



HAL
open science

Déterminants structuraux de la régulation de la compétence par le système de communication ComRS chez les streptocoques

Jordhan Thuillier

► **To cite this version:**

Jordhan Thuillier. Déterminants structuraux de la régulation de la compétence par le système de communication ComRS chez les streptocoques. Biologie structurale [q-bio.BM]. Université Paris-Saclay, 2019. Français. NNT : 2019SACLS504 . tel-02462742

HAL Id: tel-02462742

<https://theses.hal.science/tel-02462742>

Submitted on 31 Jan 2020

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

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

Déterminants structuraux de la régulation de la compétence par le système de communication ComRS chez les streptocoques

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°569
Innovation Thérapeutique du Fondamental à l'Appliqué (ITFA)
Spécialité du doctorat : Biochimie et Biologie structurale

Thèse présentée et soutenue à ORSAY, le 10 décembre 2019, par

Jordhan THUILLIER

Composition du Jury :

Dr. Yves DESSAUX DR CNRS, I2BC – Gif sur Yvette (UMR 9198)	Président
Dr. Andrea DESSEN DR CNRS, IBS - Grenoble (UMR 5075)	Rapporteur
Dr. Rémi FRONZES DR CNRS, IECB - Bordeaux (UMR 5234)	Rapporteur
Dr. Rozenn GARDAN CR INRA, Institut Micalis – Jouy en Josas (UMR 1319)	Examineur
Dr. Johann MIGNOLET Assistant de recherche, UCLouvain (Belgique)	Examineur
Dr. Sylvie NESSLER PR UPSUD, I2BC - Orsay (UMR 9198)	Directeur de thèse

Remerciements

Je tiens à remercier en premier lieu Andrea Dessen, Rémi Fronzes, Rozenn Gardan, Johann Mignolet et Yves Dessaux d'avoir accepté de juger ce travail de thèse.

Je remercie également Herman van Tilbreurgh de m'avoir accueilli au sein de son équipe.

Mes remerciements les plus sincères vont tout naturellement à ma directrice de thèse, Sylvie Nessler, qui m'a donné la possibilité de faire ma thèse avec elle. En plus de passer beaucoup trop vite, ces trois ans à tes côtés m'ont permis de progresser tant sur le plan scientifique que le plan humain.

Je souhaite également adresser un très grand Merci à Dominique Durand pour son expertise concernant le SAXS, sa disponibilité, son aide et sa gentillesse, mais aussi à Magali Aumont-Nicaise sur la plateforme PIM pour son aide en DSC et en ITC ainsi que pour sa bonne humeur permanente. Je tiens également à remercier Christophe Velours pour son expertise du SEC-MALS et sa sympathie.

Merci également à Noureddine Lazar pour m'avoir formé sur les machines de purification ainsi que pour son aide technique et à Ines Gallay de m'avoir initié à la cristallographie et m'avoir guidé lors des premiers run au synchrotron.

Je tiens spécifiquement à remercier Hélène pour les nombreuses discussions, scientifiques ou non, que l'on a pu avoir. Tu es toujours de très bon conseil et c'est un plaisir que d'échanger avec toi. De la même manière, j'aimerais remercier spécifiquement Raoudha qui a réussi à me supporter, moi et mon envie d'ouvrir la fenêtre en hiver, pendant 3ans. J'aurais difficilement pu trouver une meilleure partenaire de bureau que toi.

Une pensée particulière pour Louisa et Sophia qui savent mieux que les autres ce que l'on peut traverser durant une thèse. Il est important de bien s'entendre et de se soutenir entre doctorants et je suis très heureux d'avoir fait votre connaissance.

Je remercie également Stéphane qui en plus de m'avoir aidé sur la plateforme de cristallographie, est une personne avec qui j'ai réellement apprécié passer du temps.

Je me sens chanceux d'être tombé dans cette équipe et pour finir je tenais également à remercier toute l'équipe FAAM.

Enfin, un dernier merci plus personnel pour ma famille ainsi que pour Caroline.

Table des matières

Remerciements	3
Sommaire des figures	9
Sommaire des tableaux.....	13
I. Introduction.....	15
1. La communication intercellulaire chez les bactéries.....	15
1.1. Généralités	15
1.2. Le quorum sensing : Un cas particulier de communication intercellulaire.....	16
1.3. Intérêt thérapeutique du quorum sensing	17
2. Le quorum sensing chez les bactéries à Gram négatif	19
2.1. Autoinducteurs de type Acyl Homoserine Lactone	19
2.2. Les différents systèmes QS à AHL.....	21
2.2.1. Les systèmes de type LuxI/LuxR	21
2.2.2. Autres systèmes QS à AHL	23
3. Le quorum sensing chez les bactéries à Gram positif	24
3.1. Les autoinducteurs peptidiques (AIPs)	24
3.2. Les systèmes QS indirects	26
3.3. Les systèmes QS directs.....	28
4. Les récepteurs du QS direct de la famille RNPP	30
4.1. Structure	30
4.1.1. Le domaine TPR.....	31
4.1.2. Le domaine HTH.....	32
4.2. Régulation des RNPPs par les AIPs.....	33
4.2.1. Le système Rap/Phr	33
4.2.2. Le système PlcR/PapR.....	36
4.2.3. Le système NprR/NprX	39
4.2.4. Le système PrgX.....	43

4.2.5.	Le système ComR/ComS.....	44
5.	But de la thèse	52
II.	Résultats	55
1.	Etude de la spécificité du système ComR/XIP.....	55
1.1.	Comparaison des séquences de ComR _{Sth} et ComR _{Sve}	55
1.2.	Comparaison structurale de ComR _{Sth} et ComR _{Sve}	57
1.3	Analyse fonctionnelle par mutagenèse dirigée	60
1.3.	Manuscrit de l'article	62
1.5.	Description du mécanisme d'activation chez les ComR de type I.....	63
1.5.1.	Fixation du peptide	63
1.5.2.	Stabilisation de la conformation fermée de la poche de fixation	65
1.5.3.	Relargage des HTH	68
1.5.4.	Dimérisation	70
1.6.	Vers la caractérisation de peptides inhibiteurs.....	71
1.7.	Conformations alternatives de ComR _{Sth} apo.....	73
1.8.	Perspectives.....	75
2.	Etude structurale de deux paralogues de ComR chez <i>S. salivarius</i>	76
2.1.	Introduction.....	76
2.2.	Comparaison structurale des protéines ScuR et SarF	80
2.2.1.	Purification des deux paralogues	80
2.2.2.	Analyse DSC	82
2.2.3.	Analyse par SEC-MALS	83
2.2.4.	Analyse par SAXS.....	85
2.2.4.1.	Exploitation des données de diffusion	86
2.2.4.2.	Modélisation par homologie de ScuR.....	91
2.2.5.	Perspectives	94
III.	Discussion	97

IV.	Matériel et méthodes	105
1.	Production des protéines	105
1.1.	Plasmides utilisés	105
1.2.	Surexpression des protéines	105
1.3.	Marquage des protéines au Sélénium.....	106
2.	Purification par chromatographie d'affinité.....	107
3.	Préparation des ligands.....	110
4.	Mesures interactions peptide-protéine	111
4.1.	Calorimétrie différentielle à balayage (DSC)	111
4.2.	Calorimétrie par titration isotherme (ITC).....	112
4.3.	Size Exclusion Chromatography Multi-Angle Light Scattering (SEC-MALS) ..	113
5.	Analyse structurale par diffusion des rayons X aux petits angles (SAXS).....	114
5.1.	Principe	114
5.2.	Protocole expérimental	117
5.3.	Modélisation	118
6.	Cristallographie	119
6.1.	Cristallogenèse	119
6.1.1.	Principe de la cristallisation.....	119
6.1.2.	Criblage des conditions de cristallisation	121
6.1.3.	Optimisation des cristaux	124
6.2.	Analyse par diffraction aux rayons X	126
6.2.1.	Théorie de la diffraction	126
6.2.2.	Démarche expérimentale	131
6.3.	Affinement du modèle atomique	133
6.3.1.	Dépôt de la structure dans la PDB.....	134
V.	Références bibliographique.....	135

Sommaire des figures

Figure 1 : Schématisation de la communication bactérienne.....	15
Figure 2 : Illustration du mécanisme de quorum sensing.....	16
Figure 3 : Evolution du nombre d'articles traitant du quorum sensing.....	17
Figure 4 : Les différents modes d'action des antibiotiques.....	18
Figure 5: Structure générale d'une Acyl Homoserine Lactone.....	19
Figure 6 : Quelques exemples d'AHLs.....	20
Figure 7 : Schéma du système QS LuxI/LuxR.....	21
Figure 8 : Structure des récepteurs de type LuxR.....	22
Figure 9 : Régulation de l'opéron LuxICDABEG chez <i>V. fischeri</i> par 3 systèmes QS.....	23
Figure 10 : Structure de l'autoinducteur AI-2.....	24
Figure 11 : Différentes formes d'AIP cyclique.....	26
Figure 12 : Schéma du fonctionnement d'un système à deux composants.....	27
Figure 13 : Mécanisme d'activation du système à deux composants ComCDE.....	28
Figure 14 : Arbre phylogénétique des récepteurs Rgg.....	29
Figure 15 : Schéma de la structure des différentes protéines (R)RNPP.....	30
Figure 16 : Structure des domaines à motifs TPR.....	31
Figure 17 : Structures de différents types de domaines HTH.....	32
Figure 18 : Structure de deux domaines HTH fixés sur un ADN cible.....	33
Figure 19 : Schéma du mode d'action des systèmes Rpa/Phr chez <i>Bacillus subtilis</i>	34
Figure 20 : Mécanisme moléculaire de la régulation des systèmes Rap/Phr.....	35
Figure 21 : Schéma du mode d'action du système PlcR-PapR chez <i>B. thuringiensis</i>	36
Figure 22 : Classification des systèmes PlcR/PapR de <i>B. cereus</i>	37
Figure 23 : Mécanisme de régulation de PlcR.....	38
Figure 24 : Schéma du mode d'action bifonctionnel de NprR chez <i>B. thuringiensis</i>	39
Figure 25 : Classification des systèmes NprR/NprX de <i>B. cereus</i>	40
Figure 26 : Structure de la forme apo de NprR(Δ HTH).....	41
Figure 27 : Superposition des dimères de NprR et de RapF.....	42
Figure 28 : Modèle d'interaction entre le complexe NprR/NprX et un ADN cible.....	42
Figure 29 : Modèle du complexe PrgX-ADN.....	44
Figure 30 : Schéma ComRS chez <i>S. mutans</i>	45

Figure 31 : Classification des voies de compétence chez les Streptocoques	46
Figure 32 : Structure de la forme apo de ComR.	47
Figure 33 : Structure du complexe ternaire ComR _{Sth} /XIP _{Sth} /ADN	47
Figure 34 : Mécanisme d'activation proposé pour le système ComRS de <i>S. thermophilus</i>	48
Figure 35 : Zoom sur la zone de dimérisation dans le complexe ComR _{Sth} /XIP/ADN.....	49
Figure 36 : Activité relative sur ComR _{Sth} de 8 variants XIP.....	49
Figure 37 : Comparaison des affinités des peptides XIP _{Sth} et XIP _{Sth} -Y3A.	50
Figure 38 : Analyse de la permissivité des systèmes ComRS de type II	51
Figure 39 : Poche de fixation du peptide XIP _{Sth} sur son récepteur ComR _{Sth}	52
Figure 40 : Alignement de séquence entre ComR _{Sth} et ComR _{Sve}	55
Figure 41: Cristaux obtenus lors des essais de cristallisation sur la forme apo de ComR _{Sve} ...	57
Figure 42 : Cristal ayant permis de résoudre la structure du complexe ComR _{Sve} / XIP _{Sve}	58
Figure 43 : Carte de densité électronique du complexe binaire ComR _{Sve} / XIP _{Sve}	59
Figure 44 : Absence de cross-talk entre ComR _{Sth} et ComR _{Sve}	60
Figure 45 : Isotherme de l'interaction ComR _{Sve} / XIP _{Sve}	61
Figure 46 : Structure de l'isothiocyanate de fluorescéine (FITC).....	62
Figure 47 : Superposition des peptides de XIP _{Sve} VPFMIYY et XIP _{Sth} LPYFAGCL.....	63
Figure 48 : Comparaison des sites de fixation du peptide XIP sur ComR _{Sth} et ComR _{Sve}	64
Figure 49 : Comparaison le l'ancrage des peptides XIP dans ComR _{Sth} et de ComR _{Sve}	65
Figure 50 : Superposition des sites de fixation entre les protéines ComR _{Sth} et ComR _{Sve}	66
Figure 51 : Stabilisation de la forme active fermée.	67
Figure 52 : Shift de l'hélice α 8.....	68
Figure 53 : Perte du pont salin R39 – E117	68
Figure 54 : Rôle de la partie N-terminale du peptide.....	69
Figure 55 : Cassure de l'hélice α 9.....	70
Figure 56 : Zoom sur la dimérisation de ComR _{Sth} en complexe avec XIP et l'ADN	70
Figure 57 : Différence de dimérisation entre ComR _{Sth} et ComR _{Sve}	71
Figure 58 : DSC sur les complexes ComR _{Sth} /XIP _{Sth} -Y3A et ComR _{Sth} /XIP _{Sth} -F4A.	72
Figure 59 : Cristaux de ComR _{Sth} apo.....	73
Figure 60 : Cristaux obtenus pour le complexe ComR/XIP chez <i>S. mutans</i> et <i>S. pyogenes</i>	75
Figure 61 : Couplage entre compétence et prédation chez <i>S. salivarius</i>	76
Figure 62 : Alignement de séquence entre ComR _{Ssa} et ses paralogues ScuR et SarF.....	77
Figure 63 : Séquence consensus des 22 peptides activateurs de ScuR et SarF.	79
Figure 64 : Mesure d'interaction par polarisation de fluorescence.	79

Figure 65 : Chromatographes en sortie de colonne Strep-Tactin.....	81
Figure 66 : Gels SDS-PAGE de contrôle de la pureté pour les protéines ScuR et SarF.....	81
Figure 67 : DSC sur ScuR et SarF en absence et en présence de BI7.....	82
Figure 68 : Analyse par SEC-MALS de la protéine SarF apo	83
Figure 69 : Analyse par SEC-MALS de la protéine SarF-BI7.....	84
Figure 70 : Analyse par SEC-MALS de la protéine ScuR apo.	84
Figure 71 : Analyse par SEC-MALS de la protéine ScuR-BI7.....	85
Figure 72 : Chromatogrammes des échantillons passés en SAXS en sortie de SEC.....	86
Figure 73 : Analyse par SAXS de ScuR et SarF avec et sans peptide BI7.	87
Figure 74 : Représentation logarithmique des courbes $I(q)/c$	89
Figure 75 : Fonctions de distribution des distances r des complexes ScuR-BI7 et SarF-BI7..	89
Figure 76 : Comparaison des complexes ScuR-BI7 et SarF-BI7	90
Figure 77 : Modélisation de ScuR apo	91
Figure 78 : Modélisation de ScuR-BI7	92
Figure 79 : Courbes expérimentales et calculées pour les modèles SarF apo et SarF-BI7	93
Figure 80 : Cristal du complexe SarF-BI7-pSptA.....	94
Figure 81 : Les différentes stratégies de QQ chez les bactéries à Gram négatif.....	98
Figure 82 : Superposition de la forme apo de ComR _{Sth} et ComR _{Ssu}	100
Figure 83 : Alignement de séquence entre protéines ComR de types I, II et III.....	101
Figure 84 : Zoom sur la séquestration des domaines HTH chez ComR _{Sth} et ComR _{Ssu}	102
Figure 85 : Superposition de XIP _{Sth} dans la poche de ComR _{Ssu}	103
Figure 86 : Profil d'élution de ComR _{Sve} sur colonne d'héparine.....	110
Figure 87 : Principe de la DSC.....	111
Figure 88 : Principe de l'analyse par ITC	113
Figure 89 : Principe du SAXS.....	115
Figure 90 : Schéma d'une expérience de cristallisation en phase vapeur	120
Figure 91 : Diagramme de phase type d'une expérience de cristallisation.....	121
Figure 92 : Plaque ttp LABTECH triple sitting drop.....	122
Figure 93 : Photo du robot pipeteur TECAN	122
Figure 94 : Robots de cristallisation.....	123
Figure 95 : Capture d'écran de l'interface du logiciel CRIMS	124
Figure 96 : Robot de pipetage Hamilton.....	125
Figure 97 : Photo d'une plaque 24puits permettant de faire des optimisations manuelles. ...	125
Figure 98 : Schéma d'une maille élémentaire.....	126

