only 2338 bp downstream from *srn_3610_sprC* (1861729-1862268) and both sRNAs are expressed from the same genomic DNA strand (Figure 4, minus strand). Based on the homology between *srn_3610_sprC* and *srn_9340* promoter sequences, we hypothesized that *srn_9340* transcription might also be negatively controlled by SarA.

SarA represses *Srn_9340* transcription

Two ∼110 nt-long transcripts (Figure 5A) were detected in HG003 by northern blots using a *srn_9340*-specific probe. These transcripts were detected throughout bacterial growth (Supplementary Figure S2). In the HG003*ΔsarA* mutant, *srn_9340* expression is about 3.5× higher than in the isogenic HG003 strain (Figure 5A). Thus, *srn_9340* transcription seems to be repressed by SarA, albeit to a lesser extent than SarA represses *srn_3610_sprC*. While *srn_9340* seems to be expressed at lower level, compared to *srn_3610_sprC*, their expression patterns are very similar in HG003. They reach peak expression after 3–5 h of growth and expression levels fluctuate up to 3-fold. These observations suggest that the two sRNAs are analogously controlled. Moreover, these results also imply that other protein(s) than SarA participate(s) to the fluctuation of *srn_9340* expression during growth.

Going further, Srn_9340 levels were monitored by northern blots in the HG003*ΔsarR*, *ΔsarS*, *ΔsarT*, *ΔsarV* and *ΔsarX* strains. We observed a reduction of about 50% of *srn_9340* expression in HG003*ΔsarR,* but the levels were unaffected in the other strains (Supplementary Figure S3). These results are similar to those obtained for Srn_3610_SprC (Figure 1A), confirming the roles of SarA and SarR as antagonists for target gene transcription. We compared the HG003*ΔsarA* strain transformed with pCN36 or complemented by pCN36-*sarA,* and the results showed that the variation in *srn_9340* expression levels is controlled by SarA (Supplementary Figure S4). The Srn_9340 transcript's 5′ ends were mapped by RACE at the same genomic position (1863905), implying that they share

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

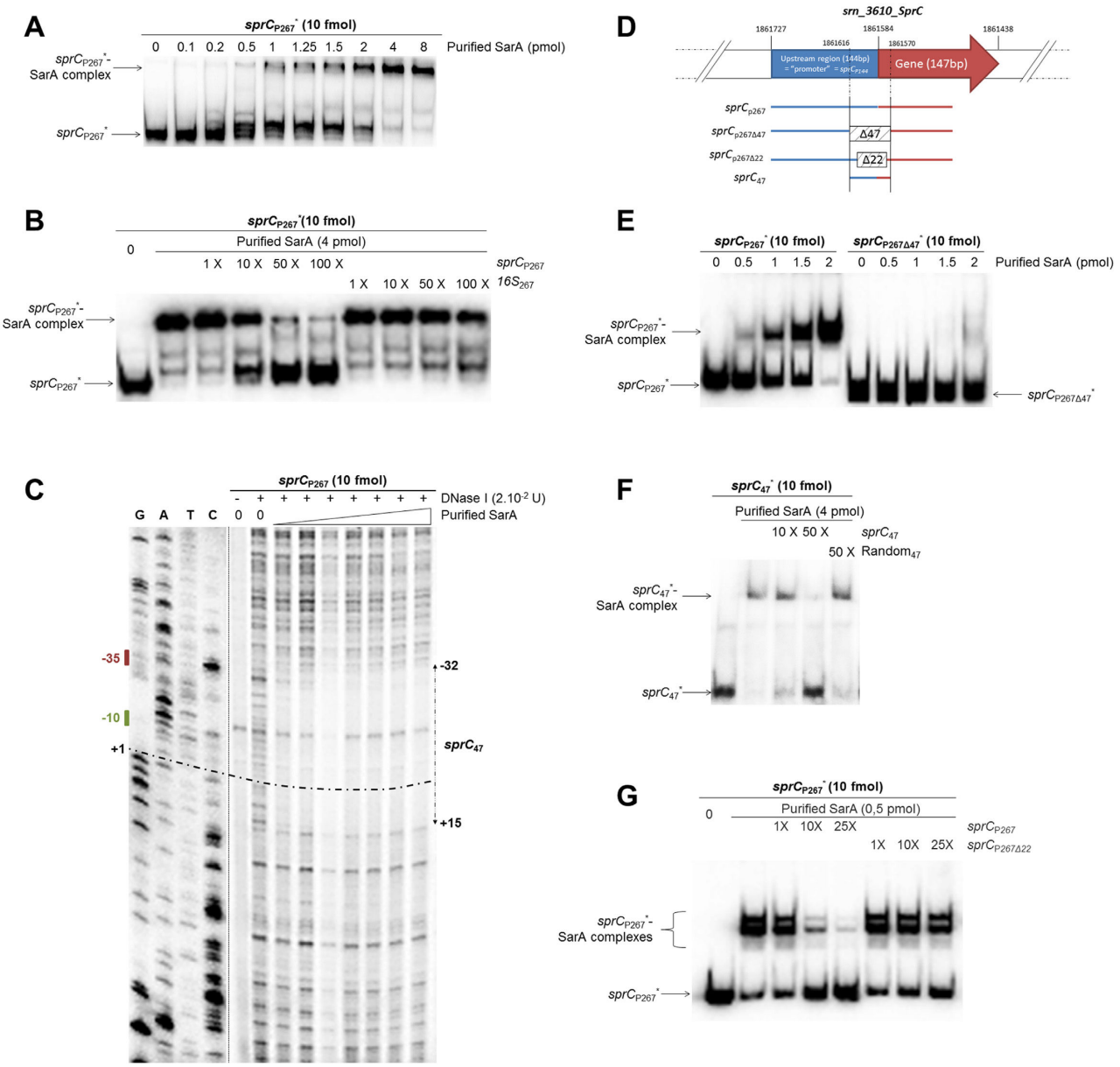


Figure 3. SarA specifically binds a region overlapping the *srn_3610_sprC* promoter and its 5′ end. (**A**) SarA binds the *srn_3610_sprC* promoter *in vitro*. Electrophoretic mobility shift assays (EMSA) were done using 10 fmol $sprC_{P267}$ probe, a ^{32}P-labeled *srn_3610_sprC* promoter fragment and increasing amounts of 0.1 to 8 pmol 6His-tagged SarA. (**B**) SarA specifically binds $sprC_{P267}$. EMSA was performed with 4 pmol of SarA and with increasing amounts of specific (unlabeled $sprC_{P267}$) or non-specific competitors (unlabeled $16S_{267}$). The $sprC_{P267}$/SarA complex is only inhibited in the presence of 50–100× excesses of the specific competitor. (**C**) The SarA binding site overlaps the *srn_3610_sprC* promoter and gene. DNase I footprinting assays were performed in the presence of 10 fmol $sprC_{P267}$, 2.10^{-2} U DNase I, and increasing amounts (0.5–2 pmol) of 6His-SarA. Lanes 1–4 correspond to $sprC_{P267}$ sequencing, and the promoter region is annotated. The vertical dotted arrow indicates the *srn_3610_sprC* region protected from DNase I degradation. The SarA binding site (−32 to +15) was named '$sprC_{47}$.' (**D**) Schematic representation of the different DNA probes used for EMSA studies. (**E**) The 47 bp region protected by SarA is necessary for SarA binding with *srn_3610_sprC*. EMSA were realized as in *A*, using ^{32}P-labeled $sprC_{P267\Delta47}$ (right) or wild-type ^{32}P-labeled $sprC_{P267}$ (left). SarA was unable to form a complex with $sprC_{P267\Delta47}$. (**F**) The 47 bp protected by SarA is sufficient for SarA binding onto *srn_3610_sprC*. EMSA were done using 10 fmol $sprC_{47}$, 4 pmol of 6His-tagged SarA, and increasing amounts of either unlabeled $sprC_{47}$ as a specific competitor or unlabeled 47-bp sequence ($Random_{47}$) as a non-specific competitor. SarA forms a complex with DNA (up-shifted band) that can only be disrupted by an excess of the specific competitor. (**G**) A 22-bp SarA binding site on $sprC_{P267}$ was found with EMSA done in the presence of 0.5 pmol of sarA and with increasing amounts of unlabeled $sprC_{P267}$ or unlabeled $sprC_{P267\Delta22}$.

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

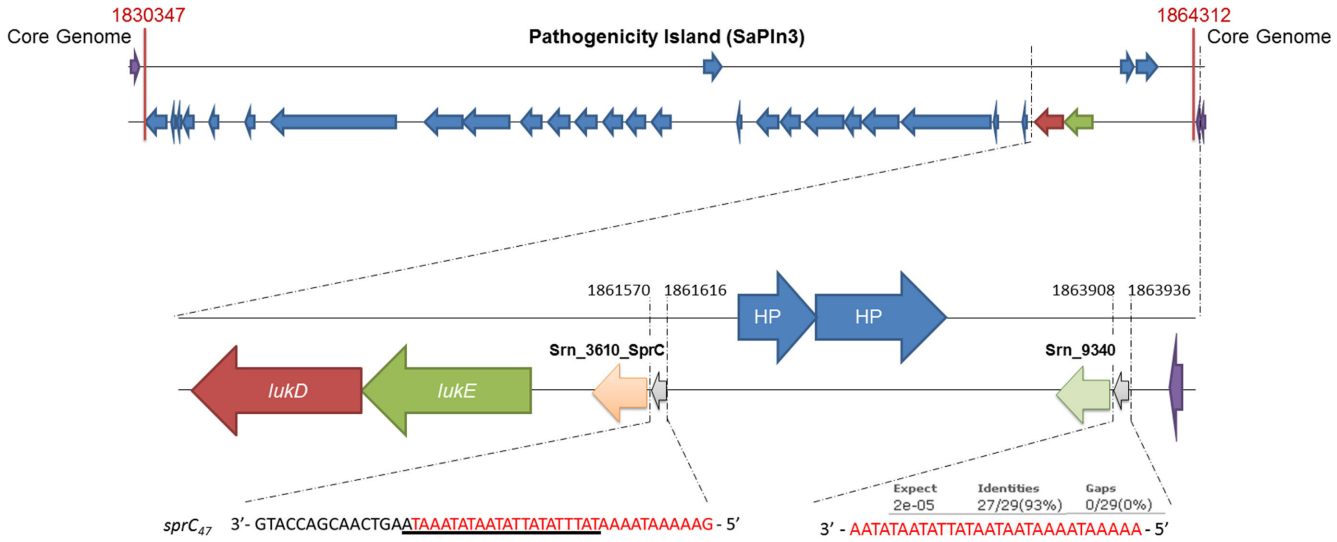


Figure 4. *srn_9340* possesses a putative binding site for SarA. *Srn_3610_sprC* belongs to the pathogenicity island SaPIn3. $SprC_{47}$, which can interact with SarA, possesses a 22-nt palindromic sequence (underlined). Blast analysis using $sprC_{47}$ as the query led to the discovery of another putative SarA binding site within SaPIn3. This sequence was identified as part of the promoter of another small RNA, Srn_9340.

the same transcription start site (TSS). The longer transcript's 3′ end maps at genomic position 1863790, whereas that of the shorter transcript ends at 1863797 (Figure 5B).

Consistent with the northern blot results (Figure 5A), the Srn_9340 transcripts are about 116 and 108 nts long (Figure 5B), with the shorter transcript lacking nucleotides at the 3′ end. No ORF was found in either sequence. The Srn_9340 transcript sizes obtained are slightly different from that of S774 (46). In addition to the size difference, Mäder *et al.* visualized a unique transcript. These differences could have been due to either genetic (HG001 versus HG003) or technical differences (gel resolution) (46).

SarA binds the *Srn_9340* promoter *in vitro*

We used EMSA for binding assays of purified SarA on the *srn_9340* promoter region (Figure 5C). We synthesized a PCR fragment of *srn_9340* nucleotides −165 to +60 (with the coding sequence starting at +1), and named this 'srn_9340_{P225}.' We observed a distinct SarA/srn_9340_{P225} complex with an apparent K_d of 11 ± 0.2 nM (Figure 5C, middle arrow). The complex has an abundance that is positively linked to SarA quantities. A second band shift then appeared which probably correspond to protein aggregation onto the DNA probe (Figure 5C, upper band). A 10-fold excess of srn_9340_{P225}effectively competed with the labeled promoter fragment for SarA binding, whereas a 100-fold excess of $16S_{225}$, an unlabeled non-specific 225 pb 16S DNA fragment, did not displace the labeled one from the complex (Figure 5D). Together, these results indicate that SarA directly and specifically binds the *srn_9340* promoter *in vitro*.