Figure 99 : Schéma expliquant le phénomène de diffraction au sein d'un réseau cristallin..	127
Figure 100 : Cliché de diffraction	128
Figure 101 : Reconstruction d'une structure protéique dans sa carte de densité associée	129

Sommaire des tableaux

Tableau 1 : Présentation de la base de données Quorumpeps.....	25
Tableau 2 : Conservation des résidus importants.....	56
Tableau 3 : Statistiques des données cristallographiques de 6QER.....	74
Tableau 4 : Tableau récapitulatif des promoteurs reconnus par ComR et ses paralogues.	78
Tableau 5 : Tableau récapitulatif du résultat des analyses des courbes SAXS.....	88
Tableau 6 : Indexation du jeu de donnée du complexe SarF-BI7-pSptA.....	94
Tableau 7 : Statistiques du jeu de données du complexe SarF-BI7-pSptA.....	95
Tableau 8: Détermination du nombre de complexes SarF-BI7-pSptA par unité asymétrique.	95
Tableau 9 : Composition du milieu synthétique pauvre.....	107
Tableau 10 : Coefficients d'extinctions molaire	109
Tableau 11 : Paramètres utilisés pour les mesures de SAXS.....	117

I. Introduction

1. La communication intercellulaire chez les bactéries

1.1. Généralités

Dans un monde unicellulaire où chaque individu se reproduit par scissiparité, la coopération entre les bactéries peut sembler improbable et, pendant longtemps, la pensée générale suggérait que les bactéries avaient une faible capacité d'interaction. Pourtant, on sait aujourd'hui que les bactéries communiquent entre elles en diffusant dans leur environnement proche de grandes quantités de molécules signal qui vont être reconnues par les bactéries avoisinantes et qui vont leur permettre de se coordonner. Ces petites molécules sont appelées phéromones ou auto-inducteurs, du fait qu'elles peuvent réguler positivement leur propre expression afin d'induire leur surproduction. Cela leur permet de partager des informations concernant leur environnement proche, comme la nature physicochimique du milieu dans lequel elles évoluent, et d'adopter une réponse coordonnée pour optimiser leur capacité à survivre et à se multiplier (WATERS & BASSLER 2005; WILLIAMS *et al.* 2007). Elles vont ainsi réguler l'expression de nombreux gènes impliqués dans des processus bactériens majeurs comme la formation de biofilm, la virulence, la sporulation ou la compétence (Figure 1). Les bactéries pourront ainsi acquérir un comportement coopératif permettant à une population bactérienne d'agir comme une entité multicellulaire (HOOSHANGI & BENTLEY 2008).

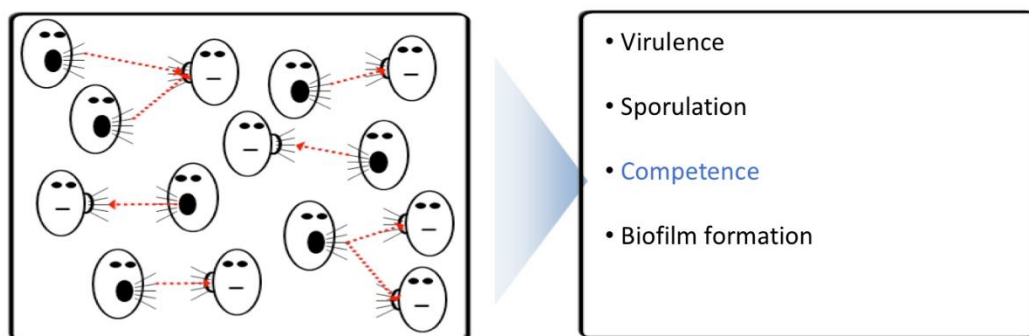


Figure 1 : Schématisation de la communication bactérienne.

En échangeant des molécules signal (flèches rouges), les bactéries émettrices et réceptrices régulent des processus bactériens majeurs, comme la compétence dont on discutera plus particulièrement dans cette thèse.

Les bactéries peuvent également communiquer avec des cellules eucaryotes. On peut citer l'exemple de la bactérie pathogène *Agrobacterium tumefaciens* qui colonise certains végétaux et qui est responsable de la galle du collet, une maladie capable de former des tumeurs au sein des plantes hôtes. *A. tumefaciens* sécrète des phéromones qui vont être perçues par les cellules eucaryotes du végétal malade et permettre à la bactérie de transférer un plasmide contenant des gènes favorables à sa multiplication (HAUDECOEUR & FAURE 2010). Chez l'homme, certaines bactéries peuvent synthétiser des substances neurochimiques, telles que les catécholamines (exemples : adrénaline ou dopamine), identiques à celles produites par les cellules eucaryotes hôtes en cas de stress (LYTE 2016).

1.2. Le quorum sensing : Un cas particulier de communication intercellulaire

Le quorum sensing (QS) est un mécanisme de communication intercellulaire particulier qui se déclenche uniquement lorsqu'une concentration seuil de molécules signal sécrétées est atteinte (SIFRI 2008). Plus la densité cellulaire va croître et plus les phéromones vont s'accumuler dans l'espace extracellulaire, jusqu'à atteindre une concentration seuil, appelée quorum, qui va permettre d'induire l'activation de récepteurs spécifiques régulant les gènes impliqués dans des processus biologiques comme la formation de biofilm, la bioluminescence ou la virulence (Figure 2).

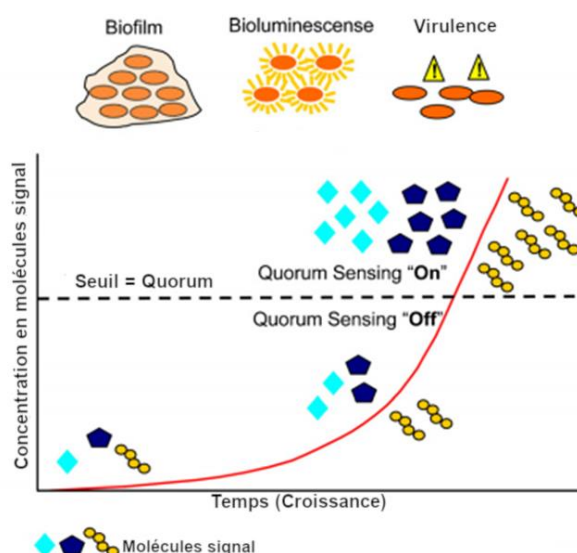


Figure 2 : Illustration du mécanisme de quorum sensing (ROY *et al.* 2011)

Au cours de la croissance bactérienne, la concentration en molécules signal (AHL, AIP ou AI-2) augmente jusqu'à atteindre un seuil qui permet de déclencher une réponse collective de la population sous forme de formation de biofilm, synthèse de facteurs de virulence, etc...

On notera que dans la littérature, de nombreuses publications utilisent abusivement du terme de quorum sensing pour parler de communication entre les bactéries sans que la notion de seuil soit vérifiée (WILLIAMS *et al.* 2007). Par soucis de cohérence, nous utiliserons également le terme générique de QS pour décrire les différents mécanismes de communication cités dans cette étude.

1.3. Intérêt thérapeutique du quorum sensing

Depuis la première description d'un mécanisme de QS impliqué dans la régulation du processus de conjugaison du plasmide Ti chez la bactérie *A. tumefaciens* en 1994 (FUQUA & WINANS 1994), le nombre d'articles académiques traitant du sujet a rapidement augmenté (Figure 3).

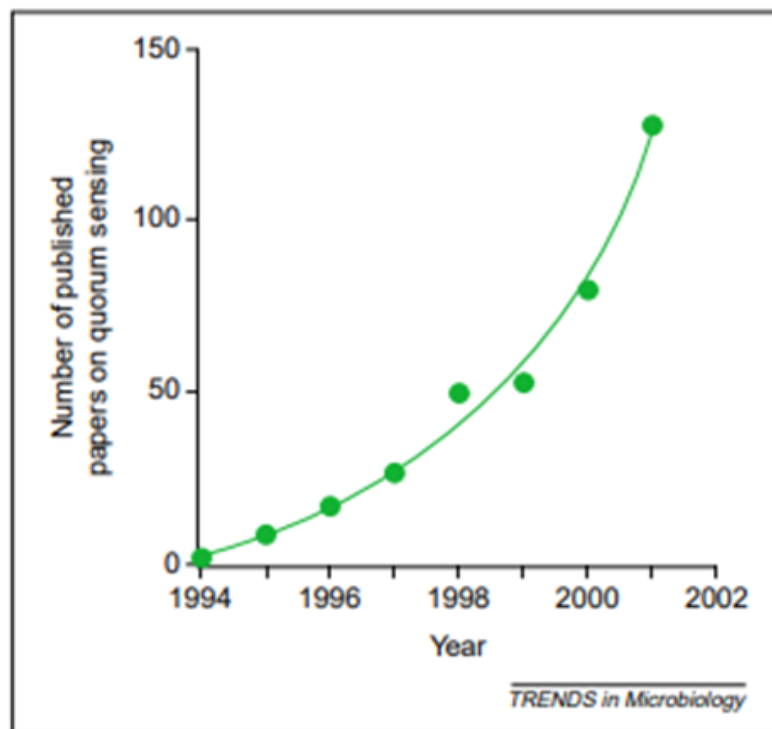


Figure 3: Evolution du nombre d'articles traitant du quorum sensing. (REDFIELD 2002)

Ces études ont ainsi démontré l'importance du QS dans la régulation du processus infectieux des 12 agents pathogènes considérés comme prioritaires pour le développement de nouveaux antibiotiques par l'Organisation Mondiale de la Santé (<https://www.who.int/fr/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>). Des systèmes QS ont en effet été décrits chez les pathogènes à Gram négatif

comme *Acinetobacter baumannii* (DOU *et al.* 2017), *Pseudomonas aeruginosa* (RUIMY & ANDREMONT 2004), *Enterobacter asburiae* (LAU *et al.* 2013), *Helicobacter pylori* (RADER *et al.* 2011), *Campylobacter* (PLUMMER 2012), *Salmonella typhimurium* (CHOI *et al.* 2012), *Neisseria gonorrhoeae* (FALSETTA *et al.* 2011), *Haemophilus influenzae* (SWORDS 2012) et *Shigella flexneri* (DAY & MAURELLI 2001) mais aussi chez les pathogènes à Gram positif comme *Enterococcus faecalis* (ALI *et al.* 2017) et *Staphylococcus aureus* (JI *et al.* 1995).

L'inhibition du QS, que l'on nomme aussi quorum quenching (QQ), comme potentielle approche antimicrobienne est donc actuellement beaucoup étudiée (MION *et al.* 2019). Comme le QQ n'affecte pas la croissance bactérienne, il devrait permettre de minimiser la pression de sélection et ainsi éviter l'apparition de résistance et notamment d'antibiorésistance, contrairement aux antibiotiques qui ciblent des processus cellulaires essentiels à la survie des bactéries (Figure 4).

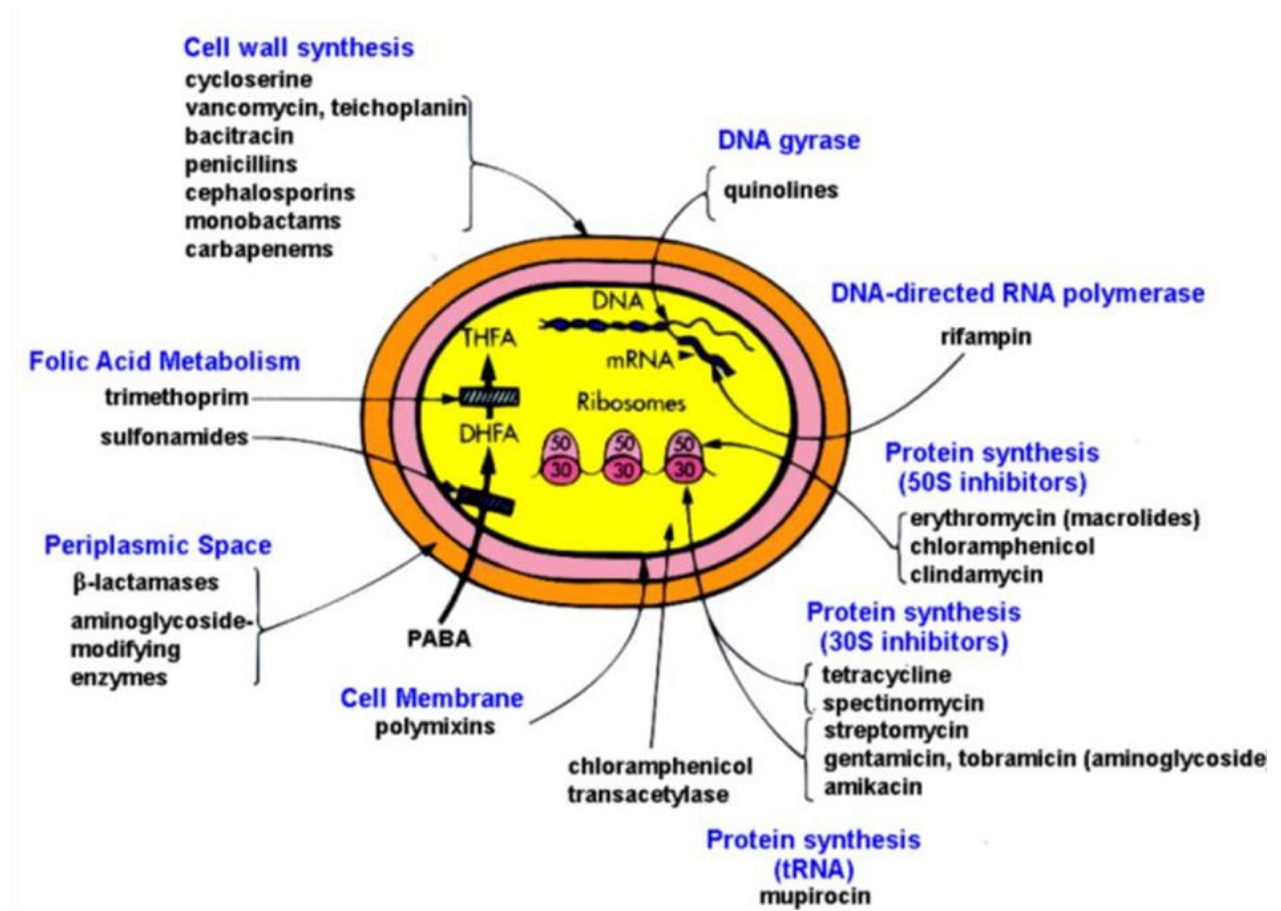


Figure 4: Les différents modes d'action des antibiotiques.

Les différents antibiotiques existant agissent à différents niveaux de la bactérie avec comme objectif principal d'altérer sa survie. Les antibiotiques (en noir) sont regroupés en fonction des processus cellulaires ciblés (en bleu). Sont ainsi illustrées : la synthèse des protéines, la transcription et la voie de biosynthèse de l'acide folique, cette dernière impliquant l'acide para-aminobenzoïque (PABA), utilisé comme précurseur par les bactéries, l'intermédiaire, l'acide dihydrofolique (DHFA) et la forme active, l'acide tétrahydrofolique (THFA).

Chez les bactéries à Gram négatif, différentes enzymes capables de dégrader les molécules signal sont considérées comme d'intéressantes cibles pour développer une stratégie de QQ (FETZNER 2015). Une approche alternative basée sur la séquestration des phéromones par des « molécules cages » comme la cyclodextrine a également été développée (KATO *et al.* 2006). Chez le pathogène à Gram positif *S. aureus*, une stratégie similaire mais basée sur l'utilisation d'anticorps a également été proposée (PARK *et al.* 2007). Enfin, il est possible d'inhiber le mécanisme du quorum sensing grâce à des molécules qui vont entrer en compétition avec la phéromone pour se fixer sur le site actif du récepteur. De nombreuses molécules inhibitrices ont déjà été découvertes, certaines naturelles (TANG & ZHANG 2014) et d'autres synthétiques, mimant la structure des molécules signal (SCUTERA *et al.* 2014).

2. Le quorum sensing chez les bactéries à Gram négatif

2.1. Autoinducteurs de type Acyl Homoserine Lactone

Chez les bactéries à Gram négatif, les molécules signal sont des Acyl Homoserine Lactones (AHL ou Acyl-HSL). Ce sont des molécules possédant un noyau lactone hydrophile sur lequel est fixé un acide gras par son groupement acyle (Figure 5).

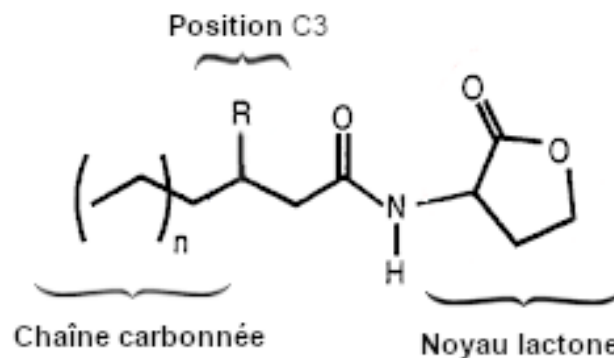


Figure 5: Structure générale d'une Acyl Homoserine Lactone.

La chaîne aliphatique de l'acide gras peut prendre des formes assez variées puisque sa longueur peut être comprise entre 4 et 18 atomes de carbone, elle peut être saturée ou non et posséder ou non un groupement hydroxyle ou cétone en position C3 (CHURCHILL & CHEN 2011).

Cette variabilité sur la chaîne latérale (Figure 6) permet d'assurer une grande spécificité de reconnaissance entre la phéromone et son récepteur (WATERS & BASSLER 2005).

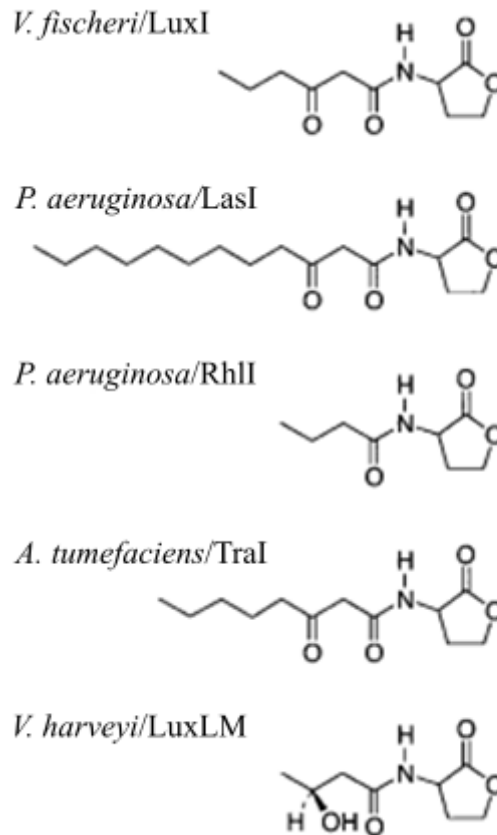


Figure 6: Quelques exemples d'AHLs (SCHAUDER *et al.* 2001).

Les AHLs sont synthétisées par des AHL synthases à partir de deux substrats : la S-adénosyl-L-méthionine (SAM) et une protéine porteuse d'un groupe acyle (R-CO-X) appelée ACP (pour Acyl Carrier Protein) (FAST & TIPTON 2012). Après synthèse, les AHLs sont sécrétées dans l'espace extracellulaire soit par simple diffusion, soit par des transporteurs actifs, probablement en fonction de la longueur et du degré de substitution de la chaîne aliphatique (PEARSON *et al.* 1999). Chez *Pseudomonas aeruginosa*, il a été montré que la pompe à efflux MexAB est responsable de l'export des AHLs à longue chaîne aliphatique mais on observe également de la sécrétion par diffusion passive d'AHLs plus courtes (KROL & BECKER 2014).

Les AHLs sont par la suite réinternalisées par les bactéries réceptrices pour interagir avec un récepteur agissant comme régulateur transcriptionnel pour réguler l'expression des gènes cibles (SALMOND *et al.* 1995). Comme pour la sécrétion, le reimport des petites AHLs se fait par simple diffusion alors que les AHLs possédant une plus grande chaîne aliphatique passent par

un mécanisme de reimport actif. Ainsi, il a été montré que chez *Sinorhizobium meliloti*, un homologue du transporteur d'acide gras FadL d'*Escherichia coli* est impliqué dans le reimport des AHLs à longues chaînes carbonées (KROL & BECKER 2014).

2.2. Les différents systèmes QS à AHL

2.2.1. Les systèmes de type LuxI/LuxR

Le premier système QS à AHL à avoir été décrit est le système LuxI/LuxR (Figure 7) qui régule la bioluminescence chez *Vibrio fischeri* (NEALSON & HASTINGS 1979). Cette bactérie symbiotique colonise les photophores des calamars et permet à son hôte de mieux se défendre contre différents prédateurs.

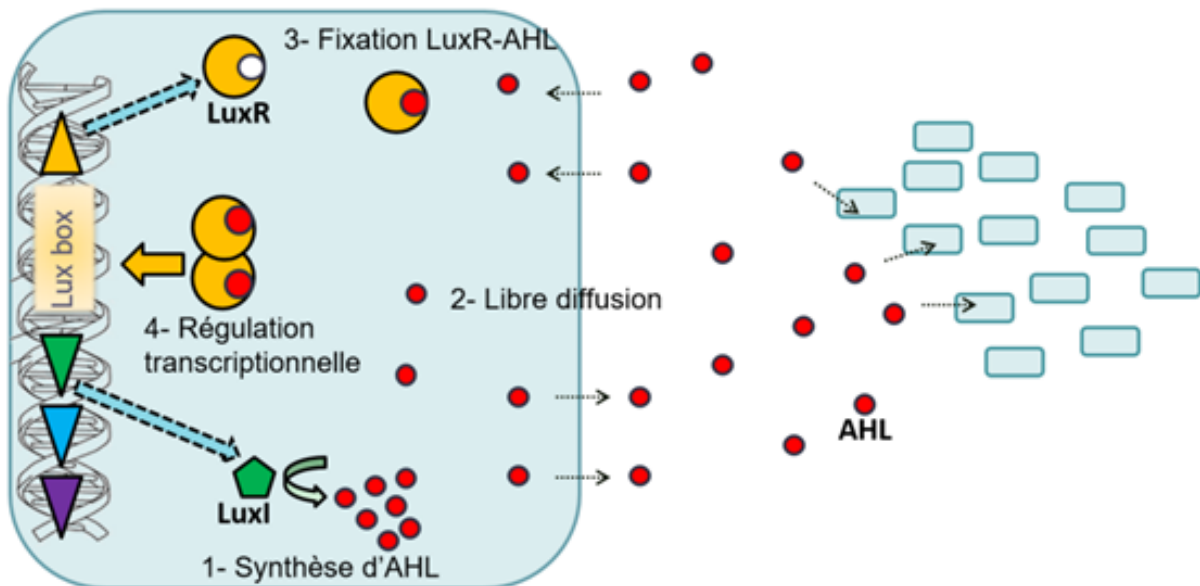


Figure 7: Schéma du système QS LuxI/LuxR. (LANDMAN 2017)

La partie gauche du schéma représente les différentes étapes du système QS LuxI/LuxR d'une bactérie émettrice (rectangle bleu). Les AHLs sécrétées par libre diffusion vont pouvoir être captées par d'autres bactéries réceptrices, à droite sur la figure.

La protéine LuxI porte l'activité AHL synthase alors que LuxR est le récepteur cytoplasmique, sur lequel l'AHL réinternalisée par les bactéries réceptrices vient se fixer, induisant sa dimérisation et son activation. Le complexe LuxR-AHL peut alors se fixer sur la séquence ADN cible appelée Lux-box et activer l'expression de l'opéron luciférase luxICDABE essentiel pour le phénomène de bioluminescence, et celle du gène codant pour le récepteur LuxR lui-même, induisant ainsi l'auto-induction du système (MILLER & BASSLER 2001).

Chez les bactéries Gram négatif, on retrouve une multitude de récepteurs à AHL de type LuxR. Ces régulateurs transcriptionnels possèdent un domaine C-terminal de fixation à l'ADN de type hélice-tour-hélice et un domaine N-terminal de fixation à l'AHL, dont la structure est très conservée (Figure 8) (CHOI & GREENBERG 1991) malgré un pourcentage d'identité de séquence assez faible (20 à 30%) (SALMOND *et al.* 1995). Dans certains cas, comme par exemple chez *Erwinia*, ces récepteurs agissent comme répresseurs (FUQUA & GREENBERG 2002).

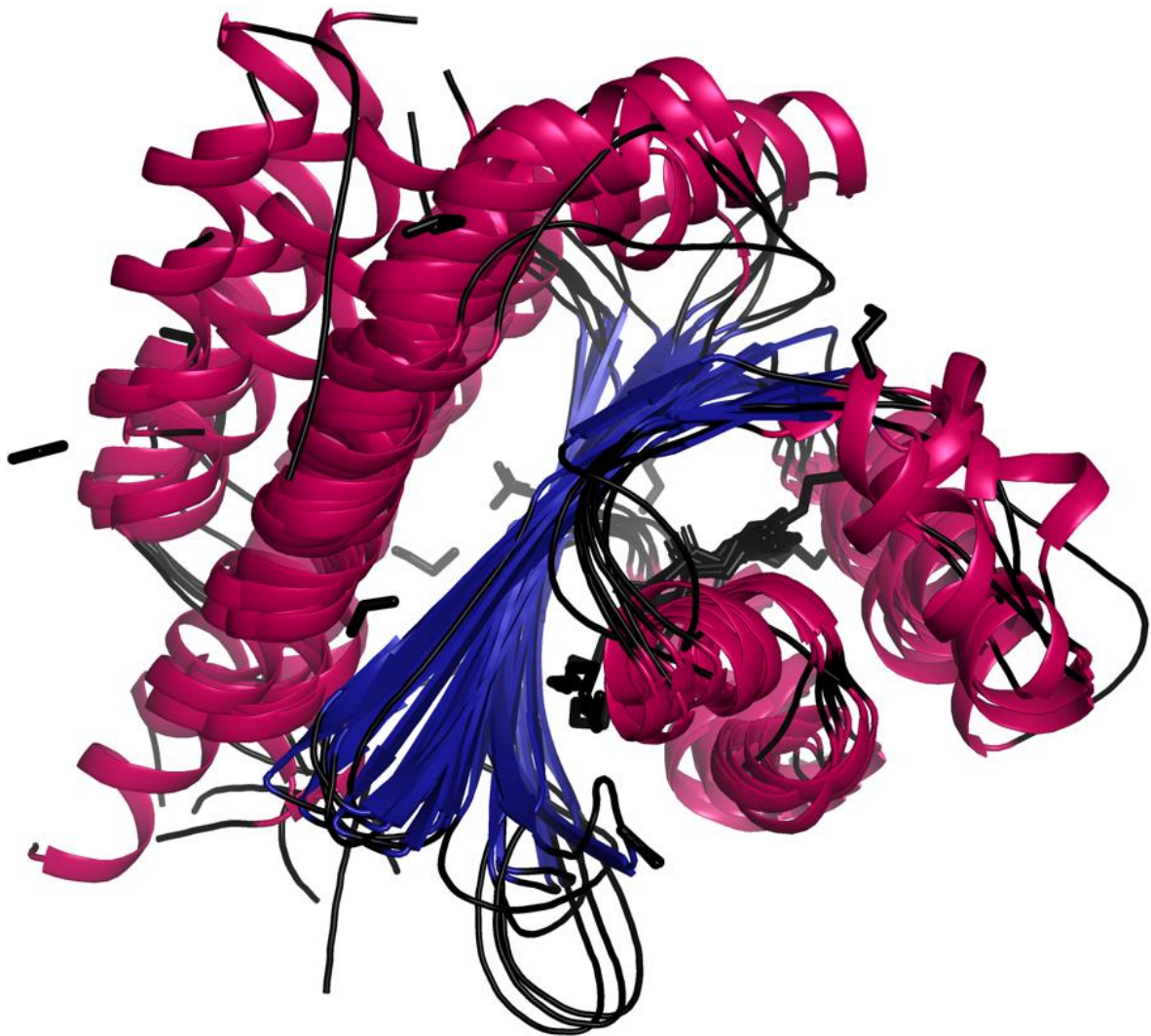


Figure 8: Structure des récepteurs de type LuxR.

Superposition du domaine de fixation de l'AHL de différents membres de la superfamille LuxR 3.30.450.80 selon la classification CATH (<http://www.cathdb.info/version/latest/superfamily/3.30.450.80/superposition>)

2.2.2. Autres systèmes QS à AHL

Alors que le système QS LuxI/LuxR reimporte des AHLs de type 3-oxo-C6-HSL pour interagir avec le récepteur cytoplasmique LuxR, il existe aussi chez *V. fischeri* deux autres systèmes de régulation de la luminescence, AinS/AinR et LuxS/LuxP/Q, qui reconnaissent leurs phéromones au niveau de récepteurs transmembranaires à activité histidine kinase (Figure 9). Une cascade de phosphorylation impliquant les protéines LuxU et LuxO régule alors la synthèse du régulateur transcriptionnel LitR qui contrôle l'expression des gènes *luxR* et *ainS*. Ces trois systèmes QS interviennent à différents niveaux de densité cellulaire, ce qui pourrait permettre une colonisation persistante de l'organisme hôte (LUPP & RUBY 2005; VERMA & MIYASHIRO 2013).