SarA binds the *srn_9340* promoter upstream from the +1 transcription start

We performed DNase I footprinting assays using a 3′-end labeled srn_9340_{P225} probe with increasing concentrations of purified SarA protein. As expected, without SarA, there was no protection against DNase I cleavage. Increasing SarA amounts revealed an *srn_9340* region that became protected from DNase I cleavages (Figure 5E). This 51 bp-long sequence is positioned between −51 and −1 within the *srn_9340* promoter (1863956–1863904), and we named it 'srn_9340_{51}' (Figure 5F).

To test whether this protected sequence is necessary and sufficient for SarA binding, we amplified $srn\_9340_{P225\Delta51}$, a 174 bp DNA fragment (nts −165 to +60) that does not include the srn_9340_{51} sequence (Figure 5F). EMSA performed on $srn\_9340_{P225\Delta51}$ showed that SarA complex formation was greatly reduced as compared to EMSA performed with the native promoter (Figure 5G). To test whether this region is sufficient for SarA binding, EMSA was performed with srn_9340_{51}. Indeed, SarA specifically binds srn_9340_{51}(Figure 5H), therefore, srn_9340_{51} is necessary and sufficient for SarA binding. Furthermore, a 10-fold excess of unlabeled srn_9340_{51} effectively competed with the labeled srn_9340_{P225} for SarA binding, whereas a 50-fold excess of $Random_{51}$, an unlabeled non-specific DNA fragment, did not disrupt the labeled one from the complex (Figure 5I). Considered together, these results confirm the requirement for srn_9340_{51} in complex formation between SarA and the *srn_9340* promoter.

srn_9340 promoter dissection

To further delineate the SarA binding site on the srn_9340 promoter, the srn_9340_{51} sequence was divided into three overlapping 30-bp sequences, then tested for SarA binding by EMSA. The results (see Supplementary Figure S5) clearly indicate the importance of a specific 23 bp sequence. A 25-fold excess of unlabeled $srn\_9340_{P225\Delta23}$ did not disrupt the srn_9340_{P225}/SarA complex, whereas the same amount of unlabeled srn_9340_{P225} did (Figure 5J). Therefore, removing this 23 nt sequence from the 225 nt-long

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

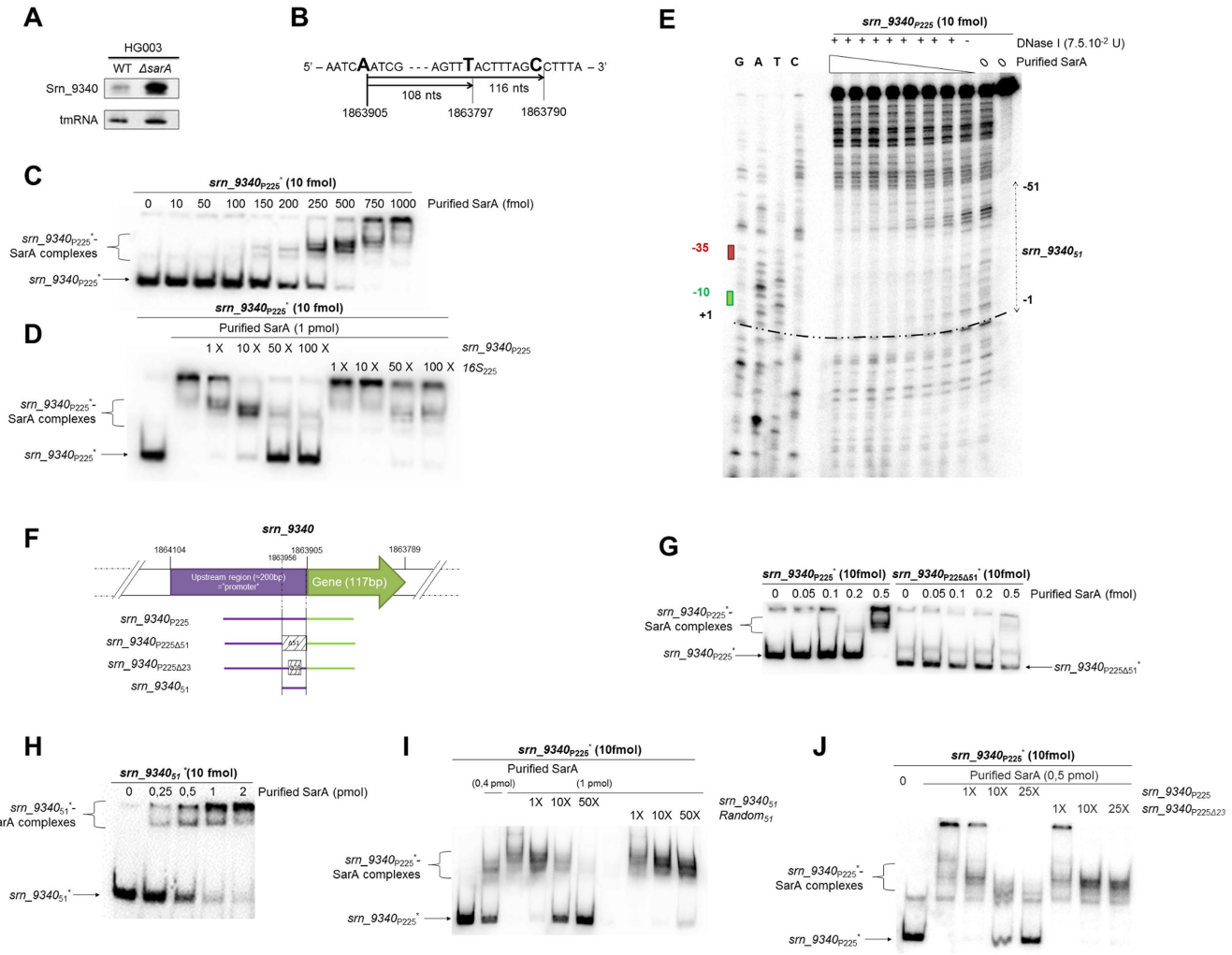


Figure 5. SarA binds *srn_9340* on its promoter and represses transcription. (**A**) SarA represses *srn_9340* expression. After 5 h of growth, 10 μg of total RNA were obtained from *Staphylococcus aureus* HG003 and HG003Δ*sarA*. Srn_9340 expression levels were monitored by northern blot. (**B**) Schematic representation of both the 5′ and 3′ ends of the Srn_9340 transcripts (long and short small RNAs), as determined by rapid amplification of cDNA ends (RACE). (**C**) SarA binds the *srn_9340* promoter *in vitro*. An EMSA was done using 10 fmol of a ^{32}P-labeled *srn_9340* promoter fragment (srn_9340_{P225}) as a probe in the presence of increasing amounts (0.01–1 pmol) of 6His-tagged SarA. Srn_9340_{P225} forms an initial complex with 0.15 pmol of SarA, and a second one at 0.75 pmol. (**D**) SarA specifically binds srn_9340_{P225} *in vitro*. EMSA was done with 10 fmol of ^{32}P-labeled srn_9340_{P225}, 1 pmol of 6His-SarA and increasing amounts of specific (unlabeled srn_9340_{P225}) or non-specific competitors (unlabeled $16S_{225}$). (**E**) DNase I footprinting assays were performed in the presence of 10 fmol srn_9340_{P225}, $7.5.10^{-2}$ U DNase I and increasing amounts (0.1–0.5 pmol) of 6His-SarA. The region protected by SarA is indicated with a vertical dotted arrow, and the numbers indicate relative positions to the previously determined transcription start site (TSS). Lanes G, A, T and C correspond to sequencing. The nucleotides from −51 to −1 (srn_9340_{51}) are protected by SarA against DNase I degradation. (**F**) Schematic representation of the DNA probes used for EMSA studies. (**G**) Deletion of the protected SarA sequence from the *srn_9340* promoter region abolishes SarA's capacity to bind the *srn_9340* promoter. (**H**) A 51 bp protected sequence is sufficient for SarA binding. EMSA were done using 10 fmol srn_9340_{51} in the presence of increasing amounts of 6His-tagged SarA (0.25–2 pmol). (**I**) Srn_9340_{51} competes with srn_9340_{225} for SarA binding. EMSA was performed with 10 fmol of ^{32}P-labeled srn_9340_{P225}, 0.4 or 1 pmol of 6His-SarA and increasing amounts of specific (unlabeled srn_9340_{51}) or non-specific competitors (unlabeled $Random_{51}$). (**J**) The 23 bp SarA binding site on srn_9340_{225} was confirmed by EMSA done in the presence of 0.5 pmol sarA and with increasing amounts of unlabeled srn_9340_{225} or unlabeled $srn\_9340_{225\Delta 23}$.

DNA promoter prevents it from competing against the preformed SarA/*srn_9340* complex, which implies that this region is mandatory for binding SarA.

SarA binds *srn_3610_sprC* and *srn_9340 in vivo*

In ChIP experiments, cellular macromolecular interactions are frozen and analyzed by protein IP and qPCR. Among other uses, this method allows for the comparison of enriched DNA sequences bound to a protein. We performed ChIP assays on HG003Δ*sarA* carrying either pCN38 or pCN38-SarA6His. We used 6His antibodies to immunoprecipitate SarA6His, with *hla* as a positive control because SarA directly regulates *hla* transcription (18,47). IP induced about a 19-fold enrichment in P_{sprC} (Figure 6A), while in the same extracts the *srn_9340* and *hla* promoters were respectively enriched about 6- and 16-fold. In the presence of SarA (HG003Δ*sarA*_ pCN38-SarA6His), the average P_{sprC} retrievals in the SarA-IP experiments were statistically higher than without it (HG003Δ*sarA*_pCN38). The obtained en-

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

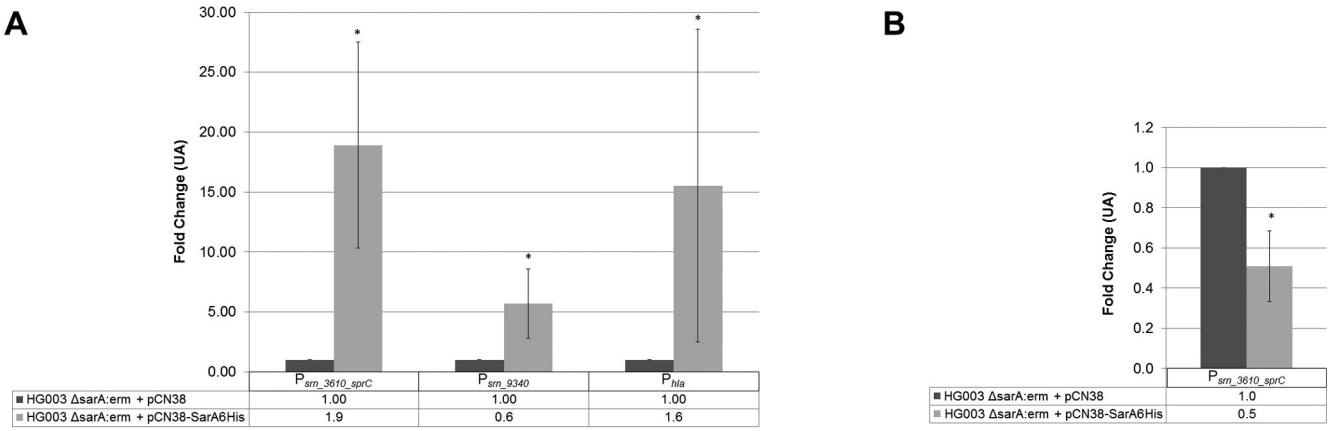


Figure 6. *In vivo* analysis of the srn_3610_sprC and srn_9340 promoters: SarA versus σ^A binding. The *Staphylococcus aureus* HG003Δ*sarA* strain was transformed either with an empty pCN38 vector (dark gray) or with pCN38-SarA6His (light gray). Chromatin immunoprecipitation (ChIP) experiments were conducted using antibodies for 6His-tag (**A**) or for the σ^{70}/σ^A RNA polymerase subunit (**B**). DNA enrichment was assessed by qPCR analysis using specific primers for the promoters of *srn_3610_sprC* (left panel), *srn_9340* (middle) or *hla* (right). Values beneath the bars indicate fold change between specific and non-specific interactions. (Mann–Whitney U test; three independent experiments; $*P < 0.05$). (A) *srn_3610_sprC*, *srn_9340* and *hla* promoters co-immunoprecipitate with SarA6His. (B) SarA6His disturbs σ^A binding on the *srn_3610_sprC* promoter.

richments of P_{hla} and P_{srn_9340} were also statistically significant. These results confirm the presence of SarA on both the *srn_3610_sprC* and *srn_9340* promoters *in vivo*.

In *S. aureus*, σ^A is the functional analog of σ^{70} in *E. coli.* This sigma factor is the RNA polymerase subunit involved in promoter recognition (48). Both *S. aureus* σ^A and *E.* coli σ^{70} RNA polymerase subunits recognize canonical promoters and are interchangeable (49). It has already been reported that *srn_9340* transcription depends on σ^A (46). After promoter sequence analysis, we predicted that the *srn_3610_sprC* promoter must be σ^A-dependent. We used *E. coli* σ^{70} antibodies to immunoprecipitate the *S. aureus* σ^A, investigating its accessibility at P_{sprC} and P_{srn_9340} with and without SarA. $sprC_{47}$ and srn_9340_{51} correspond to the DNA sequences protected by SarA, and they both contain the −10 TATA box and partial or complete −35 box RNA polymerase binding sites. Therefore, SarA should prevent binding of staphylococcal RNA polymerase onto both sRNA promoters. We found that the average amount of *srn_3610_sprC* promoter in the σ^{70}-IP with SarA (HG003*ΔsarA*_pCN38-SarA6His) was statistically lower than that of the deleted strain (HG003*ΔsarA*_pCN38) (Figure 6B). The *srn_3610_sprC* promoter amounts in the ChIP targeting σ^A were also lower in the presence of SarA than without it ($p < 0.05$). Meanwhile, the presence or absence of SarA did not make a statistical difference in the amount of the *srn_9340* promoter recovered after σ^{70}-IP enrichment (HG003*ΔsarA*_ pCN38-SarA6His and HG003*ΔsarA*_pCN38, respectively) (Supplementary Figure S6).

Since in both promoters the SarA binding site includes the RNA polymerase loading location, we expected a similar repression model. We therefore supposed that the σ^{70}-IP ChIP experiment was not useful for the study of P_{srn_9340}. This led us to further explore the genomic environment of *srn_9340*. The highly transcribed gene $tRNA^{ser}$ is 408 bp away from *srn_9340*. Such a close proximity (Supplementary Figure S6) could lead to sonicated DNA fragments containing *srn_9340* and $tRNA^{ser}$, making it difficult to use ChIP data to draw any conclusions. Thereafter, we performed EMSA studies on purified SarA and bacterial RNA polymerase *(E. coli)* together with either *srn_3610_sprC* or the *srn_9340* promoters.

SarA binding onto *srn_3610_sprC* and *srn_9340* promoters prevents RNA polymerase loading

Purified *E. coli* RNA polymerase holoenzyme binds the two $sprC_{P267}$ and srn_9340_{P225}promoters (Figure 7A and B) with about 0.75 and 1.8 nM affinities *in vitro*. To see whether SarA might prevent RNA polymerase binding onto these promoters, another set of SarA EMSA experiments was performed with and without the RNA polymerase (Figure 7C and D). Similar patterns were observed for both promoters in the presence of both SarA and RNA polymerase (Figure 7C and D). Multiple bands corresponding to SarA/DNA low molecular weight complexes appeared, and in the presence of RNA polymerase, high molecular weight complexes emerged. These latter complexes migrate similarly as DNA/polymerase complex (compare lanes 3, 4, 5 with, respectively, lanes 6, 7, 8), allowing the conclusion that these complexes only contain the RNA polymerase and the sRNA promoter. Moreover, using 100 fmol of RNA polymerase in the presence of SarA, we still observed a lower molecular weight sarA/promoter complexes, whereas 50 fmol of RNA polymerase is sufficient to bind nearly all $sprC_{P267}$ and srn_9340_{P225}(Figure 7A and B). However, even when two-fold higher RNA polymerase is added in the presence of SarA, SarA-promoter complexes are still detected (Figure 7C and D). This suggests that DNA/SarA and DNA/RNA polymerase complexes co-exist, implying that when SarA is loaded onto each of the two sRNA promoters, RNA polymerase stops being able to bind. We infer from these results that SarA binding onto $srn_9340_{P225}$and $sprC_{P267}$ promoters hampers RNA polymerase binding. To reinforce that, we performed *in vitro* transcription assay using the *E. coli* RNA polymerase and each of the DNAs en-

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

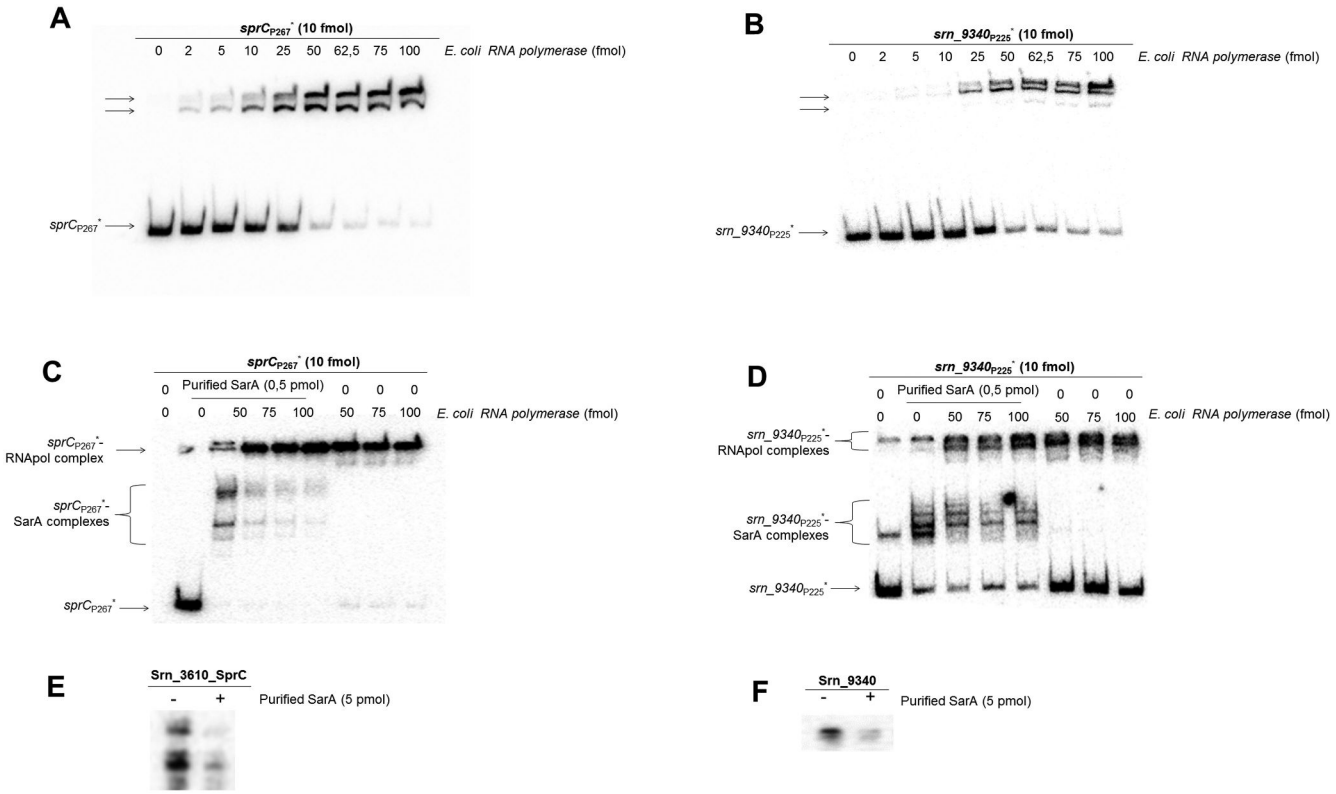


Figure 7. SarA and bacterial RNA polymerase holoenzyme cannot simultaneously bind $sprC_{P267}$ and srn_9340_{P225}. EMSA was done using 10 fmol ^{32}P-labeled $sprC_{P267}$ or srn_9340_{P225} probes with increasing amounts (2–100 fmol) of purified RNA polymerase. The binding capacity of *Escherichia coli* RNA polymerase was confirmed by EMSA experiments using $sprC_{P267}$ (**A**) or srn_9340_{P225} (**B**). To demonstrate their binding exclusivity on $sprC_{P267}$ (**C**) or srn_9340_{P225} (**D**), 0.5 pmol purified SarA and 50–100 fmol *E. coli* RNA polymerase were simultaneously added to EMSA experiments. *In vitro* transcription was realized using 10 fmol of $sprC_{456}$ (**E**) or srn_9340_{337} (**F**) and 100 fmol of *E. coli* RNA polymerase. A total of 5 pmol of the purified TF SarA were added, when indicated.

compassing the sequences from promoter to the sRNA terminators ($sprC_{456}$ and srn_9340_{337}), in the presence/absence of the SarA protein. As shown in panels E and F from Figure 7, the *E. coli* RNA polymerase is able to produce transcripts from *srn_3610_sprC* and *srn_9340* and SarA protein severely impact these transcriptions. Altogether, it is concluded that TF SarA binds *srn_3610_SprC* and *srn_9340* promoters and prevents transcription.

DISCUSSION

How sRNAs are regulated in *S. aureus* remains an essential but unanswered and barely addressed question. Most documented cases of sRNA regulation are in enterobacteria, where sRNAs are transcriptionally controlled by TFs and *vice versa* (50). In *S. aureus*, sRNAs also regulate the expression of TFs. For example, RNAIII represses Rot TF translation (30) and RsaA (Srn_1510) represses MgrA synthesis (25). To learn about sRNA gene regulation in *S. aureus*, we investigated the regulation of *srn_3610_sprC* expression. This sRNA was used as a model since it acts as a virulence attenuator, modulating host phagocytosis, therefore we expected its expression to be tightly controlled. The pleiotropic TF SarA is involved in *S. aureus* biofilm formation and pathogenesis, among other things (51–53). In our study, this TF was identified as a negative regulator of transcription initiation in *srn_3610_sprC* and *srn_9340*. Whereas the absence of SarA correlates with a change in the expression level of at least 120 genes, it has only been shown to directly control the expression of a dozen mRNAs. Here we show that SarA is directly involved in transcriptional regulation of Srn_3610_SprC and Srn_9340 sRNAs. To our knowledge this is the first identification of SarA as a transcriptional repressor of sRNA. Beyond just a mechanistic description of a direct interaction between a TF and an sRNA promoter, our work has identified two sRNAs whose expressions are under control of one general TF. Having arrived with an exogenous PI, these two sRNAs are now negatively regulated by a TF from the core genome. This demonstrates *S. aureus*' ability to take over sRNA regulation. Such a mechanism probably strengthens *S. aureus*' infectivity.

Based on the similarities between SarA repression of *srn_3610_sprC* and *srn_9340* transcription, their comparable expression patterns during bacterial growth, and their close genetic locations, we can guess that these two sRNAs are in part functionally related. Previous studies have shown that *srn_3610_sprC* expression is tightly regulated during host immune cell recognition of the bacteria for clearance during infection (33), while no functions have yet been identified for Srn_9340 (22,46). Our work connects the control of the expression of these two sRNAs to a common TF, sug-

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

gesting that Srn_9340 might also be involved in virulence and host immune response.

Without SarA, *srn_9340* and *srn_3610_sprC* transcription levels are respectively 3- and 9-folds higher, indicating that their expression should be repressed during bacterial growth. A TF downregulating sRNA gene expression could, upon TF release from the sRNA promoters, allow transient expression of sRNAs. This will enable the bacteria to respond rapidly and efficiently to environmental changes. SarA repression of *srn_9340* and *srn_3610_sprC* was confirmed by *trans*-complementation. In that experiment, higher *sarA* mRNA and protein levels were detected in the HG003*ΔsarA*_pCN36-SarA strain than in the isogenic strain. However, *srn_3610_sprC* and *srn_9340* expression levels did not decline further (Figure 2B and Supplementary Figure S4). This might be because the endogenous levels of SarA are sufficient to repress the transcription of both sRNAs. It also suggests that a phenotype linked to these sRNAs might be detected by stimulating their expression *in vivo* with an SarA-independent promoter.