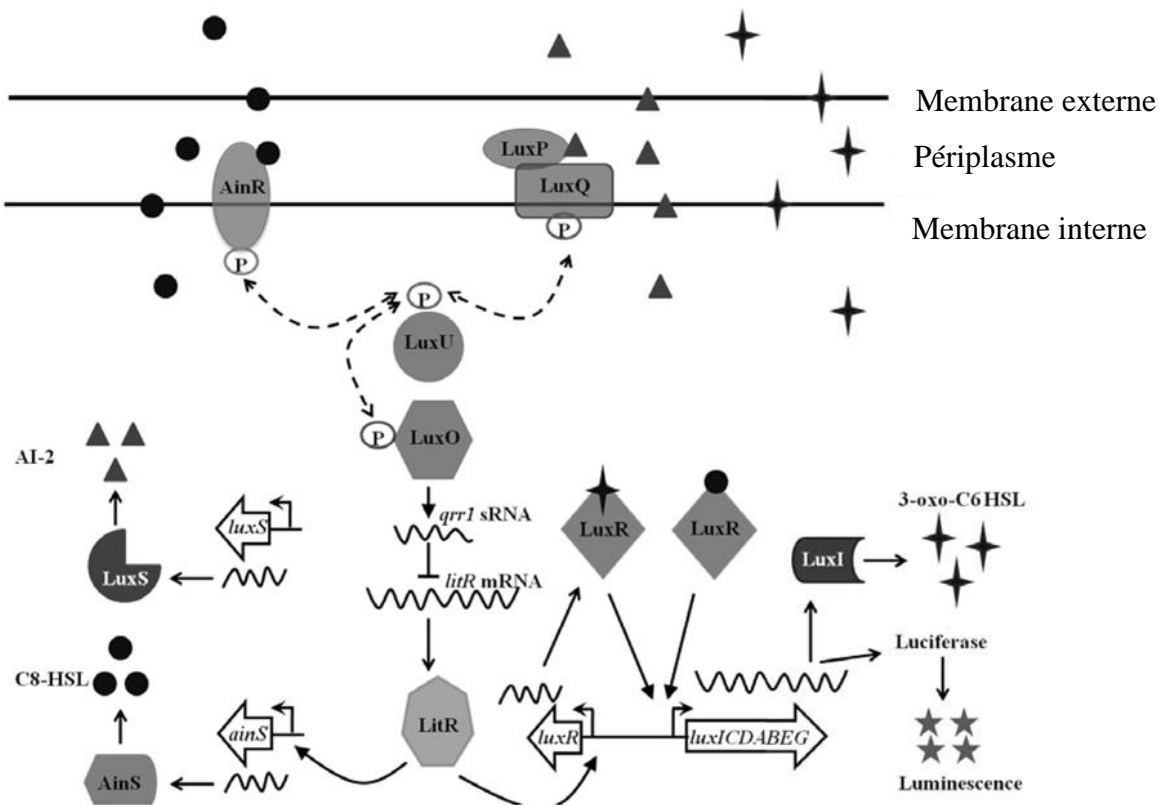


Figure 9: Régulation de l'opéron LuxICDABEG chez *V. fischeri* par 3 systèmes QS (VERMA & MIYASHIRO 2013).

Contrairement au système LuxI/LuxR, les deux systèmes de QS AinS/AinR et LuxS/LuxP/Q utilisent des phéromones qui vont être reconnues par des histidine kinases transmembranaires de systèmes à deux composants. Lorsque la densité cellulaire est faible et que le QS n'est pas activé, le régulateur LuxO est phosphorylé et va activer l'expression de l'ARNs *qrr1* qui va réguler positivement la dégradation de l'ARNm de *litR*. Lorsque la concentration cellulaire augmente, il va également y avoir une augmentation des phéromones C8-HSL (représentées par des cercles noirs) du système AinS/AinR et les phéromones AI-2 (représentées par des triangles noirs) du système LuxS/LuxP/Q. Ces deux phéromones vont se fixer sur leur histidine kinase respective et ainsi pouvoir diminuer la phosphorylation de LuxO en contrôlant la phosphorylation de l'intermédiaire LuxU. Cela va augmenter le niveau de LitR qui active notamment la transcription de LuxR et ainsi pouvoir réguler, via le système LuxI/LuxR, l'opéron LuxICDABEG.

Alors que le système LuxI/LuxR utilise une AHL de type 3-oxo-C6-HSL, AinS synthétise une C8-HSL, qui peut aussi activer LuxR. La protéine LuxS synthétise quant à elle une phéromone d'un autre type, appelé AI-2 (SCHAUDER *et al.* 2001) (Figure 10).

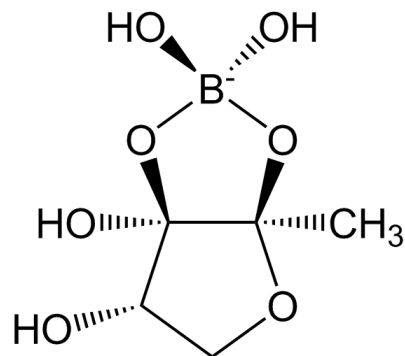


Figure 10 : Structure de l'autoinducteur AI-2.

Le gène *luxS* codant pour l'enzyme responsable de la synthèse de l'AI-2 a été identifié dans la majorité des génomes de bactéries à Gram négatif mais aussi à Gram positif (FEDERLE 2009) et l'AI-2 est considéré comme une molécule signal universelle pouvant être utilisée par les bactéries pour de la communication inter espèces (BASSLER & LOSICK 2006). Cependant, ce point de vue est assez controversé. En effet, AI-2 est un sous-produit du processus de recyclage de la S-adénosyl-L-méthionine (SAM) et n'est peut-être pas systématiquement produit à des fins de signalisation (DIGGLE *et al.* 2007; REZZONICO & DUFFY 2008).

3. Le quorum sensing chez les bactéries à Gram positif

3.1. Les autoinducteurs peptidiques (AIPs)

Les bactéries à Gram positif utilisent quant à elles presque exclusivement des phéromones de nature peptidique que l'on nomme Autoinducer peptides (AIPs) (MILLER & BASSLER 2001). La base de données Quorumpeps (<http://quorumpeps.ugent.be/>) classe tous les AIPs connus depuis 2012 et référence toutes les publications scientifiques impliquant des AIPs (WYNENDAELE *et al.* 2013). Elle permet d'avoir une multitude d'informations structurales et fonctionnelles sur les AIPs et leurs récepteurs (Tableau 1).

Home

E.g. *Staphylococcus*

Current query: *bacillus*

Sequence, Trivial name, SMILES, Molecular formula, Receptor, Method, Species, Literature ?

Search Refine ?

Compare results Clear selection Search hits: 43

ID ?	Sequence	Trivial name	Molecular formula	Species origin	Receptor	Reference ?
10	ADLPFEF	PapR7I	C41H55N7O12	<i>Bacillus thuringiensis</i> , <i>Bacillus cereus</i>	PlcRI, PlcRII, PlcRIII, PlcRIV	3, 37
15	AKDEH	PHRBST1	C24H38N8O10	<i>Bacillus stearothermophilus</i>	Rap	36
16	AKTVQ	phrANTH3	C23H43N7O8	<i>Bacillus anthracis</i>	Rap	36
18	ARNQT	PhrA	C22H40N10O9	<i>Bacillus subtilis</i>	RapA	12, 26
22	CVGIW, thiolacton linkage between C1 and W5	LamD	C27H38N6O5S1	<i>Lactobacillus plantarum</i>	LamC	24
23	ADPITRQW(Farnesyl)GD	ComX 168	C65H99N15O17	<i>Bacillus subtilis</i>	ComP	14, 20, 39
24	A(Abu)F(Abu)LPGGGGVA(Abu)L(Abu)(Dha)EAI, cyclisation between (Ala1-S-Abu2), (Abu4-S-Ala12), (Abu13-S-Ala18) and (Ile19-NH-CH=CH-S-Abu15)	Mersacidin	C80H120N20O21S4	<i>Bacillus subtilis</i>	MrsK2	22, 69
34	DRVGA	PhrI	C20H36N8O8	<i>Bacillus subtilis</i>	Rap	11, 12
44	EKMIG	PhrG	C24H44N6O8S	<i>Bacillus subtilis</i>	Rap	12, 15
49	EQLSFTSIGILQLLTIGTRSCWFFYCRY	PltA	C156H233N37O41S2	<i>Lactobacillus plantarum</i>	PltK	24
50	ERGMT	Competence and Sporulation Factor, CSF, PhrC	C22H40N8O9S	<i>Bacillus subtilis</i> , <i>Bacillus mojavensis</i>	RapC	4, 8, 12

Tableau 1: Présentation de la base de données Quorumpeps. Résultats d'une recherche en utilisant le mot clé Bacillus.

Les AIPs sont synthétisés par les bactéries sous forme de précurseurs qui vont être sécrétés et maturés à l'extérieur des bactéries. On observe dans la base Quorumpeps que la majorité des AIPs sont des peptides linéaires non modifiés. Cependant, certaines bactéries utilisent des AIPs cycliques, ce qui les protègent de la protéolyse. C'est le cas en particulier du système Agr qui régule l'expression des facteurs de virulence chez le pathogène *Staphylococcus aureus* (THOENDEL *et al.* 2011). Le peptide AgrD est synthétisé sous forme d'un pro-peptide linéaire qui est exporté et maturé par le complexe transmembranaire formé par la peptidase signal SpsB et l'endopeptidase AgrB pour donner un peptide mature de 7 à 9 résidus dont les 5 derniers forment une thiolactone cyclique (Figure 11). Dans ce cas, la cyclisation résulte de l'interaction entre le groupement carboxyl C-terminal du peptide et une cystéine conservée, (NG & BASSLER 2009). Le système Agr a également été identifié chez *Clostridium acetobutylicum* (STEINER *et*

al. 2012) et *Listeria monocytogenes* (ZETZMANN *et al.* 2016). Le système Fsr d'*Enterococcus faecalis* utilise quant à lui un peptide cyclique de 11 acides aminés, nommé GBPA (pour Gelatinase biosynthesis-activating pheromone) dont la cyclisation se fait *via* une sérine (NISHIGUCHI *et al.* 2009) (Figure 11).

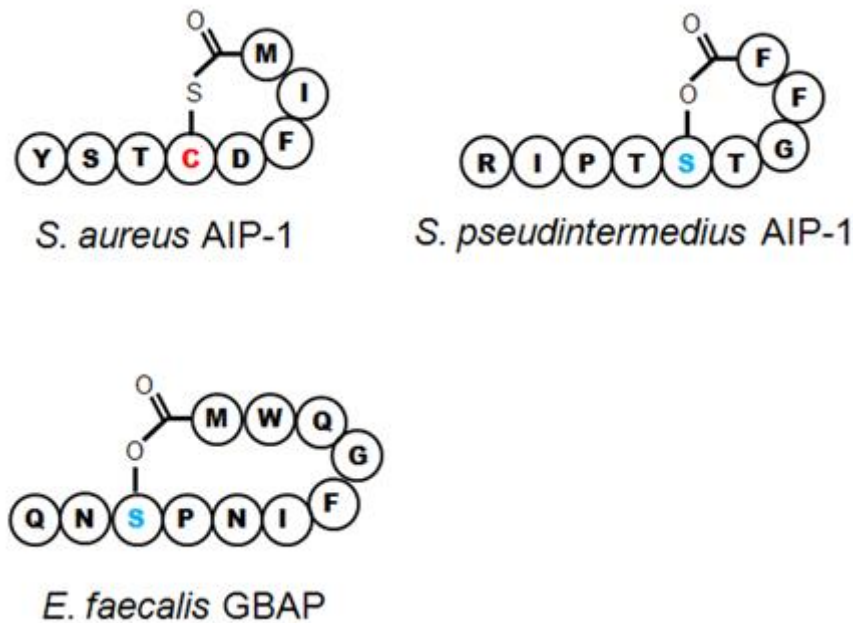


Figure 11 : Différentes formes d'AIP cyclique (GRAY *et al.* 2013).

Dans les systèmes QS dits indirects, les AIPs sont reconnus par des récepteurs transmembranaires à activité histidine kinase qui transmettent le signal au régulateur transcriptionnel *via* une cascade de phosphorylation. Dans les systèmes QS directs, les AIPs sont réinternalisés pour se fixer directement sur des récepteurs cytoplasmiques agissant comme régulateurs transcriptionnels.

3.2. Les systèmes QS indirects

Dans les systèmes QS indirects, les AIPs sont sécrétés dans l'espace extracellulaire par des transporteurs de type ABC (pour ATP Binding Cassette) qui assurent un transport actif ATP-dépendant (KUAN *et al.* 1995; MILLER & BASSLER 2001). La forme mature des AIPs est reconnue par des récepteurs transmembranaires qui, lorsque l'AIP se fixe sur le domaine senseur extracellulaire (Input Domain), s'autophosphorylent de manière ATP-dépendante sur

une histidine du domaine transmetteur cytoplasmique. Dans les cas les plus simples, appelés systèmes à deux composants, le régulateur transcriptionnel qui régule l'expression des gènes cibles (Response Regulator) est directement activé par transfert du phosphate sur un aspartate de son domaine receveur. Il peut alors transmettre le signal en se fixant sur les gènes cibles par son domaine de fixation à l'ADN (Output Domain) (GROSS *et al.* 1989) (Figure 12).

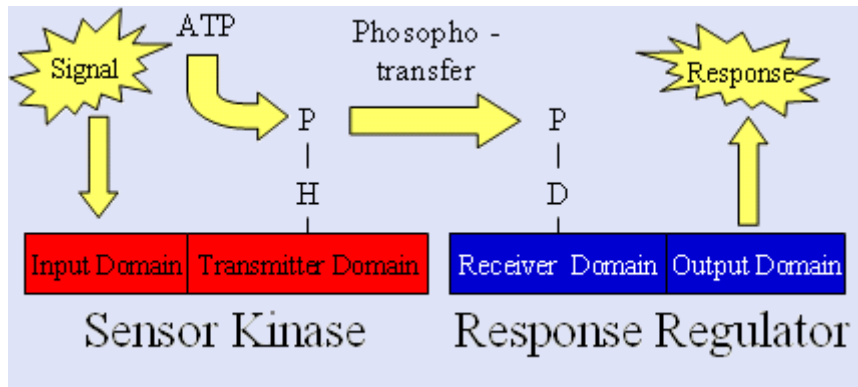


Figure 12: Schéma du fonctionnement d'un système à deux composants.

Un exemple en lien avec mes travaux est le système ComCDE, impliqué dans la régulation de la compétence chez les streptocoques (Figure 13). Le gène *comC* code pour le précurseur du peptide qui est sécrété par l'ABC transporteur ComAB. Lors de cette étape de sécrétion, le peptide subit une première coupure protéolytique pour donner un peptide intermédiaire. Une deuxième étape de maturation au niveau d'un motif Gly-Gly conservé est effectuée par la peptidase de surface SepM. On obtient ainsi un peptide mature pouvant être reconnu par le récepteur transmembranaire ComD. Celui-ci s'autophosphoryle alors de manière ATP-dépendante puis active, par transfert de phosphate, le facteur de transcription ComE. Ce dernier régule l'expression, soit des gènes cibles contrôlant la compétence comme *cipB*, soit ceux codant pour la synthèse de bactériocines. On note que l'expression de *comC* est également activée par ComE, créant ainsi une boucle de rétro activation positive (Figure 13).

Chez *Streptococcus pneumoniae*, ComC est synthétisé sous la forme d'un précurseur de 41 résidus puis mûré en un intermédiaire de 21 résidus avant d'obtenir le peptide mature actif de 17 résidus (HÅVARSTEIN *et al.* 1995).