Our study points out SarA binding differences and similarities for both promoters. Purified SarA protein specifically binds *srn_3610_sprC* and *srn_9340* promoters *in vitro*, with apparent binding constants of about 90 and 10 nM, respectively. These K_ds are within the same range as those previously reported for other SarA mRNA targets (9,54,55). This suggests that, *in vitro*, SarA binds target sRNA and mRNA gene promoters with similar efficiencies. Gel retardation (EMSA) assays between SarA and srn_9340_{P225}revealed multiple retarded bands, as previously observed for SarA and other target mRNA promoters, e.g. on the *tst* toxic shock promoter (10,54). Depending on gel composition and migration conditions, $sprC_{P267}$/SarA complexes appear as either double or single bands (Figures 3E, G and 6C). This suggests the coexistence of multiple conformations of the DNA/SarA complexes, as expected for a dimeric protein with a binding site made up of two anchor points (56).The multiple complexes detected *in vitro* could also be related to several SarA binding sites or SarA could act as an architectural accessory protein rather than as a canonical TF (57). Architectural DNA-binding proteins like these influence genomic DNA superhelicities and modify the number of base pairs per helical turn. Depending on the location of its binding sites, SarA could modify the spacing between the −10 and −35 boxes. This could lead the promoter to either have an optimal conformation for transcription initiation, or, conversely, an unfavorable spacing inducing transcription repression (58,59).

DNase I protection assays between SarA and $sprC_{P267}$ identified a 47-bp protected DNA sequence located at positions −35 to +12. EMSA confirmed that this sequence contains the SarA binding site. When this section was deleted, SarA did not bind to the *srn_3610_sprC* promoter region, while the DNA fragment was sufficient for SarA binding *in vitro*, indicating that it is necessary and sufficient for SarA recruitment. As for SarA binding onto the *srn_9340* promoter, the protected section is a 51 bp sequence located upstream from the +1 transcriptional start site. This sequence is necessary and sufficient for SarA binding onto the *srn_9340* promoter. Previous studies revealed that the SarA region protected against DNAse I cleavages ranges from 31 to 144 bp (10,18,20). Such differences can be excused: some protected sequences corresponded to multiple SarA binding sites, thus the 47/51 bp SarA protected regions are within the ranges previously reported.

The SarA binding sequences were able to be reduced to about 22/23 nts long after EMSA studies using 30-nt overlapping DNA sequences together with longer sequences interrupted by internal deletion (Figures 3G and 5J). Figure 8 sums up the experimental evidence we collected showing SarA TF binding onto the two sRNA promoters, along with the existing information from the literature. The SarA binding site identified on the sarA P3 promoter is 26 bp-long, a similar size to the 22/23 bp sequence identified on the two sRNA promoters. Moreover, the 22/23 nt sequences are similar to a previously reported 26 bp consensus sequence that is required for SarA binding (18). This consensus sequence is not always detected within the promoters of the genes directly regulated by SarA. SarA binding studies used SELEX to identify a 7 bp segment required for SarA binding *in vitro* (56). That same sequence was present (with a mismatch tolerance of one) at least once within the 150 bp upstream from the TSS in 72 of 102 SarA-regulated genes (16). The analysis of the $sprC_{47}$ and the srn_9340_{51} SarA-protected sequences revealed two adjacent palindromes (Figure 8). These palindromes could anchor the SarA dimer onto the *srn_9340* promoter (59). SarA binds DNA as a dimer (20), and although the presence of a palindromic sequence is not a prerequisite for dimer binding, it may facilitate sarA recruitment. In *S. aureus*, the CodY TF acts as a dimer and also binds palindromic sequences (60). Furthermore, the SarA-regulated sRNAs promoters $sprC_{22}$ and srn_9340_{23} each contains part of a palindrome, and their absence in $sprC_{P267\Delta22}$and $srn\_9340_{P225\Delta23}$ prevents SarA from binding *in vitro*. Altogether, our data are consistent with previous studies on SarA binding and provide further evidence that sarA binding requires an ATTTTAT sequence in its target's promoter (56). Moreover, SarA binding covers the TATA box and −35 region of *sarA* mRNA (10), both necessary sequences for σ^A subunit RNA polymerase binding prior to transcription initiation (61). SarA may repress sRNA promoter transcription through its ability to compete with the RNA polymerase for promoter binding, as proposed for mRNAs regulated by SarA (56).

We showed with ChIP that SarA binds the *srn_3610_sprC* and *srn_9340* promoter regions *in vivo* (Figure 6A). These experiments also revealed that when the RNA polymerase subunit σ^A binds the *srn_3610_sprC* promoter, P_{sprC} enrichment with anti σ^{70} antibodies was statistically higher when SarA was deleted. Therefore, SarA and σ^A probably bind P_{sprC} in a mutually exclusive manner. We performed EMSA competition studies with the *srn_3610_sprC* and *srn_9340* promoters, purified bacterial RNA polymerase and SarA. These experiments suggested that RNA polymerase holoenzyme and SarA cannot bind the two promoters simultaneously. This is consistent with the fact that they both bind to the same area within P_{sprC}. ChIP data for the *srn_9340* promoter region using anti-σ^{70} antibodies were variable from one experiment to another (Supplementary Figure S6). The σ^A-ChIP experimental procedure was not useful for the *srn_9340* promoter due to the proximity of an-

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017



Figure 8. Analysis of the SarA-protected sequences on the *srn_3610_sprC* and *srn_9340* promoters. The alignment of the $sprC_{47}$ and srn_9340_{51} sequences was performed using Clustal Omega software (http://www.ebi.ac.uk/Tools/msa/clustalo/). Boxes outline the −35 (yellow) and −10 (red) boxes. The TSS is indicated with a black arrow; the SarA binding site defined by Sterba *et al.* (5′ATTTTAT3′) is shown with dotted arrows (reverse complementary sequences); palindromic sequences are depicted by large arrows; and identical nucleotides are identified with stars. The 22 nt of $sprC_{47}$ and the 23 nt of *srn_9340* that are necessary for SarA binding are bold.

other highly transcribed gene ($tRNA^{ser}$) within the 400 bp surrounding *srn_9340*.Through σ^A binding onto $tRNA^{ser}$, immunoprecipitation with anti-σ^{70} antibodies resulted in a DNA fragment containing *srn_9340*. Therefore, the link between σ^A enrichment of the *srn_9340* promoter and the presence or absence of SarA is not relevant. This highlights the difficulty of ChIP analysis using antibodies directed against housekeeping proteins such as σ factor when done in bacterial genomes with closely located genes. We propose that SarA binds the *srn_3610_sprC* and *srn_9340* promoters and prevents RNA polymerase binding. *In vitro* transcription assays, in the presence/absence of purified SarA, at each sRNA promoter, confirmed that hypothesis since the presence of SarA drastically reduced the synthesis of each sRNA.

We report here that sarA-mediated transcriptional repression of P_{sprC} and P_{srn_9340} is effective throughout bacterial growth. External signals, possibly during infection and phagocytosis, may affect SarA binding onto the two sRNA promoters. This could be through post-translational modifications (62,63) or targeted degradation, with these changes inducing the release of SarA from the sRNA promoters, thus allowing their fast and efficient transcription. In the presence of oxidants, high levels of *srn_3610_sprC* hamper survival (33), implying that the SarA-mediated transcriptional control of this sRNA is essential during oxidative stress. These recent observations suggest that *srn_3610_sprC* expression levels must be tightly regulated, requiring timely transcription derepression for tasks that remain unidentified.

In conclusion, as far as we know these two *S. aureus* sRNAs gene negatively regulated by SarA are the first reported examples of joint transcriptional regulation of sRNAs. This may however represent just the tip of the iceberg, especially since coordinated regulation of several sRNAs by a single TF would be a significant advantage for bacterial fitness. We have already begun investigations aimed at exploring the functioning of Srn_9340 and at uncovering the physiological signals that remove SarA from the sRNA promoters to allow their optimal expression.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank our lab colleagues, particularly Y. Augagneur; G. Pascreau, for reading the manuscript and giving advice; M. Hallier, for the strains, plasmids and technical help with SarA purification; and H. Le Pabic for her shared interest in Srn_3610_SprC. We also would like to thank A. Cheung for the SarA family mutants and E. Charpentier for the pCN plasmids.

FUNDING

Institut National de la Santé et de la Recherche Médicale; University of Rennes 1; Région Bretagne (to T.M.); Direction Générale de l'Armement (DGA) (to T.M.). Funding for open access charge: Inserm.
Conflict of interest statement. None declared.

REFERENCES

1. Lowy,F.D. (1998) Staphylococcus aureus infections. *N. Engl. J. Med.*, **339**, 520–532.
2. van Belkum,A., Verkaik,N.J., de Vogel,C.P., Boelens,H.A., Verveer,J., Nouwen,J.L., Verbrugh,H.A. and Wertheim,H.F. (2009) Reclassification of Staphylococcus aureus nasal carriage types. *J. Infect. Dis.*, **199**, 1820–1826.
3. Chambers,H.F. and Deleo,F.R. (2009) Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat. Rev. Microbiol.*, **7**, 629–641.
4. Fechter,P., Caldelari,I., Lioliou,E. and Romby,P. (2014) Novel aspects of RNA regulation in Staphylococcus aureus. *FEBS Lett.*, **588**, 2523–2529.
5. Priest,N.K., Rudkin,J.K., Feil,E.J., van den Elsen,J.M., Cheung,A., Peacock,S.J., Laabei,M., Lucks,D.A., Recker,M. and Massey,R.C. (2012) From genotype to phenotype: can systems biology be used to predict Staphylococcus aureus virulence? *Nat. Rev. Microbiol.*, **10**, 791–797.
6. Thomas,M.S. and Wigneshweraraj,S. (2014) Regulation of virulence gene expression. *Virulence*, **5**, 832–834.
7. Cheung,A.L. and Zhang,G. (2002) Global regulation of virulence determinants in Staphylococcus aureus by the SarA protein family. *Front. Biosci.*, **7**, d1825–1842.
8. Browning,D.F. and Busby,S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57–65.
9. Ballal,A. and Manna,A.C. (2010) Control of thioredoxin reductase gene (trxB) transcription by SarA in Staphylococcus aureus. *J. Bacteriol.*, **192**, 336–345.
10. Cheung,A.L., Nishina,K. and Manna,A.C. (2008) SarA of Staphylococcus aureus binds to the sarA promoter to regulate gene expression. *J. Bacteriol.*, **190**, 2239–2243.
11. Liu,Y., Manna,A.C., Pan,C.H., Kriksunov,I.A., Thiel,D.J., Cheung,A.L. and Zhang,G. (2006) Structural and function analyses of the global regulatory protein SarA from Staphylococcus aureus. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2392–2397.

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

12. Manna,A.C. and Cheung,A.L. (2006) Transcriptional regulation of the agr locus and the identification of DNA binding residues of the global regulatory protein SarR in Staphylococcus aureus. *Mol. Microbiol.*, **60**, 1289–1301.
13. Cheung,A.L., Bayer,M.G. and Heinrichs,J.H. (1997) sar Genetic determinants necessary for transcription of RNAII and RNAIII in the agr locus of Staphylococcus aureus. *J. Bacteriol.*, **179**, 3963–3971.
14. Cheung,A.L., Koomey,J.M., Butler,C.A., Projan,S.J. and Fischetti,V.A. (1992) Regulation of exoprotein expression in Staphylococcus aureus by a locus (sar) distinct from agr. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6462–6466.
15. Cheung,A.L., Nishina,K.A., Trotonda,M.P. and Tamber,S. (2008) The SarA protein family of Staphylococcus aureus. *Int. J. Biochem. Cell Biol.*, **40**, 355–361.
16. Dunman,P.M., Murphy,E., Haney,S., Palacios,D., Tucker-Kellogg,G., Wu,S., Brown,E.L., Zagursky,R.J., Shlaes,D. and Projan,S.J. (2001) Transcription profiling-based identification of Staphylococcus aureus genes regulated by the agr and/or sarA loci. *J. Bacteriol.*, **183**, 7341–7353.
17. Blevins,J.S., Gillaspy,A.F., Rechtin,T.M., Hurlburt,B.K. and Smeltzer,M.S. (1999) The Staphylococcal accessory regulator (sar) represses transcription of the Staphylococcus aureus collagen adhesin gene (cna) in an agr-independent manner. *Mol. Microbiol.*, **33**, 317–326.
18. Chien,Y., Manna,A.C., Projan,S.J. and Cheung,A.L. (1999) SarA, a global regulator of virulence determinants in Staphylococcus aureus, binds to a conserved motif essential for sar-dependent gene regulation. *J. Biol. Chem.*, **274**, 37169–37176.
19. Manna,A.C., Ingavale,S.S., Maloney,M., van Wamel,W. and Cheung,A.L. (2004) Identification of sarV (SA2062), a new transcriptional regulator, is repressed by SarA and MgrA (SA0641) and involved in the regulation of autolysis in Staphylococcus aureus. *J. Bacteriol.*, **186**, 5267–5280.
20. Rechtin,T.M., Gillaspy,A.F., Schumacher,M.A., Brennan,R.G., Smeltzer,M.S. and Hurlburt,B.K. (1999) Characterization of the SarA virulence gene regulator of Staphylococcus aureus. *Mol. Microbiol.*, **33**, 307–316.
21. Guillet,J., Hallier,M. and Felden,B. (2013) Emerging functions for the Staphylococcus aureus RNome. *PLoS Pathog.*, **9**, e1003767.
22. Sassi,M., Augagneur,Y., Mauro,T., Ivain,L., Chabelskaya,S., Hallier,M., Sallou,O. and Felden,B. (2015) SRD: a Staphylococcus regulatory RNA database. *RNA*, **21**, 1005–1017.
23. Pichon,C. and Felden,B. (2005) Small RNA genes expressed from Staphylococcus aureus genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 14249–14254.
24. Chabelskaya,S., Gaillot,O. and Felden,B. (2010) A Staphylococcus aureus small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. *PLoS Pathog.*, **6**, e1000927.
25. Romilly,C., Lays,C., Tomasini,A., Caldelari,I., Benito,Y., Hammann,P., Geissmann,T., Boisset,S., Romby,P. and Vandenesch,F. (2014) A non-coding RNA promotes bacterial persistence and decreases virulence by regulating a regulator in Staphylococcus aureus. *PLoS Pathog.*, **10**, e1003979.
26. Romilly,C., Caldelari,I., Parmentier,D., Lioliou,E., Romby,P. and Fechter,P. (2012) Current knowledge on regulatory RNAs and their machineries in Staphylococcus aureus. *RNA Biol.*, **9**, 402–413.
27. Koenig,R.L., Ray,J.L., Maleki,S.J., Smeltzer,M.S. and Hurlburt,B.K. (2004) Staphylococcus aureus AgrA binding to the RNAIII-agr regulatory region. *J. Bacteriol.*, **186**, 7549–7555.
28. Reyes,D., Andrey,D.O., Monod,A., Kelley,W.L., Zhang,G. and Cheung,A.L. (2011) Coordinated regulation by AgrA, SarA, and SarR to control agr expression in Staphylococcus aureus. *J. Bacteriol.*, **193**, 6020–6031.
29. Xue,T., Zhang,X., Sun,H. and Sun,B. (2014) ArtR, a novel sRNA of Staphylococcus aureus, regulates alpha-toxin expression by targeting the 5′ UTR of sarT mRNA. *Med. Microbiol. Immunol.*, **203**, 1–12.
30. Boisset,S., Geissmann,T., Huntzinger,E., Fechter,P., Bendridi,N., Possedko,M., Chevalier,C., Helfer,A.C., Benito,Y., Jacquier,A. *et al.* (2007) Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.*, **21**, 1353–1366.
31. Geisinger,E., Adhikari,R.P., Jin,R., Ross,H.F. and Novick,R.P. (2006) Inhibition of rot translation by RNAIII, a key feature of agr function. *Mol. Microbiol.*, **61**, 1038–1048.
32. Gupta,R.K., Luong,T.T. and Lee,C.Y. (2015) RNAIII of the Staphylococcus aureus agr system activates global regulator MgrA by stabilizing mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14036–14041.
33. Le Pabic,H., Germain-Amiot,N., Bordeau,V. and Felden,B. (2015) A bacterial regulatory RNA attenuates virulence, spread and human host cell phagocytosis. *Nucleic Acids Res.*, **43**, 9232–9248.
34. Krausz,K.L. and Bose,J.L. (2016) Bacteriophage transduction in Staphylococcus aureus: broth-based method. *Methods Mol. Biol.*, **1373**, 63–68.
35. Herbert,S., Ziebandt,A.K., Ohlsen,K., Schafer,T., Hecker,M., Albrecht,D., Novick,R. and Gotz,F. (2010) Repair of global regulators in Staphylococcus aureus 8325 and comparative analysis with other clinical isolates. *Infect. Immun.*, **78**, 2877–2889.
36. Kreiswirth,B.N., Lofdahl,S., Betley,M.J., O'Reilly,M., Schlievert,P.M., Bergdoll,M.S. and Novick,R.P. (1983) The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature*, **305**, 709–712.
37. Charpentier,E., Anton,A.I., Barry,P., Alfonso,B., Fang,Y. and Novick,R.P. (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl. Environ. Microbiol.*, **70**, 6076–6085.
38. Pinel-Marie,M.L., Brielle,R. and Felden,B. (2014) Dual toxic-peptide-coding Staphylococcus aureus RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep.*, **7**, 424–435.
39. Eyraud,A., Tattevin,P., Chabelskaya,S. and Felden,B. (2014) A small RNA controls a protein regulator involved in antibiotic resistance in Staphylococcus aureus. *Nucleic Acids Res.*, **42**, 4892–4905.
40. Carey,M.F., Peterson,C.L. and Smale,S.T. (2013) DNase I footprinting. *Cold Spring Harb. Protoc.*, **2013**, 469–478.
41. Slatko,B.E., Albright,L.M., Tabor,S. and Ju,J. (2001) DNA sequencing by the dideoxy method. *Curr. Protoc. Mol. Biol.*, doi:10.1002/0471142727.mb0704as47.
42. Faith,J.J., Hayete,B., Thaden,J.T., Mogno,I., Wierzbowski,J., Cottarel,G., Kasif,S., Collins,J.J. and Gardner,T.S. (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, **5**, e8.
43. Manna,A. and Cheung,A.L. (2001) Characterization of sarR, a modulator of sar expression in Staphylococcus aureus. *Infect. Immun.*, **69**, 885–896.
44. Schmidt,K.A., Manna,A.C., Gill,S. and Cheung,A.L. (2001) SarT, a repressor of alpha-hemolysin in Staphylococcus aureus. *Infect. Immun.*, **69**, 4749–4758.
45. Howden,B.P., Beaume,M., Harrison,P.F., Hernandez,D., Schrenzel,J., Seemann,T., Francois,P. and Stinear,T.P. (2013) Analysis of the small RNA transcriptional response in multidrug-resistant Staphylococcus aureus after antimicrobial exposure. *Antimicrob. Agents Chemother.*, **57**, 3864–3874.
46. Mader,U., Nicolas,P., Depke,M., Pane-Farre,J., Debarbouille,M., van der Kooi-Pol,M.M., Guerin,C., Derozier,S., Hiron,A., Jarmer,H. *et al.* (2016) Staphylococcus aureus transcriptome architecture: from laboratory to infection-mimicking conditions. *PLoS Genet.*, **12**, e1005962.
47. Oscarsson,J., Kanth,A., Tegmark-Wisell,K. and Arvidson,S. (2006) SarA is a repressor of hla (alpha-hemolysin) transcription in Staphylococcus aureus: its apparent role as an activator of hla in the prototype strain NCTC 8325 depends on reduced expression of sarS. *J. Bacteriol.*, **188**, 8526–8533.
48. Mondal,R., Ganguly,T., Chanda,P.K., Bandhu,A., Jana,B., Sau,K., Lee,C.Y. and Sau,S. (2010) Stabilization of the primary sigma factor of Staphylococcus aureus by core RNA polymerase. *BMB Rep.*, **43**, 176–181.
49. Belcheva,A., Verma,V., Korenevsky,A., Fridman,M., Kumar,K. and Golemi-Kotra,D. (2012) Roles of DNA sequence and sigma A factor in transcription of the vraSR operon. *J. Bacteriol.*, **194**, 61–71.
50. Wagner,E.G. and Romby,P. (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.*, **90**, 133–208.
51. Beenken,K.E., Blevins,J.S. and Smeltzer,M.S. (2003) Mutation of sarA in Staphylococcus aureus limits biofilm formation. *Infect. Immun.*, **71**, 4206–4211.

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

52. Blevins,J.S., Elasri,M.O., Allmendinger,S.D., Beenken,K.E., Skinner,R.A., Thomas,J.R. and Smeltzer,M.S. (2003) Role of sarA in the pathogenesis of Staphylococcus aureus musculoskeletal infection. *Infect. Immun.*, **71**, 516–523.
53. Valle,J., Toledo-Arana,A., Berasain,C., Ghigo,J.M., Amorena,B., Penades,J.R. and Lasa,I. (2003) SarA and not sigmaB is essential for biofilm development by Staphylococcus aureus. *Mol. Microbiol.*, **48**, 1075–1087.
54. Andrey,D.O., Renzoni,A., Monod,A., Lew,D.P., Cheung,A.L. and Kelley,W.L. (2010) Control of the Staphylococcus aureus toxic shock tst promoter by the global regulator SarA. *J. Bacteriol.*, **192**, 6077–6085.
55. Zhao,Y., Verma,V., Belcheva,A., Singh,A., Fridman,M. and Golemi-Kotra,D. (2012) Staphylococcus aureus methicillin-resistance factor fmtA is regulated by the global regulator SarA. *PLoS One*, **7**, e43998.
56. Sterba,K.M., Mackintosh,S.G., Blevins,J.S., Hurlburt,B.K. and Smeltzer,M.S. (2003) Characterization of Staphylococcus aureus SarA binding sites. *J. Bacteriol.*, **185**, 4410–4417.
57. Fujimoto,D.F., Higginbotham,R.H., Sterba,K.M., Maleki,S.J., Segall,A.M., Smeltzer,M.S. and Hurlburt,B.K. (2009) Staphylococcus aureus SarA is a regulatory protein responsive to redox and pH that can support bacteriophage lambda integrase-mediated excision/recombination. *Mol. Microbiol.*, **74**, 1445–1458.
58. Bronner,S., Monteil,H. and Prevost,G. (2004) Regulation of virulence determinants in Staphylococcus aureus: complexity and applications. *FEMS Microbiol. Rev.*, **28**, 183–200.
59. Schumacher,M.A., Hurlburt,B.K. and Brennan,R.G. (2001) Crystal structures of SarA, a pleiotropic regulator of virulence genes in S. aureus. *Nature*, **409**, 215–219.
60. Stenz,L., Francois,P., Whiteson,K., Wolz,C., Linder,P. and Schrenzel,J. (2011) The CodY pleiotropic repressor controls virulence in gram-positive pathogens. *FEMS Immunol. Med. Microbiol.*, **62**, 123–139.
61. Decker,K.B. and Hinton,D.M. (2013) Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu. Rev. Microbiol.*, **67**, 113–139.
62. Didier,J.P., Cozzone,A.J. and Duclos,B. (2010) Phosphorylation of the virulence regulator SarA modulates its ability to bind DNA in Staphylococcus aureus. *FEMS Microbiol. Lett.*, **306**, 30–36.
63. Sun,F., Ding,Y., Ji,Q., Liang,Z., Deng,X., Wong,C.C., Yi,C., Zhang,L., Xie,S., Alvarez,S. *et al.* (2012) Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15461–15466.