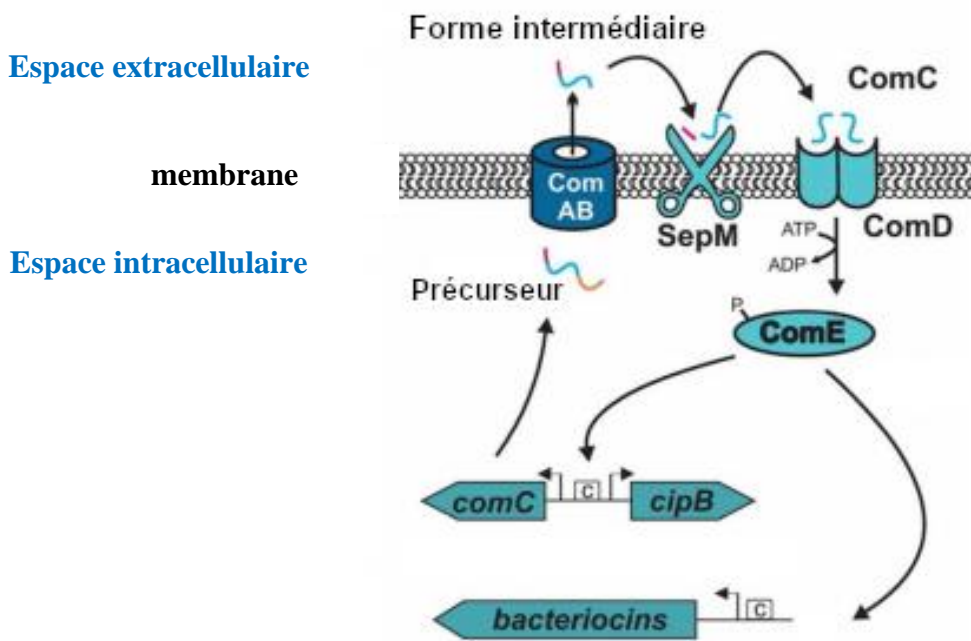


Figure 13: Mécanisme d'activation du système à deux composants ComCDE (COOK & FEDERLE 2014)
 Le peptide ComC, synthétisé sous forme de précurseur, est représenté par trois couleurs. Il va subir une première étape de maturation lors de sa sécrétion par le transporteur ComAB, menant à la sécrétion d'une forme intermédiaire qui subit une deuxième étape de maturation par la peptidase SepM pour donner le peptide mature ComC qui va se fixer sur ComD, un récepteur transmembranaire qui va phosphoryler le régulateur de réponse ComE. Ce dernier va ainsi pouvoir réguler l'expression de différents gènes impliqués dans la synthèse de bactériocines comme cipB.

3.3. Les systèmes QS directs

Les systèmes QS directs, plus récemment identifiés, sont moins bien caractérisés. Il semblerait que les précurseurs d'AIPs soient sécrétés, non pas par des ABC transporteurs comme dans les systèmes QS indirects, mais via le système de transport général appelé système Sec (MONNET *et al.* 2014). Le mécanisme de maturation semble quant à lui être spécifique à l'espèce. Ainsi, la zinc-métalloprotéase Eep participe à la maturation du peptide SHP (short hydrophobic peptide) chez *Streptococcus salivarius* (GARDAN *et al.* 2013) alors que son homologue n'est pas capable de le faire chez *Streptococcus mutans* (KHAN *et al.* 2012). Dans tous les cas, on obtient un peptide mature de 6 à 8 résidus qui va être réinternalisé par des oligopeptides perméases (Opp) de la famille des ABC transporteurs (POTTATHIL & LAZZAZZERA 2003). Les Opp sont composées de 5 sous unités : une protéine de liaison aux oligopeptides extracellulaires qui capture les substrats et qui est ancrée à la membrane, deux protéines transmembranaires qui forment le pore et deux protéines responsables de l'hydrolyse de l'ATP. La protéine de liaison

aux oligopeptides est capable de capturer dans l'environnement proche de la bactérie des peptides de 2 à 18 résidus. Ainsi, chez *Streptococcus thermophilus*, le reimport du peptide XIP est assuré par l'Opp nommée Ami (GARAUULT *et al.* 2002). Les peptides ainsi réinternalisés vont se fixer sur un récepteur cytoplasmique pour activer directement la transcription des gènes cibles. Ces récepteurs forment la famille protéique que l'on nomme famille RNPP (DECLERCK *et al.* 2007) en référence à la première lettre des premiers membres caractérisés : les protéines Rap (pour Response-regulator Aspartate Phosphatases) initialement identifiées chez *Bacillus subtilis* (PEREGO & HOCH 1996; SOLOMON *et al.* 1996), le régulateur bifonctionnel NprR (pour Neutral protease Regulator) (PERCHAT *et al.* 2011), le régulateur majeur de la virulence chez *Bacillus cereus* PlcR (pour Phospholipase C Regulator) (LERECLUS *et al.* 1996) et PrgX, le récepteur de la phéromone sexuelle chez *Enterococcus faecalis* (SHI *et al.* 2005). Plus récemment, on a identifié chez de nombreuses espèces de streptocoques de nouveaux membres de la famille RNPP qui ont été appelés Rgg ou Rgg-like. On trouve des homologues de Rgg uniquement chez les *Bacilli* de l'ordre des *Lactobacillales* et chez la famille *Listeriaceae* de l'ordre des *Bacillales* (Figure 14). Il a été proposé de renommer la famille RRNPP (PEREZ-PASCUAL *et al.* 2016).

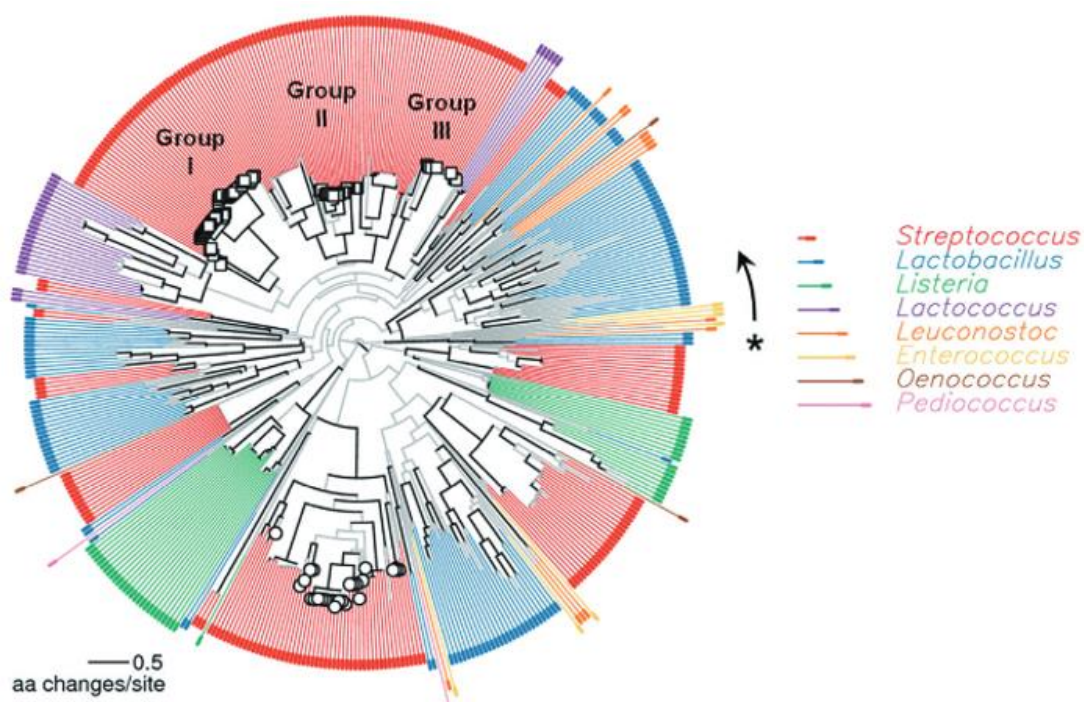


Figure 14 : Arbre phylogénétique des récepteurs Rgg.

La distribution des Rgg est représentée en fonction de leur distance génétique (FLEUCHOT *et al.* 2011).

4. Les récepteurs du QS direct de la famille RNPP

4.1. Structure

Les membres de la famille RNPP sont des régulateurs transcriptionnels qui contrôlent l'expression des gènes cibles, sauf les protéines Rap qui les régulent indirectement (ROCHA-ESTRADA *et al.* 2010). Selon la première définition il faut, pour qu'une protéine soit considérée comme étant un membre de la famille RNPP, (i) que les homologues de cette protéine ne soient présents que chez les *firmicutes*, (ii) que la protéine interagisse directement avec son peptide signal à l'intérieur de la bactérie et (iii) que le peptide soit, d'une part, codé par un gène en aval du gène codant pour la protéine régulatrice, et d'autre part, qu'il soit synthétisé sous la forme d'un précurseur devant être clivé pour devenir actif (DECLERCK *et al.* 2007).

Tous les membres de la famille RNPP possèdent un domaine C-terminal de liaison au peptide constitué d'une succession de 6 à 7 motifs appelés TetratricoPeptide Repeat (TPR). Ce domaine TPR est associé à un domaine N-terminal de liaison à l'ADN de type Helix-Turn-Helix (HTH), à l'exception des protéines Rap chez lesquelles le domaine HTH est remplacé par deux motifs TPR supplémentaires. Ces derniers permettent de fixer une protéine qui va permettre aux Rap de réguler les gènes cibles de manière indirecte. Chez NprR, qui est bifonctionnel, on trouve à la fois ces deux motifs TPR additionnels et un domaine HTH de fixation à l'ADN (ROCHA-ESTRADA *et al.* 2010) (Figure 15).

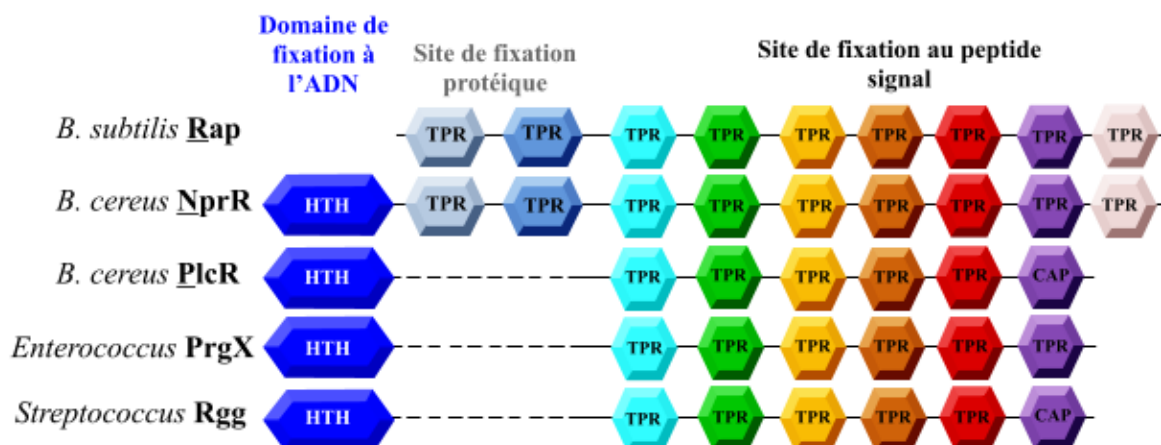


Figure 15 : Schéma de la structure des différentes protéines (R)RNPP.

En bleu est représenté le domaine HTH et chaque motif TPR est représenté avec une couleur différente.

Le dernier motif TPR du régulateur transcriptionnel PlcR ne contient qu'une seule hélice nommée capping helix (CAP) (DECLERCK *et al.* 2007). On retrouve la même configuration dans certains Rgg et Rgg-like (TALAGAS *et al.* 2016; MIGNOLET *et al.* 2019).

4.1.1. Le domaine TPR

Les motifs TPR (Tetratricopeptide Repeat) forment un domaine d'interaction protéine – protéine observé aussi bien chez les procaryotes que chez les eucaryotes (BLATCH & LÄSSLE 1999). Un motif TPR est composé de 34 acides aminés formant deux hélices α antiparallèles (Figure 16 – A). Les séquences des TPR sont peu conservées et difficilement identifiables (Prosite documentation PDOC50005). Cependant, une séquence signature a quand même été définie (SIKORSKI *et al.* 1990). On y trouve notamment des acides aminés hydrophobes qui permettent d'empiler chaque motif TPR sur son voisin avec un décalage d'environ 24° . On obtient ainsi un domaine en forme de superhélice droite au cœur duquel le peptide vient se fixer (Figure 16 – B).

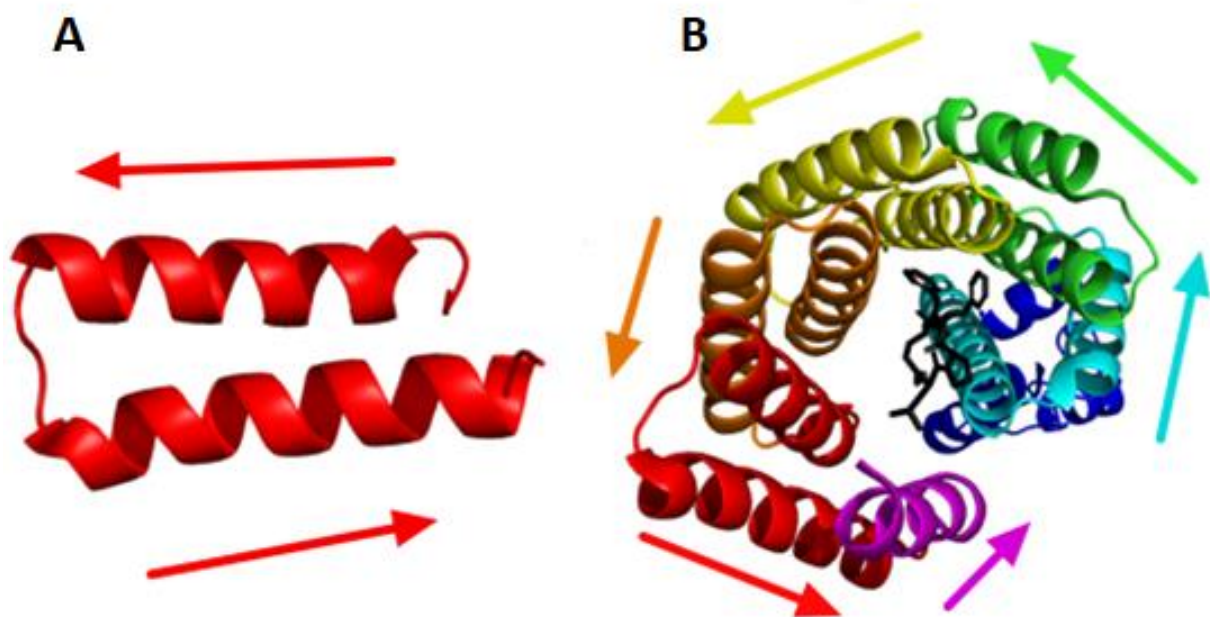


Figure 16 : Structure des domaines à motifs TPR.

A- Motif TPR composé de deux hélices antiparallèles. B- La superhélice du domaine TPR de PlcR (Code PDB : 2QFC) (DECLERCK *et al.* 2007) représentée en cartoon, coloré par motif TPR, avec le peptide fixé et représenté en stick et en noir.

4.1.2. Le domaine HTH

Les domaines de liaison à l'ADN de type Helix-Turn-Helix (HTH) se replient rapidement et indépendamment du reste de la protéine (RELIGA *et al.* 2007). Initialement identifiés chez de nombreux procaryotes (WINTJENS & ROOMAN 1996), les domaines HTH peuvent avoir des architectures assez différentes (Figure 17) mais sont toujours composés d'au minimum 2 hélices α (RELIGA *et al.* 2007). L'hélice de reconnaissance (nommée α_3) se fixe dans le grand sillon de l'ADN, établissant des interactions spécifiques avec les paires de bases de la séquence cible.