Downloaded from http://nar.oxfordjournals.org/ at BibliothÃ¨que de l'IRMAR on January 17, 2017

RÉFÉRENCES

REFERENCES

- (2009). Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother *53*, 4961-4967.
- Aboshkiwa, M., al-Ani, B., Coleman, G., and Rowland, G. (1992). Cloning and physical mapping of the Staphylococcus aureus rplL, rpoB and rpoC genes, encoding ribosomal protein L7/L12 and RNA polymerase subunits beta and beta'. J Gen Microbiol *138*, 1875-1880.
- Andrey, D.O., Jousselin, A., Villanueva, M., Renzoni, A., Monod, A., Barras, C., Rodriguez, N., and Kelley, W.L. (2015). Impact of the Regulators SigB, Rot, SarA and sarS on the Toxic Shock Tst Promoter and TSST-1 Expression in Staphylococcus aureus. PLoS One *10*, e0135579.
- Andrey, D.O., Renzoni, A., Monod, A., Lew, D.P., Cheung, A.L., and Kelley, W.L. (2010). Control of the Staphylococcus aureus toxic shock tst promoter by the global regulator SarA. Journal of bacteriology *192*, 6077-6085.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev *29*, 231-262.
- Baba, T., Bae, T., Schneewind, O., Takeuchi, F., and Hiramatsu, K. (2008). Genome sequence of Staphylococcus aureus strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. J Bacteriol *190*, 300-310.
- Ballal, A., and Manna, A.C. (2009a). Expression of the sarA family of genes in different strains of Staphylococcus aureus. Microbiology *155*, 2342-2352.
- Ballal, A., and Manna, A.C. (2009b). Regulation of superoxide dismutase (sod) genes by SarA in Staphylococcus aureus. J Bacteriol *191*, 3301-3310.
- Ballal, A., and Manna, A.C. (2010). Control of thioredoxin reductase gene (trxB) transcription by SarA in Staphylococcus aureus. J Bacteriol *192*, 336-345.
- Ballal, A., Ray, B., and Manna, A.C. (2009). sarZ, a sarA family gene, is transcriptionally activated by MgrA and is involved in the regulation of genes encoding exoproteins in Staphylococcus aureus. J Bacteriol *191*, 1656-1665.
- Beaume, M., Hernandez, D., Farinelli, L., Deluen, C., Linder, P., Gaspin, C., Romby, P., Schrenzel, J., and Francois, P. (2010). Cartography of methicillin-resistant S. aureus transcripts: detection, orientation and temporal expression during growth phase and stress conditions. PLoS One *5*, e10725.

- Belcheva, A., Verma, V., and Golemi-Kotra, D. (2009). DNA-binding activity of the vancomycin resistance associated regulator protein VraR and the role of phosphorylation in transcriptional regulation of the vraSR operon. Biochemistry *48*, 5592-5601.
- Benito, Y., Kolb, F.A., Romby, P., Lina, G., Etienne, J., and Vandenesch, F. (2000). Probing the structure of RNAIII, the Staphylococcus aureus agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. RNA *6*, 668-679.
- Bischoff, M., Entenza, J.M., and Giachino, P. (2001). Influence of a functional sigB operon on the global regulators sar and agr in Staphylococcus aureus. J Bacteriol *183*, 5171-5179.
- Biswas, I., Jha, J.K., and Fromm, N. (2008). Shuttle expression plasmids for genetic studies in Streptococcus mutans. Microbiology *154*, 2275-2282.
- Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994). Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell *78*, 889-896.
- Bohn, C., Rigoulay, C., Chabelskaya, S., Sharma, C.M., Marchais, A., Skorski, P., Borezee-Durant, E., Barbet, R., Jacquet, E., Jacq, A.*, et al.* (2010). Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism. Nucleic Acids Res *38*, 6620-6636.
- Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A.*, et al.* (2007). Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev *21*, 1353-1366.
- Bosi, E., Monk, J.M., Aziz, R.K., Fondi, M., Nizet, V., and Palsson, B.O. (2016). Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci U S A *113*, E3801-3809.
- Brosse, A., Korobeinikova, A., Gottesman, S., and Guillier, M. (2016). Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. Nucleic Acids Res *44*, 9650-9666.
- Brunskill, E.W., and Bayles, K.W. (1996). Identification and molecular characterization of a putative regulatory locus that affects autolysis in Staphylococcus aureus. J Bacteriol *178*, 611-618.
- Carey, M.F., Peterson, C.L., and Smale, S.T. (2013). DNase I footprinting. Cold Spring Harb Protoc *2013*, 469-478.

- Case, C.C., Simons, E.L., and Simons, R.W. (1990). The IS10 transposase mRNA is destabilized during antisense RNA control. EMBO J *9*, 1259-1266.
- Ceri, H., Olson, M.E., Stremick, C., Read, R.R., Morck, D., and Buret, A. (1999). The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. J Clin Microbiol *37*, 1771-1776.
- Chabelskaya, S., Bordeau, V., and Felden, B. (2014). Dual RNA regulatory control of a Staphylococcus aureus virulence factor. Nucleic Acids Res *42*, 4847-4858.
- Chabelskaya, S., Gaillot, O., and Felden, B. (2010). A Staphylococcus aureus small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. PLoS Pathog *6*, e1000927.
- Chan, P.F., Foster, S.J., Ingham, E., and Clements, M.O. (1998). The Staphylococcus aureus alternative sigma factor sigmaB controls the environmental stress response but not starvation survival or pathogenicity in a mouse abscess model. Journal of bacteriology *180*, 6082-6089.
- Chen, P.R., Bae, T., Williams, W.A., Duguid, E.M., Rice, P.A., Schneewind, O., and He, C. (2006). An oxidation-sensing mechanism is used by the global regulator MgrA in Staphylococcus aureus. Nat Chem Biol *2*, 591-595.
- Chen, P.R., Nishida, S., Poor, C.B., Cheng, A., Bae, T., Kuechenmeister, L., Dunman, P.M., Missiakas, D., and He, C. (2009). A new oxidative sensing and regulation pathway mediated by the MgrA homologue SarZ in Staphylococcus aureus. Mol Microbiol *71*, 198-211.
- Cheung, A.L., Bayer, A.S., Zhang, G., Gresham, H., and Xiong, Y.Q. (2004). Regulation of virulence determinants in vitro and in vivo in Staphylococcus aureus. FEMS Immunol Med Microbiol *40*, 1-9.
- Cheung, A.L., Chien, Y.T., and Bayer, A.S. (1999). Hyperproduction of alpha-hemolysin in a sigB mutant is associated with elevated SarA expression in Staphylococcus aureus. Infect Immun *67*, 1331-1337.
- Cheung, A.L., Koomey, J.M., Butler, C.A., Projan, S.J., and Fischetti, V.A. (1992). Regulation of exoprotein expression in Staphylococcus aureus by a locus (sar) distinct from agr. Proc Natl Acad Sci U S A *89*, 6462-6466.
- Cheung, A.L., Nishina, K., and Manna, A.C. (2008a). SarA of Staphylococcus aureus binds to the sarA promoter to regulate gene expression. J Bacteriol *190*, 2239-2243.
- Cheung, A.L., Nishina, K.A., Trotonda, M.P., and Tamber, S. (2008b). The SarA protein family of Staphylococcus aureus. Int J Biochem Cell Biol *40*, 355-361.

- Cheung, A.L., Schmidt, K., Bateman, B., and Manna, A.C. (2001). SarS, a SarA homolog repressible by agr, is an activator of protein A synthesis in Staphylococcus aureus. Infect Immun *69*, 2448-2455.
- Cheung, A.L., and Zhang, G. (2001). Are the structures of SarA and SarR similar? Trends Microbiol *9*, 570-573.
- Chevalier, C., Boisset, S., Romilly, C., Masquida, B., Fechter, P., Geissmann, T., Vandenesch, F., and Romby, P. (2010). Staphylococcus aureus RNAIII binds to two distant regions of coa mRNA to arrest translation and promote mRNA degradation. PLoS Pathog *6*, e1000809.
- Chien, Y., and Cheung, A.L. (1998). Molecular interactions between two global regulators, sar and agr, in Staphylococcus aureus. J Biol Chem *273*, 2645-2652.
- Chien, Y., Manna, A.C., Projan, S.J., and Cheung, A.L. (1999). SarA, a global regulator of virulence determinants in Staphylococcus aureus, binds to a conserved motif essential for sar-dependent gene regulation. J Biol Chem *274*, 37169-37176.
- Chunhua, M., Yu, L., Yaping, G., Jie, D., Qiang, L., Xiaorong, T., and Guang, Y. (2012). The expression of LytM is down-regulated by RNAIII in Staphylococcus aureus. J Basic Microbiol *52*, 636-641.
- Cordwell, S.J., Larsen, M.R., Cole, R.T., and Walsh, B.J. (2002). Comparative proteomics of Staphylococcus aureus and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. Microbiology *148*, 2765-2781.
- Costerton, J.W., Geesey, G.G., and Cheng, K.J. (1978). How bacteria stick. Sci Am *238*, 86-95.
- Cramton, S.E., Gerke, C., Schnell, N.F., Nichols, W.W., and Gotz, F. (1999). The intercellular adhesion (ica) locus is present in Staphylococcus aureus and is required for biofilm formation. Infect Immun *67*, 5427-5433.
- Debarbouille, M., Dramsi, S., Dussurget, O., Nahori, M.A., Vaganay, E., Jouvion, G., Cozzone, A., Msadek, T., and Duclos, B. (2009). Characterization of a serine/threonine kinase involved in virulence of Staphylococcus aureus. J Bacteriol *191*, 4070-4081.
- deHaseth, P.L., and Helmann, J.D. (1995). Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol Microbiol *16*, 817-824.
- Deora, R., and Misra, T.K. (1996a). Characterization of the primary sigma factor of Staphylococcus aureus. J Biol Chem *271*, 21828-21834.
- Deora, R., and Misra, T.K. (1996b). Characterization of the primary sigma factor of Staphylococcus aureus. The Journal of biological chemistry *271*, 21828-21834.

- Didier, J.P., Cozzone, A.J., and Duclos, B. (2010). Phosphorylation of the virulence regulator SarA modulates its ability to bind DNA in Staphylococcus aureus. FEMS Microbiol Lett *306*, 30-36.
- Dubrac, S., Boneca, I.G., Poupel, O., and Msadek, T. (2007). New insights into the WalK/WalR (YycG/YycF) essential signal transduction pathway reveal a major role in controlling cell wall metabolism and biofilm formation in Staphylococcus aureus. J Bacteriol *189*, 8257-8269.
- Dunman, P.M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., Brown, E.L., Zagursky, R.J., Shlaes, D., and Projan, S.J. (2001). Transcription profiling-based identification of Staphylococcus aureus genes regulated by the agr and/or sarA loci. J Bacteriol *183*, 7341-7353.
- Eyraud, A., Tattevin, P., Chabelskaya, S., and Felden, B. (2014). A small RNA controls a protein regulator involved in antibiotic resistance in Staphylococcus aureus. Nucleic Acids Res *42*, 4892-4905.
- Fridman, M., Williams, G.D., Muzamal, U., Hunter, H., Siu, K.W., and Golemi-Kotra, D. (2013). Two unique phosphorylation-driven signaling pathways crosstalk in Staphylococcus aureus to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. Biochemistry *52*, 7975-7986.
- Fujimoto, D.F., Higginbotham, R.H., Sterba, K.M., Maleki, S.J., Segall, A.M., Smeltzer, M.S., and Hurlburt, B.K. (2009). Staphylococcus aureus SarA is a regulatory protein responsive to redox and pH that can support bacteriophage lambda integrase-mediated excision/recombination. Mol Microbiol *74*, 1445-1458.
- Geiger, T., and Wolz, C. (2014). Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria. Int J Med Microbiol *304*, 150-155.
- Geisinger, E., Adhikari, R.P., Jin, R., Ross, H.F., and Novick, R.P. (2006). Inhibition of rot translation by RNAIII, a key feature of agr function. Mol Microbiol *61*, 1038-1048.
- Gertz, S., Engelmann, S., Schmid, R., Ohlsen, K., Hacker, J., and Hecker, M. (1999). Regulation of sigmaB-dependent transcription of sigB and asp23 in two different Staphylococcus aureus strains. Mol Gen Genet *261*, 558-566.
- Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., Deboy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M.*, et al.* (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. Journal of bacteriology *187*, 2426-2438.

- Gupta, R.K., Luong, T.T., and Lee, C.Y. (2015). RNAIII of the Staphylococcus aureus agr system activates global regulator MgrA by stabilizing mRNA. Proc Natl Acad Sci U S A *112*, 14036-14041.
- Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. Mol Cell *3*, 495-504.
- Gustafsson, E., and Oscarsson, J. (2008). Maximal transcription of aur (aureolysin) and sspA (serine protease) in Staphylococcus aureus requires staphylococcal accessory regulator R (sarR) activity. FEMS Microbiol Lett *284*, 158-164.
- Herbert, S., Ziebandt, A.K., Ohlsen, K., Schafer, T., Hecker, M., Albrecht, D., Novick, R., and Gotz, F. (2010). Repair of global regulators in Staphylococcus aureus 8325 and comparative analysis with other clinical isolates. Infect Immun *78*, 2877-2889.
- Homerova, D., Bischoff, M., Dumolin, A., and Kormanec, J. (2004). Optimization of a two-plasmid system for the identification of promoters recognized by RNA polymerase containing Staphylococcus aureus alternative sigma factor sigmaB. FEMS Microbiol Lett *232*, 173-179.
- Howden, B.P., Beaume, M., Harrison, P.F., Hernandez, D., Schrenzel, J., Seemann, T., Francois, P., and Stinear, T.P. (2013). Analysis of the small RNA transcriptional response in multidrug-resistant Staphylococcus aureus after antimicrobial exposure. Antimicrob Agents Chemother *57*, 3864-3874.
- Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C.*, et al.* (2005). Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression. The EMBO journal *24*, 824-835.
- Hurwitz, J. (2005). The discovery of RNA polymerase. J Biol Chem *280*, 42477-42485.
- Ibarra, J.A., Perez-Rueda, E., Carroll, R.K., and Shaw, L.N. (2013). Global analysis of transcriptional regulators in Staphylococcus aureus. BMC Genomics *14*, 126.
- Imashimizu, M., Shimamoto, N., Oshima, T., and Kashlev, M. (2014). Transcription elongation. Heterogeneous tracking of RNA polymerase and its biological implications. Transcription *5*, e28285.
- Ingavale, S., van Wamel, W., Luong, T.T., Lee, C.Y., and Cheung, A.L. (2005). Rat/MgrA, a regulator of autolysis, is a regulator of virulence genes in Staphylococcus aureus. Infect Immun *73*, 1423-1431.
- Jensen, S.O., and Lyon, B.R. (2009). Genetics of antimicrobial resistance in Staphylococcus aureus. Future Microbiol *4*, 565-582.

- Juang, Y.L., and Helmann, J.D. (1994). The delta subunit of Bacillus subtilis RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. J Mol Biol *239*, 1-14.
- Kabata, H., Kurosawa, O., Arai, I., Washizu, M., Margarson, S.A., Glass, R.E., and Shimamoto, N. (1993). Visualization of single molecules of RNA polymerase sliding along DNA. Science *262*, 1561-1563.
- Karlsson, A., and Arvidson, S. (2002). Variation in extracellular protease production among clinical isolates of Staphylococcus aureus due to different levels of expression of the protease repressor sarA. Infect Immun *70*, 4239-4246.
- Keller, A.N., Yang, X., Wiedermannova, J., Delumeau, O., Krasny, L., and Lewis, P.J. (2014). epsilon, a new subunit of RNA polymerase found in gram-positive bacteria. J Bacteriol *196*, 3622-3632.
- Kim, S., Reyes, D., Beaume, M., Francois, P., and Cheung, A. (2014). Contribution of teg49 small RNA in the 5' upstream transcriptional region of sarA to virulence in Staphylococcus aureus. Infect Immun *82*, 4369-4379.
- Kluytmans, J., van Belkum, A., and Verbrugh, H. (1997). Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. Clinical microbiology reviews *10*, 505-520.
- Komatsuzawa, H., Sugai, M., Shirai, C., Suzuki, J., Hiramatsu, K., and Suginaka, H. (1995). Triton X-100 alters the resistance level of methicillin-resistant Staphylococcus aureus to oxacillin. FEMS Microbiol Lett *134*, 209-212.
- Kropinski, A.M. (2006). Phage Therapy - Everything Old is New Again. Can J Infect Dis Med Microbiol *17*, 297-306.
- Kullik, I., Giachino, P., and Fuchs, T. (1998). Deletion of the alternative sigma factor sigmaB in Staphylococcus aureus reveals its function as a global regulator of virulence genes. Journal of bacteriology *180*, 4814-4820.
- Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y.*, et al.* (2001). Whole genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet *357*, 1225-1240.
- Lamichhane-Khadka, R., Cantore, S.A., Riordan, J.T., Delgado, A., Norman, A.E., Duenas, S., Zaman, S., Horan, S., Wilkinson, B.J., and Gustafson, J.E. (2009). sarA inactivation reduces vancomycin-intermediate and ciprofloxacin resistance expression by Staphylococcus aureus. Int J Antimicrob Agents *34*, 136-141.
- Le Pabic, H., Germain-Amiot, N., Bordeau, V., and Felden, B. (2015). A bacterial regulatory RNA attenuates virulence, spread and human host cell phagocytosis. Nucleic Acids Res *43*, 9232-9248.

- Li, L., Cheung, A., Bayer, A.S., Chen, L., Abdelhady, W., Kreiswirth, B.N., Yeaman, M.R., and Xiong, Y.Q. (2016). The Global Regulon sarA Regulates beta-Lactam Antibiotic Resistance in Methicillin-Resistant Staphylococcus aureus In Vitro and in Endovascular Infections. J Infect Dis *214*, 1421-1429.
- Li, X.Y., and McClure, W.R. (1998). Characterization of the closed complex intermediate formed during transcription initiation by Escherichia coli RNA polymerase. J Biol Chem *273*, 23549-23557.
- Liang, X., Zheng, L., Landwehr, C., Lunsford, D., Holmes, D., and Ji, Y. (2005). Global regulation of gene expression by ArlRS, a two-component signal transduction regulatory system of Staphylococcus aureus. J Bacteriol *187*, 5486-5492.
- Lindsay, J.A. (2014). Staphylococcus aureus genomics and the impact of horizontal gene transfer. Int J Med Microbiol *304*, 103-109.
- Lindsay, J.A., and Holden, M.T. (2004). Staphylococcus aureus: superbug, super genome? Trends Microbiol *12*, 378-385.
- Lisser, S., and Margalit, H. (1993). Compilation of E. coli mRNA promoter sequences. Nucleic Acids Res *21*, 1507-1516.
- Liu, J., and Turnbough, C.L., Jr. (1994). Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the pyrC promoter in Escherichia coli. J Bacteriol *176*, 2938-2945.
- Liu, P.F., Cheng, J.S., Sy, C.L., Huang, W.C., Yang, H.C., Gallo, R.L., Huang, C.M., and Shu, C.W. (2015). IsaB Inhibits Autophagic Flux to Promote Host Transmission of Methicillin-Resistant Staphylococcus aureus. J Invest Dermatol *135*, 2714-2722.
- Liu, Y., Manna, A.C., Pan, C.H., Kriksunov, I.A., Thiel, D.J., Cheung, A.L., and Zhang, G. (2006). Structural and function analyses of the global regulatory protein SarA from Staphylococcus aureus. Proc Natl Acad Sci U S A *103*, 2392-2397.
- Liu, Y., Mu, C., Ying, X., Li, W., Wu, N., Dong, J., Gao, Y., Shao, N., Fan, M., and Yang, G. (2011). RNAIII activates map expression by forming an RNA-RNA complex in Staphylococcus aureus. FEBS Lett *585*, 899-905.
- Mader, U., Nicolas, P., Depke, M., Pane-Farre, J., Debarbouille, M., van der Kooi-Pol, M.M., Guerin, C., Derozier, S., Hiron, A., Jarmer, H.*, et al.* (2016). Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. PLoS genetics *12*, e1005962.
- Malachowa, N., and DeLeo, F.R. (2010). Mobile genetic elements of Staphylococcus aureus. Cell Mol Life Sci *67*, 3057-3071.

- Mandal, S., Mahapa, A., Biswas, A., Jana, B., Polley, S., Sau, K., and Sau, S. (2016). A Surfactant-Induced Functional Modulation of a Global Virulence Regulator from Staphylococcus aureus. PLoS One *11*, e0151426.
- Mandin, P., and Guillier, M. (2013). Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. Curr Opin Microbiol *16*, 125-132.
- Manna, A., and Cheung, A.L. (2001a). Characterization of sarR, a modulator of sar expression in Staphylococcus aureus. Infect Immun *69*, 885-896.
- Manna, A., and Cheung, A.L. (2001b). Characterization of sarR, a modulator of sar expression in Staphylococcus aureus. Infection and immunity *69*, 885-896.
- Manna, A.C., and Cheung, A.L. (2003). sarU, a sarA homolog, is repressed by SarT and regulates virulence genes in Staphylococcus aureus. Infect Immun *71*, 343-353.
- Manna, A.C., and Cheung, A.L. (2006a). Expression of SarX, a negative regulator of agr and exoprotein synthesis, is activated by MgrA in Staphylococcus aureus. J Bacteriol *188*, 4288-4299.
- Manna, A.C., and Cheung, A.L. (2006b). Transcriptional regulation of the agr locus and the identification of DNA binding residues of the global regulatory protein SarR in Staphylococcus aureus. Mol Microbiol *60*, 1289-1301.
- Manna, A.C., Ingavale, S.S., Maloney, M., van Wamel, W., and Cheung, A.L. (2004). Identification of sarV (SA2062), a new transcriptional regulator, is repressed by SarA and MgrA (SA0641) and involved in the regulation of autolysis in Staphylococcus aureus. J Bacteriol *186*, 5267-5280.
- Mauro, T., Rouillon, A., and Felden, B. (2016). Insights into the regulation of small RNA expression: SarA represses the expression of two sRNAs in Staphylococcus aureus. Nucleic Acids Res *44*, 10186-10200.
- Menzies, B.E., and Kourteva, I. (2000). Staphylococcus aureus alpha-toxin induces apoptosis in endothelial cells. FEMS Immunol Med Microbiol *29*, 39-45.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.
- Moon, B.Y., Park, J.Y., Robinson, D.A., Thomas, J.C., Park, Y.H., Thornton, J.A., and Seo, K.S. (2016). Mobilization of Genomic Islands of Staphylococcus aureus by Temperate Bacteriophage. PLoS One *11*, e0151409.

- Morfeldt, E., Taylor, D., von Gabain, A., and Arvidson, S. (1995). Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII. EMBO J *14*, 4569-4577.
- Morikawa, K., Inose, Y., Okamura, H., Maruyama, A., Hayashi, H., Takeyasu, K., and Ohta, T. (2003). A new staphylococcal sigma factor in the conserved gene cassette: functional significance and implication for the evolutionary processes. Genes Cells *8*, 699-712.
- Morrison, J.M., Anderson, K.L., Beenken, K.E., Smeltzer, M.S., and Dunman, P.M. (2012). The staphylococcal accessory regulator, SarA, is an RNA-binding protein that modulates the mRNA turnover properties of late-exponential and stationary phase Staphylococcus aureus cells. Front Cell Infect Microbiol *2*, 26.
- Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol *13*, 31-39.
- Murakami, K.S., Masuda, S., and Darst, S.A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science *296*, 1280-1284.
- Negishi, T., Fujita, N., and Ishihama, A. (1995). Structural map of the alpha subunit of Escherichia coli RNA polymerase: structural domains identified by proteolytic cleavage. J Mol Biol *248*, 723-728.
- Newsom, S.W. (2008). Ogston's coccus. J Hosp Infect *70*, 369-372.
- Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. EMBO J *12*, 3967-3975.
- Oliva, G., Sahr, T., and Buchrieser, C. (2015). Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. FEMS Microbiol Rev *39*, 331-349.
- Oscarsson, J., Harlos, C., and Arvidson, S. (2005). Regulatory role of proteins binding to the spa (protein A) and sarS (staphylococcal accessory regulator) promoter regions in Staphylococcus aureus NTCC 8325-4. Int J Med Microbiol *295*, 253-266.
- Oscarsson, J., Tegmark-Wisell, K., and Arvidson, S. (2006). Coordinated and differential control of aureolysin (aur) and serine protease (sspA) transcription in Staphylococcus aureus by sarA, rot and agr (RNAIII). Int J Med Microbiol *296*, 365-380.
- Park, C.S., Wu, F.Y., and Wu, C.W. (1982). Molecular mechanism of promoter selection in gene transcription. II. Kinetic evidence for promoter search by a one-dimensional diffusion of RNA polymerase molecule along the DNA template. The Journal of biological chemistry *257*, 6950-6956.