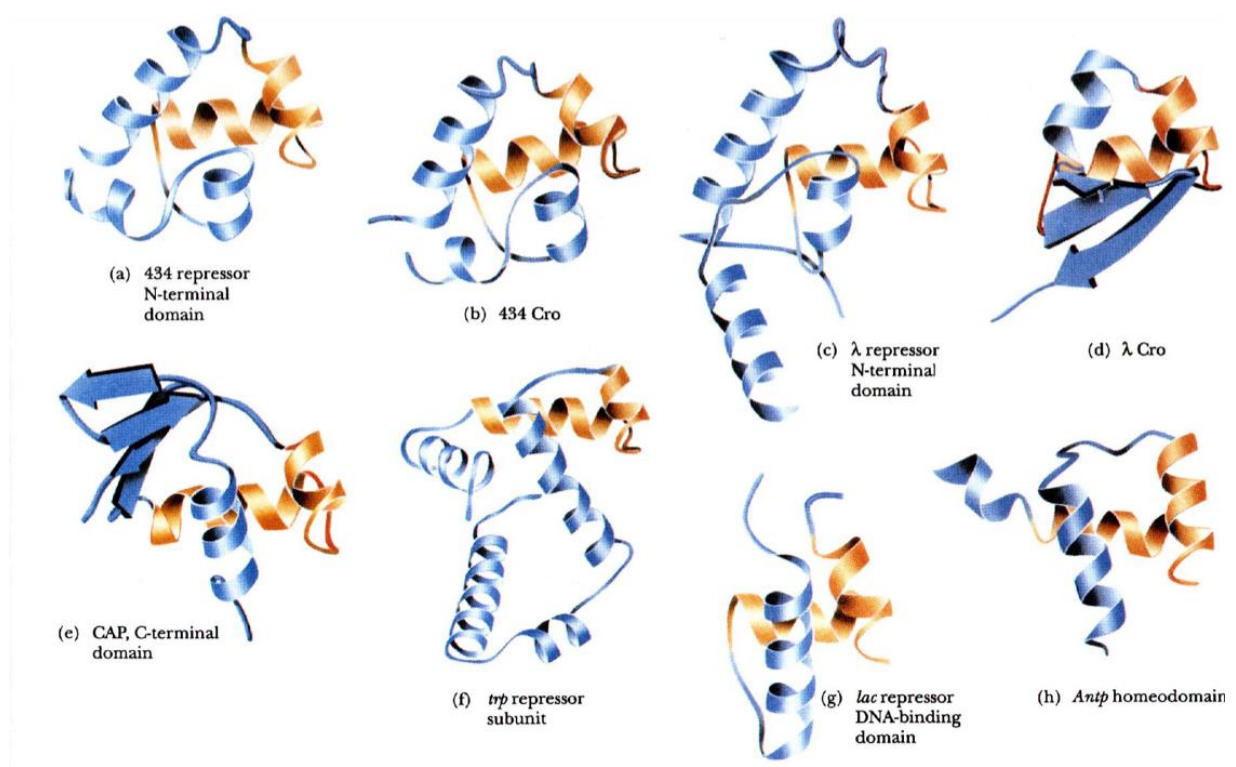


Figure 17 : Structures de différents types de domaines HTH

Les différents types de domaine HTH sont représentés en cartoon avec le motif conservé (hélice-tour-hélice) jaune et la partie variable en bleu (PELLEGRINI-CALACE & THORNTON 2005).

Les domaines HTH se fixent à l'ADN sous forme de dimères, ce qui va permettre une reconnaissance spécifique d'une séquence palindromique ou pseudo palindromique de l'ADN (ARAVIND *et al.* 2005). Les deux demi-sites de fixation sont éloignés d'un tour de double hélice, c'est-à-dire environ 34 Å (Figure 18).

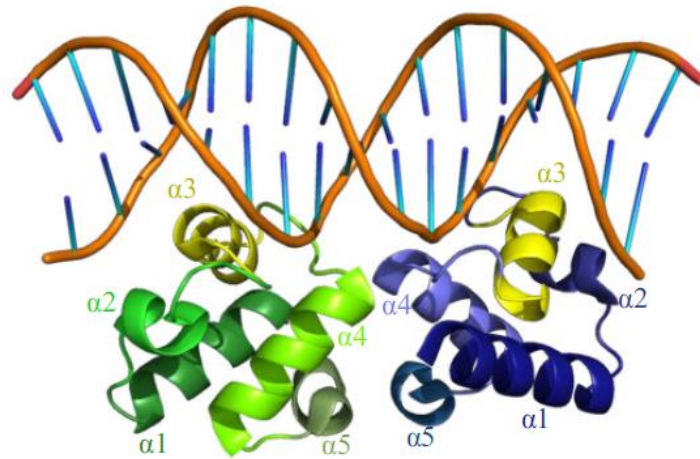


Figure 18 : Structure de deux domaines HTH fixés sur un ADN cible.
 Les deux HTH représentés dans un gradient de vert ou de bleu sont fixés sur l'ADN double brin cible (Code PDB : 5JUB) (TALAGAS *et al.* 2016). Dans chaque HTH, l'hélice $\alpha 3$ est représentée en jaune.

Les domaines HTH des RNPPs appartiennent à la sous-famille de type Cro/C1 (Prosite entry PS50943). Ils sont composés de 5 hélices α et comportent une soixantaine de résidus. On retrouve ce type de domaines de fixation à l'ADN à l'extrémité N-terminale de nombreux régulateurs transcriptionnels, aussi bien eucaryotes que procaryotes.

4.2. Régulation des RNPPs par les AIPs

4.2.1. Le système Rap/Phr

Les protéines Rap forment une large famille très conservée (CORE & PEREGO 2003). Chez *B. subtilis* où elles ont initialement été identifiées et caractérisées, on compte 11 isoformes, appelées RapA à RapK, présentant plus de 25% d'identité de séquence. Elles sont régulées par des pentapeptides de séquences très variables que l'on nomme Phr (pour Phosphatases regulator).

Les protéines Rap ne portent pas toutes une activité phosphatase et reconnaissent différents partenaires protéiques, contrôlant ainsi des processus bactériens divers (Figure 19) (POTTATHIL & LAZAZZERA 2003). En effet, alors que RapA, RapB, RapE, RapH, RapI, et RapJ sont capables de déphosphoryler Spo0F, la phosphotransférase impliquée dans l'initiation de la sporulation (PARASHAR *et al.* 2013), les isoformes RapC, RapD, RapF, RapG, RapH, et RapK séquestrent le domaine de fixation de l'ADN du régulateur transcriptionnel ComA, qui régule la

compétence (OGURA & FUJITA 2007; PARASHAR *et al.* 2013). Enfin, RapG contrôle la production de protéases extracellulaires (OGURA *et al.* 2003) et RapI, en plus de déphosphoryler Spo0F, peut également avoir un rôle dans la régulation de la mobilité de l'élément génétique ICEBs1 (AUCHTUNG *et al.* 2005).

La fixation du peptide Phr entraîne une perte d'affinité entre la protéine Rap et son partenaire protéique, ce qui a pour conséquence une activation de la sporulation ou de la compétence dans les cas de Spo0F ou ComA (SOLOMON *et al.* 1996; PEREGO & BRANNIGAN 2001).

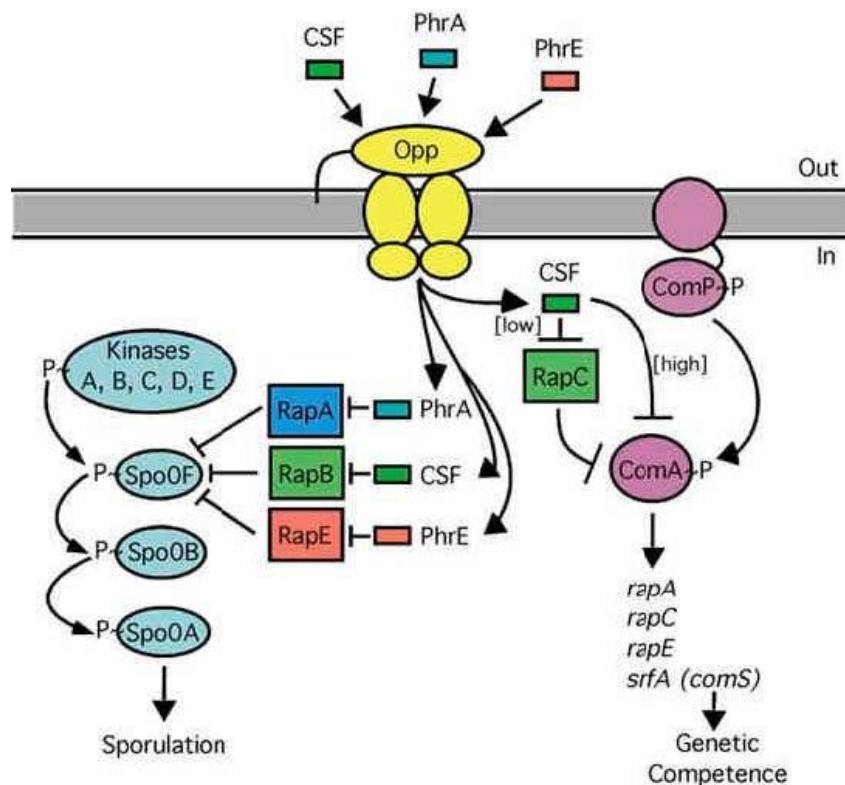


Figure 19 : Schéma du mode d'action des systèmes Rpa/Phr chez *Bacillus subtilis* (POTTATHIL & LAZZERA 2003).

Une fois réinternalisées par des oligopeptides perméases (Opp), les phéromones peptidiques PhrA, CSF et PhrE, représentées respectivement en vert, bleu et rouge vont respectivement inhiber les phosphatases RapA, RapB et RapE impliquées dans la cascade de phosphorylation qui régule la sporulation. D'autre part, le peptide CSF et la protéine RapC sont également impliqués dans la régulation de la compétence via le système à deux composants ComP/ComA.

De nombreuses structures cristallines de protéines Rap ont été résolues. La comparaison des formes apo de RapF (Code PDB : 4I9E) (GALLEGO DEL SOL & MARINA 2013) et RapI (Code PDB : 4I1A) (PARASHAR *et al.* 2013) révèle, qu'en absence de peptide, le domaine TPR à 9 motifs des Rap est particulièrement flexible. Alors qu'il adopte une conformation en superhélice dans la structure de RapI, les 2 motifs TPR additionnels du N-terminus ont totalement basculé pour adopter une conformation en faisceau de 3 hélices dans la structure de RapF (Figure 20). Les structures des complexes RapF/ComA (Code PDB : 3ULQ) (BAKER &

NEIDITCH 2011) et RapH/Spo0F (Code PDB : 3Q15) (PARASHAR *et al.* 2011) montrent que cette conformation en faisceau de 3 hélices est directement impliquée dans la fixation des protéines cibles, qui, de manière intéressante, ne se fixent pas sur la même face de la protéine.

Enfin, la structure du complexe RapJ/PhrC (Code PDB : 4GYO) (PARASHAR *et al.* 2013) montre que la fixation du peptide dans la poche du domaine TPR stabilise la conformation en superhélice incompatible avec le basculement de la partie N-terminale et la fixation de la protéine cible. Toutes ces structures permettent de décrire le mécanisme d'activation des protéines Rap (Figure 20).

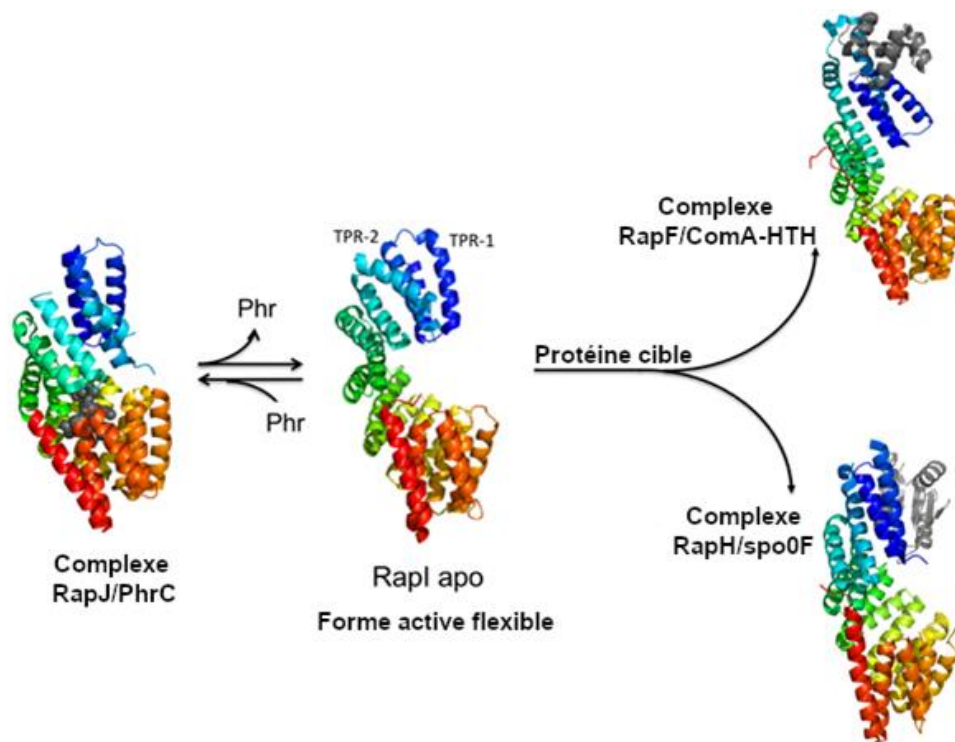


Figure 20 : Mécanisme moléculaire de la régulation des systèmes Rap/Phr.

La protéine est représentée en ruban et est colorée dans un gradient de couleur du bleu (Nter) au rouge (Cter).

Les protéines cibles, Spo0F et le domaine HTH de ComA, sont représentés sous forme de ruban gris. Les atomes du peptide Phr sont représentés sous forme de sphères noires. La structure du complexe RapJ/PhrC (Code PDB : 4GYO) (PARASHAR *et al.* 2013) montre que la fixation du peptide stabilise une conformation fermée inactive de la protéine. La structure de la forme apo de RapI (Code PDB : 4I1A) (PARASHAR *et al.* 2013) montre qu'en absence de peptide la protéine adopte une conformation ouverte très flexible qui permet aux motifs TPR 1 et 2 de basculer complètement pour adopter une conformation dite en « 3-helix bundle » compatible avec la fixation des protéines cibles, comme le montre les structures des complexes RapF/ComA (Code PDB : 3ULQ) (BAKER & NEIDITCH 2011) et RapH/Spo0F (Code PDB : 3Q15) (PARASHAR *et al.* 2011).

De manière surprenante, alors que les formes apo des protéines Rap sont décrites comme monomériques en solution (BAKER & NEIDITCH 2011; PARASHAR *et al.* 2011), la structure cristalline de RapF (Code PDB : 4I9E) (GALLEGO DEL SOL & MARINA 2013) suggère la formation d'un dimère stable possédant une interface de 2342.1Å² et le même mode de

dimérisation est observé dans la structure de RapI (Code PDB : 4I1A) (PARASHAR *et al.* 2013), avec néanmoins une interface beaucoup plus faible de 1423.7Å².

4.2.2. Le système PlcR/PapR

PlcR est un régulateur majeur de la virulence identifié chez *Bacillus cereus* (LERECLUS *et al.* 1996). En présence du peptide mature PapR (SLAMTI & LERECLUS 2002), il régule 45 gènes codant majoritairement pour des enzymes de dégradation, des toxines, des protéines membranaires de protection et des senseurs permettant de détecter la concentration de phéromones dans le milieu extracellulaire (AGAISSE *et al.* 1999; GOHAR *et al.* 2008). Il régule également positivement le gène *papR* qui code pour un peptide précurseur de 48 acides aminés qui, après sécrétion par le système SEC et une première protéolyse donnant une forme intermédiaire de 27 résidus, va être clivé par la Neutral protease B (NprB) pour donner un peptide mature de 7 résidus (BOUILLAUT *et al.* 2008) qui va être réinternalisé par une oligopeptide perméase pour venir se fixer directement sur PlcR (Figure 21).

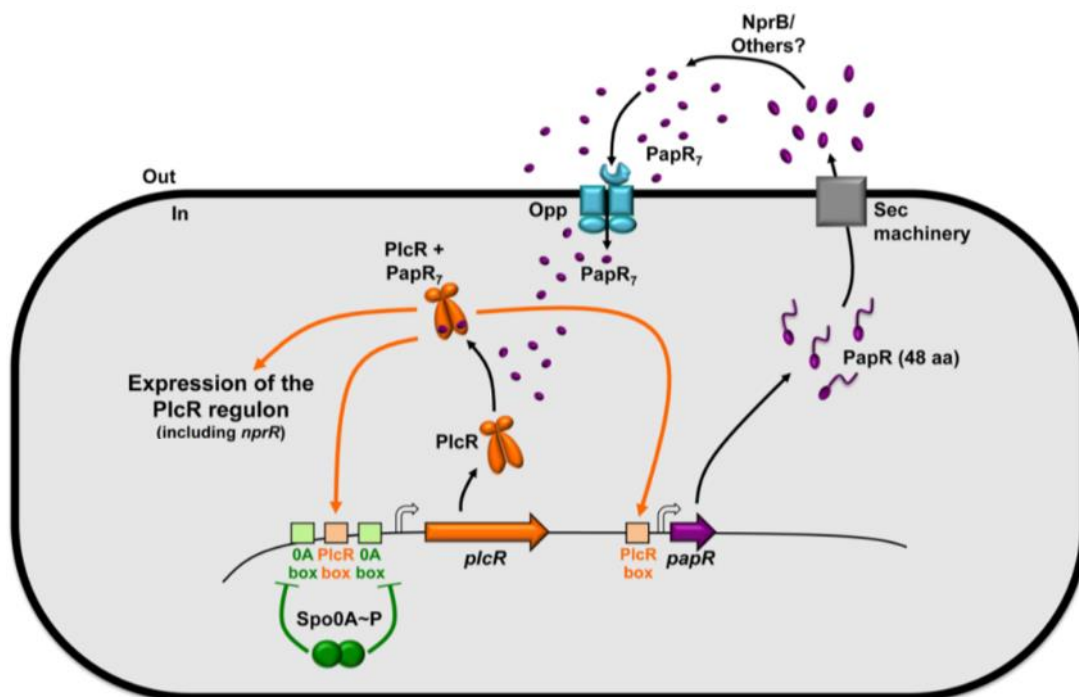


Figure 21 : Schéma du mode d'action du système PlcR-PapR chez *B. thuringiensis*

Le gène *papR* (en violet) code pour un peptide précurseur de 48AA qui va être sécrété par des systèmes SEC dans le milieu extracellulaire où il va être clivé par la peptidase NprB. Le peptide mature de 7 AA, PapR₇, va être réinternalisé par les oligopeptides perméases des bactéries réceptrices. Sa fixation sur le récepteur PlcR (en orange) permettra d'activer l'expression des gènes cibles (régulon PlcR) mais aussi les gènes *plcR* et *papR*. Il est à noter que pendant la phase stationnaire, l'inhibition des gènes *plcR* et *papR* est assurée par Spo0A-P, le régulateur de la sporulation.

Les systèmes PlcR/PapR ont été classés en 4 phérogroupe sur la base de la séquence du peptide PapR (SLAMTI & LERECLUS 2005). Bien que la forme mature biologique de PapR semble être un heptapeptide (BOUILLAUT *et al.* 2008), cette classification a été réalisée sur la base de la forme minimale active de 5 résidus provenant de 29 souches différentes de *B. cereus* et *B. thuringiensis* (Figure 22).

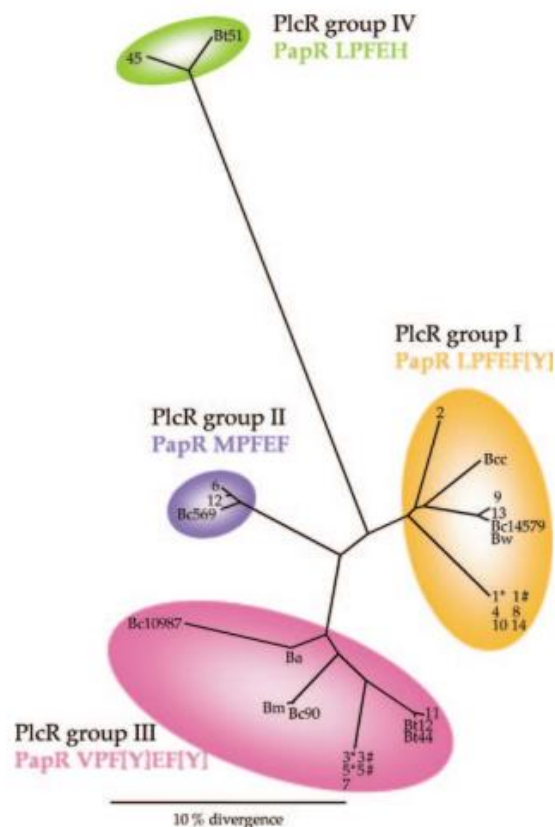


Figure 22 : Classification des systèmes PlcR/PapR de *B. cereus* (SLAMTI & LERECLUS 2005)

La comparaison de séquences de peptide papR a permis de proposer une classification en 4 phérogroupe dont l'arbre phylogénétique est présenté sur la figure. Chaque couleur correspond à un groupe distinct. Le groupe I (en orange) possède un peptide LPFE(F/Y), le groupe II (en violet) un peptide MPFEF, le groupe III (en rose) indique un peptide VP(F/Y)E(F/Y) et le groupe IV (en vert) un peptide LPFEH.

Chez *B. cereus*, un paralogue de PlcR, nommé PlcRa, a été identifié, ainsi que son peptide associé, PapRa. Les deux protéines présentent 29% d'identité de séquence. PlcRa est essentiellement impliqué dans la régulation du métabolisme du soufre et la réponse au stress oxydatif (HUILLET *et al.* 2012). Chose intéressante, alors que PlcRa est activé aussi bien par PapR (ADLPFEF) que par PapRa (CSIPYEY), PlcR ne reconnaît que son propre peptide.

La structure de PlcR a également été résolue sous différentes formes, ce qui a permis d'élucider le mécanisme d'activation de PlcR par PapR (Figure 23). La protéine sous forme apo (Code PDB : 4FSC) (GRENHA *et al.* 2013) est caractérisée par une longue hélice linker reliant les domaines TPR et HTH. Celle-ci est formée par la jonction des hélices $\alpha 5$ du domaine HTH et la première hélice du motif TPR-1. La protéine forme un dimère très compact avec une surface d'interaction de 3140\AA^2 dans laquelle les hélices linker s'enroulent en coiled-coil. Chaque domaine HTH interagit par son hélice $\alpha 1$ avec le domaine TPR de l'autre sous-unité et les hélices $\alpha 3$ ne sont pas correctement orientées pour pouvoir interagir avec l'ADN. La structure du complexe binaire PlcR-PapR (Code PDB : 2QFC) (DECLERCK *et al.* 2007) montre que la fixation du peptide n'entraîne qu'un subtil changement de conformation. La superposition du dimère du complexe sur la forme apo montre en effet un rmsd de 1.916 sur 247 C α . Le principal changement concerne l'interface de dimérisation qui diminue à 2400\AA^2 , en particulier à cause des hélices linker qui ne s'enroulent plus en coiled-coil, ce qui donne plus de flexibilité à la protéine. Enfin, la structure du complexe ternaire PlcR-PapR-ADN (Code PDB : 3U3W) (GRENHA *et al.* 2013) montre que le relâchement du complexe induit par la fixation de PapR permet, en présence d'ADN, d'introduire une cassure franche de l'hélice linker au niveau du résidu Ile 68, à l'extrémité de l'hélice $\alpha 5$. Les deux domaines HTH peuvent alors se réorienter de façon à positionner les hélices $\alpha 3$ dans les deux demi-sites de la séquence ADN cible.

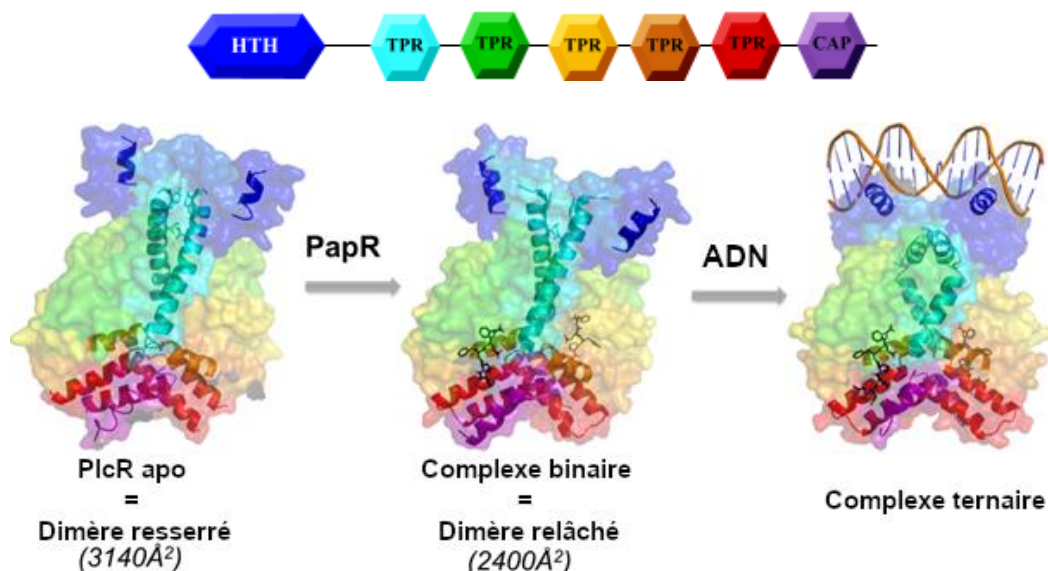


Figure 23 : Mécanisme de régulation de PlcR. Adapté de (GRENHA *et al.* 2013).

La protéine est représentée en surface transparente et chaque sous-unité du dimère est coloré selon le code couleur indiqué sur le schéma en haut de la figure. Les hélices directement impliquées dans le changement de conformation responsable de l'activation de PlcR sont montrées en ruban. Les grandes hélices linker qui se cassent en deux lors de la fixation de l'ADN sont colorées en cyan. Le peptide est représenté en sticks noirs.

4.2.3. Le système NprR/NprX

Comme indiqué dans la (Figure 15), le récepteur NprR possède à la fois un domaine HTH de fixation à l'ADN et les motifs TPR supplémentaires impliqués dans la fixation des protéines chez les protéines Rap. Il a été montré que cela reflète sa bi-fonctionnalité (PERCHAT, TALAGAS, ZOUHIR, *et al.* 2016) (Figure 24). En présence de son peptide NprX, NprR agit comme régulateur transcriptionnel et régule 41 gènes impliqués dans le nécrotrophisme de la bactérie, c'est-à-dire sa survie au sein du cadavre de l'hôte infecté (DUBOIS *et al.* 2012). En absence de son peptide, NprR régule négativement la sporulation de *B. thuringiensis* en se liant à la phosphotransférase Spo0F et en la déphosphorylant, à la manière d'une Rap phosphatase (PERCHAT, TALAGAS, PONCET, *et al.* 2016).

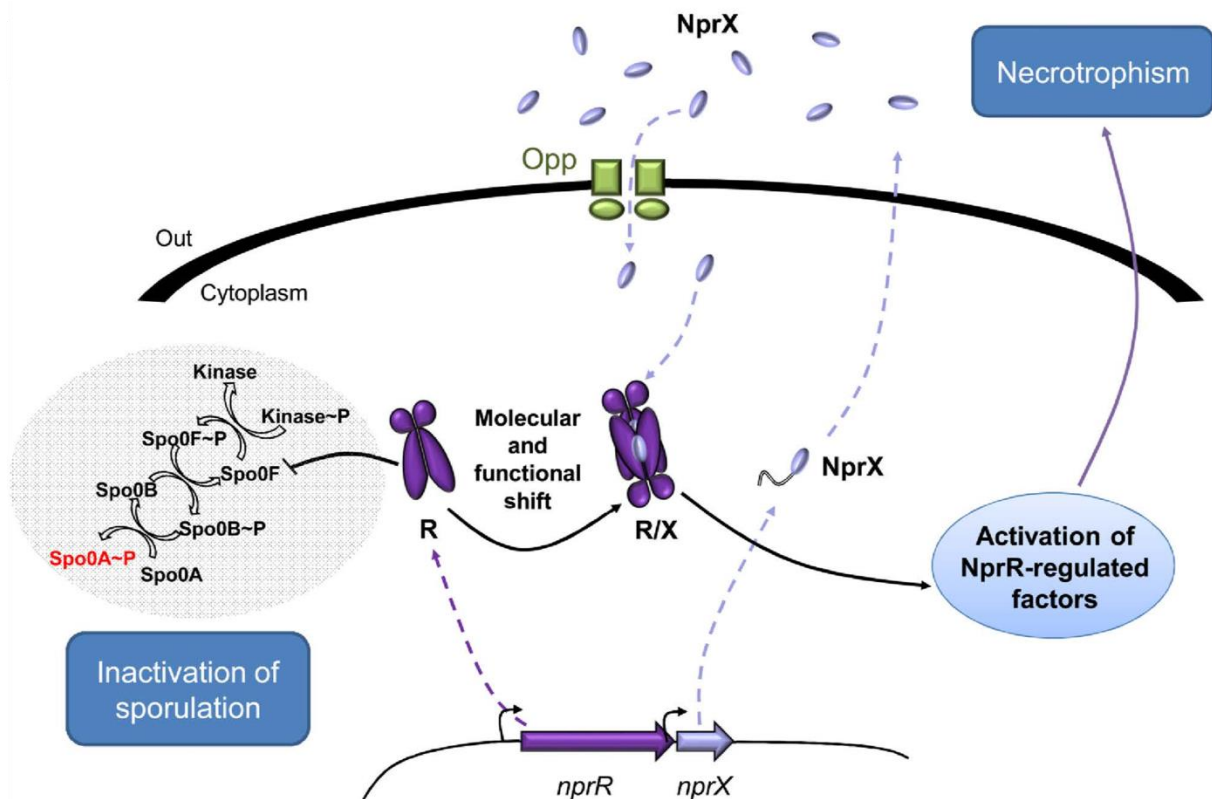


Figure 24 : Schéma du mode d'action bifonctionnel de NprR chez *B. thuringiensis* (PERCHAT, TALAGAS, PONCET, *et al.* 2016)

En absence du peptide NprX, la forme apo de la protéine NprR est dimérique. Elle peut alors se fixer sur la phosphotransférase Spo0F, ce qui va entraîner l'inactivation de la sporulation. Lorsque le peptide NprX, synthétisé sous forme de précurseur avant d'être sécrété puis maturé en une forme active de 8AA est réinternalisé par des oligopeptides perméases, le complexe qu'il forme avec NprR va adopter une structure tétramérique incompatible avec la fixation de Spo0F mais permettant sa fixation sur l'ADN. NprR passe ainsi d'une fonction inhibiteur de la sporulation à une activité facteur de transcription pour réguler l'expression de gènes impliqués dans le nécrotrophisme.

Comme pour le système PlcR/PapR, par comparaison des séquences du peptide NprX dans différentes souches du groupe *Bacillus cereus*, 7 phérogroupe ont été proposés (Figure 25) (PERCHAT *et al.* 2011). Cette étude a de plus révélée l'existence de cross-talk entre les différents phérogroupe. Ainsi, la protéine NprR-I du groupe I peut être activée par le peptide NprX-I (SKPDIVG) mais aussi par NprX-II (SKPDTYG). Cependant, elle n'arrive pas à fixer les peptides NprX-IV (SRPDVLT) et NprX-VII (WRPDMSI) qui ont des séquences plus éloignées de NprX-I. Fait intéressant, le peptide NprX-III (SNPDIYG) qui diffère seulement de 2 AA par rapport à NprX-I (SKPDIVG), peut se fixer sur NprR-I sans l'activer. Ce peptide est donc un quorum quencheur du système NprR/NprX du phérogroupe I.