- Pichon, C., and Felden, B. (2005). Small RNA genes expressed from Staphylococcus aureus genomic and pathogenicity islands with specific expression among pathogenic strains. Proc Natl Acad Sci U S A *102*, 14249-14254.
- Pinel-Marie, M.L., Brielle, R., and Felden, B. (2014). Dual toxic-peptide-coding Staphylococcus aureus RNA under antisense regulation targets host cells and bacterial rivals unequally. Cell Rep *7*, 424-435.
- Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet *32*, 508-522.
- Prados, J., Linder, P., and Redder, P. (2016). TSS-EMOTE, a refined protocol for a more complete and less biased global mapping of transcription start sites in bacterial pathogens. BMC Genomics *17*, 849.
- Proshkin, S., Mironov, A., and Nudler, E. (2014). Riboswitches in regulation of Rho-dependent transcription termination. Biochimica et biophysica acta *1839*, 974-977.
- Queck, S.Y., Jameson-Lee, M., Villaruz, A.E., Bach, T.H., Khan, B.A., Sturdevant, D.E., Ricklefs, S.M., Li, M., and Otto, M. (2008). RNAIII-independent target gene control by the agr quorum-sensing system: insight into the evolution of virulence regulation in Staphylococcus aureus. Mol Cell *32*, 150-158.
- Rechtin, T.M., Gillaspy, A.F., Schumacher, M.A., Brennan, R.G., Smeltzer, M.S., and Hurlburt, B.K. (1999). Characterization of the SarA virulence gene regulator of Staphylococcus aureus. Mol Microbiol *33*, 307-316.
- Reddie, K.G., and Carroll, K.S. (2008). Expanding the functional diversity of proteins through cysteine oxidation. Curr Opin Chem Biol *12*, 746-754.
- Reyes, D., Andrey, D.O., Monod, A., Kelley, W.L., Zhang, G., and Cheung, A.L. (2011). Coordinated regulation by AgrA, SarA, and SarR to control agr expression in Staphylococcus aureus. J Bacteriol *193*, 6020-6031.
- Rhodius, V.A., and Mutalik, V.K. (2010). Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE. Proc Natl Acad Sci U S A *107*, 2854-2859.
- Roberts, C., Anderson, K.L., Murphy, E., Projan, S.J., Mounts, W., Hurlburt, B., Smeltzer, M., Overbeek, R., Disz, T., and Dunman, P.M. (2006). Characterizing the effect of the Staphylococcus aureus virulence factor regulator, SarA, on log-phase mRNA half-lives. J Bacteriol *188*, 2593-2603.
- Ross, W., Vrentas, C.E., Sanchez-Vazquez, P., Gaal, T., and Gourse, R.L. (2013). The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. Mol Cell *50*, 420-429.

- Ruff, E.F., Record, M.T., Jr., and Artsimovitch, I. (2015). Initial events in bacterial transcription initiation. Biomolecules *5*, 1035-1062.
- Sakata-Sogawa, K., and Shimamoto, N. (2004). RNA polymerase can track a DNA groove during promoter search. Proceedings of the National Academy of Sciences of the United States of America *101*, 14731-14735.
- Santangelo, T.J., and Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. Nat Rev Microbiol *9*, 319-329.
- Sassi, M., Augagneur, Y., Mauro, T., Ivain, L., Chabelskaya, S., Hallier, M., Sallou, O., and Felden, B. (2015). SRD: a Staphylococcus regulatory RNA database. RNA *21*, 1005-1017.
- Sassi, M., Felden, B., and Augagneur, Y. (2014). Draft Genome Sequence of Staphylococcus aureus subsp. aureus Strain HG003, an NCTC8325 Derivative. Genome Announc *2*.
- Sayed, N., Jousselin, A., and Felden, B. (2012). A cis-antisense RNA acts in trans in Staphylococcus aureus to control translation of a human cytolytic peptide. Nat Struct Mol Biol *19*, 105-112.
- Schmidt, K.A., Manna, A.C., and Cheung, A.L. (2003). SarT influences sarS expression in Staphylococcus aureus. Infect Immun *71*, 5139-5148.
- Schmidt, K.A., Manna, A.C., Gill, S., and Cheung, A.L. (2001). SarT, a repressor of alpha-hemolysin in Staphylococcus aureus. Infect Immun *69*, 4749-4758.
- Schumacher, M.A., Hurlburt, B.K., and Brennan, R.G. (2001). Crystal structures of SarA, a pleiotropic regulator of virulence genes in S. aureus. Nature *409*, 215-219.
- Senn, M.M., Giachino, P., Homerova, D., Steinhuber, A., Strassner, J., Kormanec, J., Fluckiger, U., Berger-Bachi, B., and Bischoff, M. (2005). Molecular analysis and organization of the sigmaB operon in Staphylococcus aureus. Journal of bacteriology *187*, 8006-8019.
- Shaw, L., Golonka, E., Potempa, J., and Foster, S.J. (2004). The role and regulation of the extracellular proteases of Staphylococcus aureus. Microbiology *150*, 217-228.
- Shaw, L.N., Lindholm, C., Prajsnar, T.K., Miller, H.K., Brown, M.C., Golonka, E., Stewart, G.C., Tarkowski, A., and Potempa, J. (2008). Identification and characterization of sigma, a novel component of the Staphylococcus aureus stress and virulence responses. PLoS One *3*, e3844.
- Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference--a crash course. Trends Genet *21*, 339-345.

- Solecki, O., Mosbah, A., Baudy Floc'h, M., and Felden, B. (2015). Converting a Staphylococcus aureus toxin into effective cyclic pseudopeptide antibiotics. Chem Biol *22*, 329-335.
- Stenz, L., Francois, P., Whiteson, K., Wolz, C., Linder, P., and Schrenzel, J. (2011). The CodY pleiotropic repressor controls virulence in gram-positive pathogens. FEMS Immunol Med Microbiol *62*, 123-139.
- Sterba, K.M., Mackintosh, S.G., Blevins, J.S., Hurlburt, B.K., and Smeltzer, M.S. (2003). Characterization of Staphylococcus aureus SarA binding sites. J Bacteriol *185*, 4410-4417.
- Sun, F., Ding, Y., Ji, Q., Liang, Z., Deng, X., Wong, C.C., Yi, C., Zhang, L., Xie, S., Alvarez, S., *et al.* (2012). Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. Proc Natl Acad Sci U S A *109*, 15461-15466.
- Tamber, S., and Cheung, A.L. (2009). SarZ promotes the expression of virulence factors and represses biofilm formation by modulating SarA and agr in Staphylococcus aureus. Infect Immun *77*, 419-428.
- Tao, L., Wu, X., and Sun, B. (2010). Alternative sigma factor sigmaH modulates prophage integration and excision in Staphylococcus aureus. PLoS pathogens *6*, e1000888.
- Tegmark, K., Karlsson, A., and Arvidson, S. (2000). Identification and characterization of SarH1, a new global regulator of virulence gene expression in Staphylococcus aureus. Mol Microbiol *37*, 398-409.
- Thompson, J.K., and Hart, M.G. (1981). Novel patterns of ultraviolet mutagenesis and Weigle reactivation in Staphylococcus aureus and phage phi 11. J Gen Microbiol *124*, 147-157.
- Tong, S.Y., Davis, J.S., Eichenberger, E., Holland, T.L., and Fowler, V.G., Jr. (2015). Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev *28*, 603-661.
- Trotonda, M.P., Manna, A.C., Cheung, A.L., Lasa, I., and Penades, J.R. (2005). SarA positively controls bap-dependent biofilm formation in Staphylococcus aureus. J Bacteriol *187*, 5790-5798.
- Trotonda, M.P., Xiong, Y.Q., Memmi, G., Bayer, A.S., and Cheung, A.L. (2009). Role of mgrA and sarA in methicillin-resistant Staphylococcus aureus autolysis and resistance to cell wall-active antibiotics. J Infect Dis *199*, 209-218.
- Valle, J., Toledo-Arana, A., Berasain, C., Ghigo, J.M., Amorena, B., Penades, J.R., and Lasa, I. (2003). SarA and not sigmaB is essential for biofilm development by Staphylococcus aureus. Mol Microbiol *48*, 1075-1087.

- Villet, R.A., Truong-Bolduc, Q.C., Wang, Y., Estabrooks, Z., Medeiros, H., and Hooper, D.C. (2014). Regulation of expression of abcA and its response to environmental conditions. J Bacteriol *196*, 1532-1539.
- Voyich, J.M., Braughton, K.R., Sturdevant, D.E., Whitney, A.R., Said-Salim, B., Porcella, S.F., Long, R.D., Dorward, D.W., Gardner, D.J., Kreiswirth, B.N.*, et al.* (2005). Insights into mechanisms used by Staphylococcus aureus to avoid destruction by human neutrophils. J Immunol *175*, 3907-3919.
- Watson, S.P., Antonio, M., and Foster, S.J. (1998). Isolation and characterization of Staphylococcus aureus starvation-induced, stationary-phase mutants defective in survival or recovery. Microbiology *144 (Pt 11)*, 3159-3169.
- Weiss, A., Ibarra, J.A., Paoletti, J., Carroll, R.K., and Shaw, L.N. (2014). The delta subunit of RNA polymerase guides promoter selectivity and virulence in Staphylococcus aureus. Infect Immun *82*, 1424-1435.
- Wen, J., and Fozo, E.M. (2014). sRNA antitoxins: more than one way to repress a toxin. Toxins (Basel) *6*, 2310-2335.
- Wertheim, H.F., Melles, D.C., Vos, M.C., van Leeuwen, W., van Belkum, A., Verbrugh, H.A., and Nouwen, J.L. (2005). The role of nasal carriage in Staphylococcus aureus infections. Lancet Infect Dis *5*, 751-762.
- Wolz, C., Pohlmann-Dietze, P., Steinhuber, A., Chien, Y.T., Manna, A., van Wamel, W., and Cheung, A. (2000). Agr-independent regulation of fibronectin-binding protein(s) by the regulatory locus sar in Staphylococcus aureus. Mol Microbiol *36*, 230-243.
- Wu, S., de Lencastre, H., and Tomasz, A. (1996). Sigma-B, a putative operon encoding alternate sigma factor of Staphylococcus aureus RNA polymerase: molecular cloning and DNA sequencing. Journal of bacteriology *178*, 6036-6042.
- Xue, T., Zhang, X., Sun, H., and Sun, B. (2014). ArtR, a novel sRNA of Staphylococcus aureus, regulates alpha-toxin expression by targeting the 5' UTR of sarT mRNA. Med Microbiol Immunol *203*, 1-12.
- Yan, J., Liu, Y., Gao, Y., Dong, J., Mu, C., Li, D., and Yang, G. (2015). RNAIII suppresses the expression of LtaS via acting as an antisense RNA in Staphylococcus aureus. J Basic Microbiol *55*, 255-261.
- Zhang, Z., Wang, J., Zhang, B., Liu, H., Song, W., He, J., Lv, D., Wang, S., and Xu, X. (2013). Activity of antibacterial protein from maggots against Staphylococcus aureus in vitro and in vivo. International journal of molecular medicine *31*, 1159-1165.
- Zhao, Y., Verma, V., Belcheva, A., Singh, A., Fridman, M., and Golemi-Kotra, D. (2012). Staphylococcus aureus methicillin-resistance factor fmtA is regulated by the global regulator SarA. PLoS One *7*, e43998.

- Zheng, W., Liang, Y., Zhao, H., Zhang, J., and Li, Z. (2015). 5,5'-Methylenedisalicylic Acid (MDSA) Modulates SarA/MgrA Phosphorylation by Targeting Ser/Thr Phosphatase Stp1. Chembiochem *16*, 1035-1040.

Liens internets :

[1]:http://www.inserm.fr/thematiques/immunologie-inflammation-infectiologie-et-microbiologie/dossiers-d-information/infections-nosocomiales

[2]:http://www.euro.who.int/__data/assets/pdf_file/0006/113829/Patient_safety_pres_jakab.pdf? ua=1