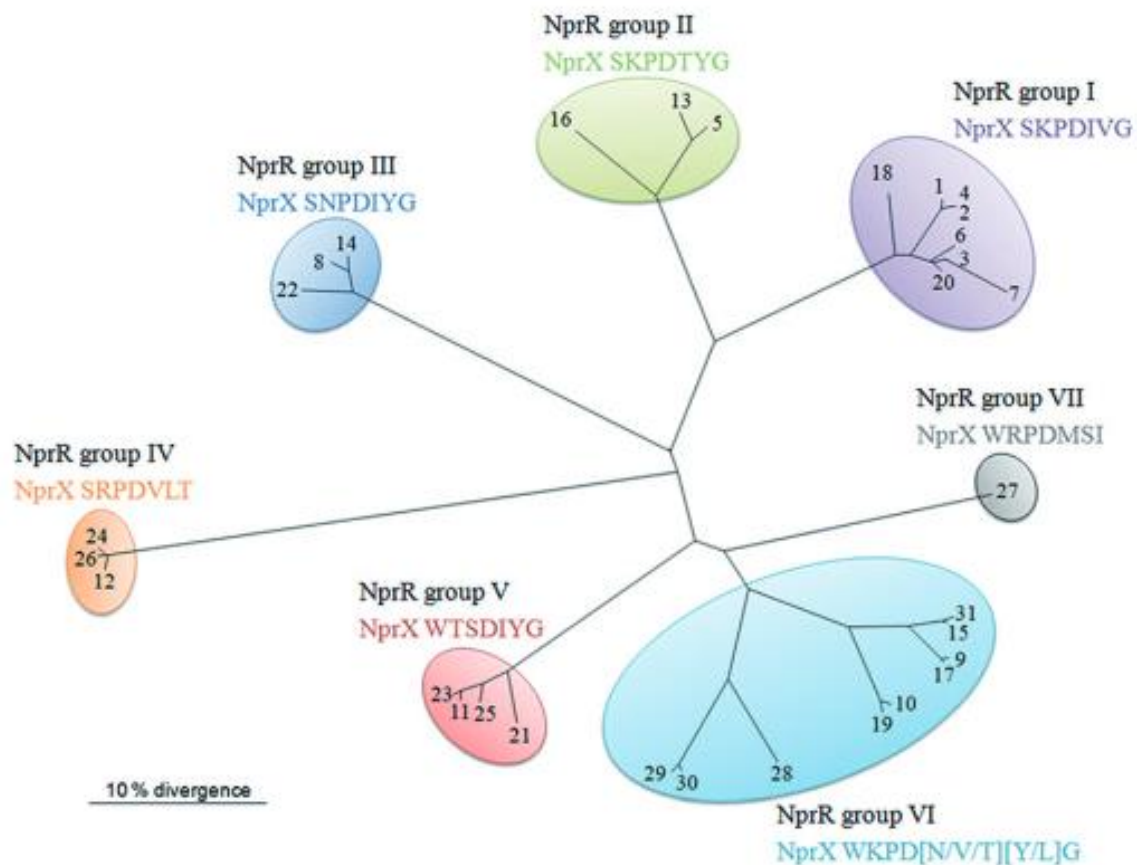


Figure 25 : Classification des systèmes NprR/NprX de *B. cereus* (PERCHAT *et al.* 2011).

La comparaison des séquences des peptides NprX a permis de proposer une classification des systèmes NprR/NprX de *B. cereus* en 7 phérogroupe dont l'arbre phylogénétique est présenté sur la figure. Chaque couleur correspond à un phérogroupe différent. Le groupe I (en violet) possède un peptide de séquence SKPDIVG, le groupe II (en vert) un peptide SKPDTYG, le groupe III (en bleu) un peptide SNPDIYG, le groupe IV (en orange) un peptide SRPDVLT, le groupe V (en rose) un peptide WTSDIYG et enfin le dernier groupe VI un peptide légèrement plus variable qui possède comme séquence consensus WKPD(N/V/T)(Y/L)G.

La forme entière de NprR n'ayant pas pu être cristallisée, l'analyse structurale a été réalisée à l'aide d'une forme tronquée de la protéine, NprR(Δ HTH), déletée de son domaine HTH. Des analyses en solution ont montré qu'en absence de peptide, NprR est dimérique (ZOUHIR *et al.* 2013). La structure cristalline du dimère de NprR(Δ HTH) a été résolue (Code PDB : 5DBK) (PERCHAT, TALAGAS, PONCET, *et al.* 2016) (Figure 26). Elle révèle des conformations très différentes pour les 2 sous-unités, démontrant que, tout comme les protéines Rap, le domaine TPR de NprR est flexible en absence de peptide. Il a donc été proposé que les 2 motifs TPR additionnels de NprR puissent adopter la conformation en faisceau de 3 hélices observée chez les Rap pour fixer Spo0F. Malgré l'absence des domaines HTH dans la structure, on voit que le mode de dimérisation des domaines TPR n'est pas compatible avec la dimérisation des domaines HTH nécessaire à l'interaction avec l'ADN.

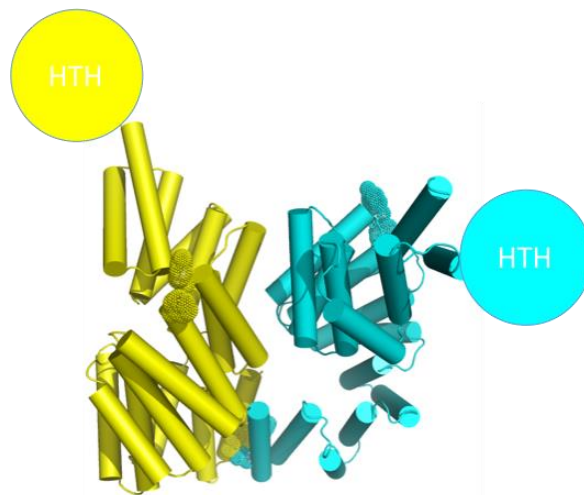


Figure 26 : Structure de la forme apo de NprR(Δ HTH)

La protéine est représentée en cartoon et colorée par chaîne. Les domaines HTH tronqués sont représentés par des cercles de la même couleur que le domaine TPR associé (PERCHAT, TALAGAS, PONCET, *et al.* 2016).

De manière intéressante, le mode de dimérisation de NprR est similaire à celui observé dans la structure cristalline de la protéine RapF (Code PDB : 4I9E) et RapI (Code PDB : 4I1A) (Figure 27). Ceci suggère que ces dimères cristallins de Rap pourraient avoir une réalité biologique, malgré les données en solution qui suggèrent au contraire une forme monomérique.

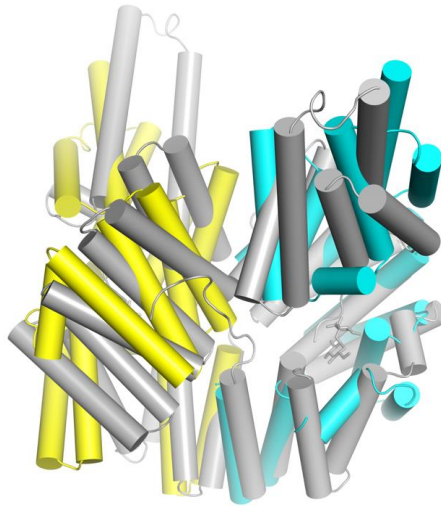


Figure 27 : Superposition des dimères de NprR et de RapF.
 NprR apo (Code PDB : 5DBK) est coloré par chaîne comme dans la Figure 26. Le dimère RapF (Code PDB : 4I9E), également représenté en cartoon, est coloré en gris.

Les données en solution montrent d'autre part que NprR adopte une structure tétramérique en présence du peptide NprX. La structure du complexe NprR(Δ HTH)-NprX (Code PDB : 4GPK) (ZOUHIR *et al.* 2013) montre que la fixation du peptide stabilise une conformation compacte du domaine TPR de NprR, permettant à deux dimères de s'associer en un tétramère compatible avec la dimérisation des domaines HTH et une interaction avec 2 sites ADN (Figure 28).

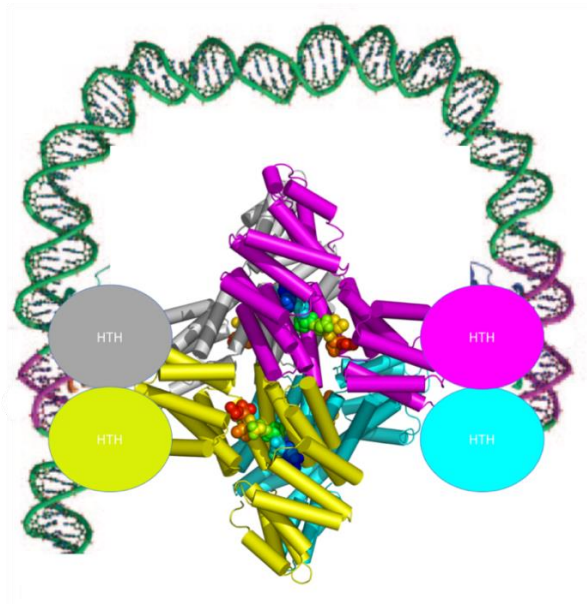


Figure 28 : Modèle d'interaction entre le complexe NprR/NprX et un ADN cible.
 La structure du tétramère de la forme tronquée NprR(Δ HTH) en complexe avec NprX (Code PDB : 4GPK) (ZOUHIR *et al.* 2013) est représentée en cartoon avec une couleur par chaîne. Le peptide fixé sur chacune des 4 sous-unités est représenté en sphères et coloré en arc-en-ciel. Les domaines HTH tronqués sont schématisés par des cercles de la même couleur que le domaine TPR correspondant. Un fragment d'ADN interagissant avec les 2 dimères de HTH est modélisé en vert

L'interaction avec 2 sites ADN explique probablement la difficulté rencontrée pour identifier la NprR-box. La structure du complexe ternaire NprR/NprX/ADN n'a donc pas encore pu être résolue. D'autre part, on ne sait pas si le complexe NprR-Spo0F existe sous forme monomérique ou dimérique. Un autre point qu'il reste à élucider concerne la position du domaine HTH, dans la forme apo et en présence de Spo0F. Les difficultés rencontrées pour cristalliser la protéine entière suggère que le domaine HTH pourrait être relié au domaine TPR par un linker flexible donnant lieu à une hétérogénéité de conformations incompatible avec l'obtention de cristaux. Des analyses complémentaires doivent donc encore être menées pour pouvoir proposer un mécanisme moléculaire complet de la régulation du système NprR/NprX.

4.2.4. Le système PrgX

Le récepteur PrgX est un régulateur transcriptionnel qui régule l'opéron *prgQ* codant pour des gènes impliqués dans le transfert conjugatif chez *Enterococcus faecalis* (KOZLOWICZ *et al.* 2006). PrgX présente la spécificité d'être régulé positivement et négativement par deux peptides différents, cCF10 et iCF10. En présence de iCF10, PrgX va réprimer l'expression de l'opéron *prgQ* alors que le complexe PrgX/cCF10 va lever la répression, ce qui va permettre le mécanisme de conjugaison (CHANDLER & DUNNY 2004). Les cellules dépourvues du plasmide pCF10 portant le gène du peptide iCF10 sont appelées cellules réceptrices. Elles vont produire et sécréter le peptide cCF10 qui va être reimporté et produire une réponse chez les cellules possédant le peptide pCF10 appelées cellules donneuses (CHANDLER & DUNNY 2004).

La structure cristalline de la forme apo de PrgX (Code PDB : 2AXU) (SHI *et al.* 2005) a révélé la formation d'un tétramère, présentant une structure quaternaire différente de celle observée avec le complexe NprR/NprX (Figure 29).

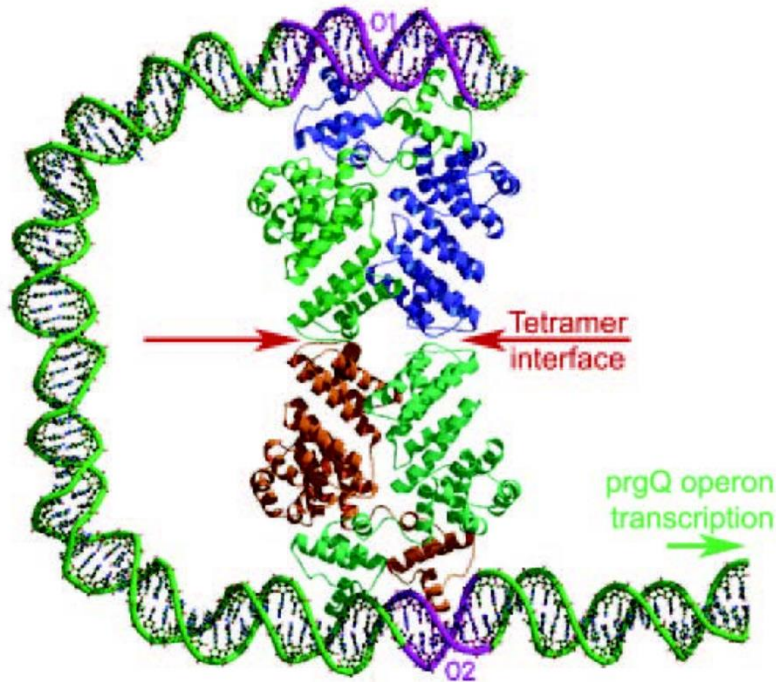


Figure 29 : Modèle du complexe PrgX-ADN.

La structure cristalline du tétramère de PrgX (Code PDB : 2AXU) suggère la reconnaissance de 2 sites distincts sur l'ADN (SHI *et al.* 2005) montrée en cartoon, avec un dimère vert/bleu et un dimère vert/rouge. L'interface du tétramère est indiquée par 2 flèches rouges. Un modèle d'interaction du tétramère, avec un fragment hypothétique molécule d'ADN (en vert) et les deux opérateurs (en violet), permet de montrer comment chacun des deux dimères pourraient interagir via leur domaines HTH avec un site opérateur de l'opéron prgQ (O1 et O2, en violet).

Les structures de PrgX en complexe avec le peptide activateur cCF10 (Code PDB : 2AXZ) (SHI *et al.* 2005) et le peptide répresseur iCF10 (Code PDB : 2GRM) (KOZLOWICZ *et al.* 2006) montrent que cCF10 et iCF10 se fixent sur le même site mais provoquent des changements de conformation différents. Alors que la fixation de iCF10 ne semble pas déstabiliser le tétramère, celle de cCF10 provoque un changement de conformation au niveau de la partie C-terminale de la protéine qui pourrait conduire à une rupture des tétramères et à l'ouverture de la boucle d'ADN (SHI *et al.* 2005; KOZLOWICZ *et al.* 2006). Cette étude a permis de proposer un mécanisme de régulation basé sur la reconnaissance de 2 sites distincts sur l'ADN par le tétramère de PrgX, provoquant la formation d'une boucle d'ADN qui empêcherait l'accès à l'ARN polymérase.

4.2.5. Le système ComR/ComS

Le récepteur ComR, initialement nommé Ster_0316 fait partie de la famille des RNPP de streptocoque appelés Rgg-like (FONTAINE, BOUTRY, *et al.* 2010). Avec son peptide ComS,

ComR constitue un système QS alternatif au système à 2 composants ComCDE (Figure 13) pour réguler la compétence *via* le facteur sigma ComX/SigX mais aussi la synthèse de bactériocine pour lier la compétence à la prédation (WENDERSKA *et al.* 2012) (MIGNOLET *et al.* 2018) (Figure 30). La majorité des espèces ne possèdent que l'un des systèmes ComCDE ou ComRS, mais certaines d'entre elles, comme *S. mutans*, possèdent les deux (Martin *et al.* 2006)

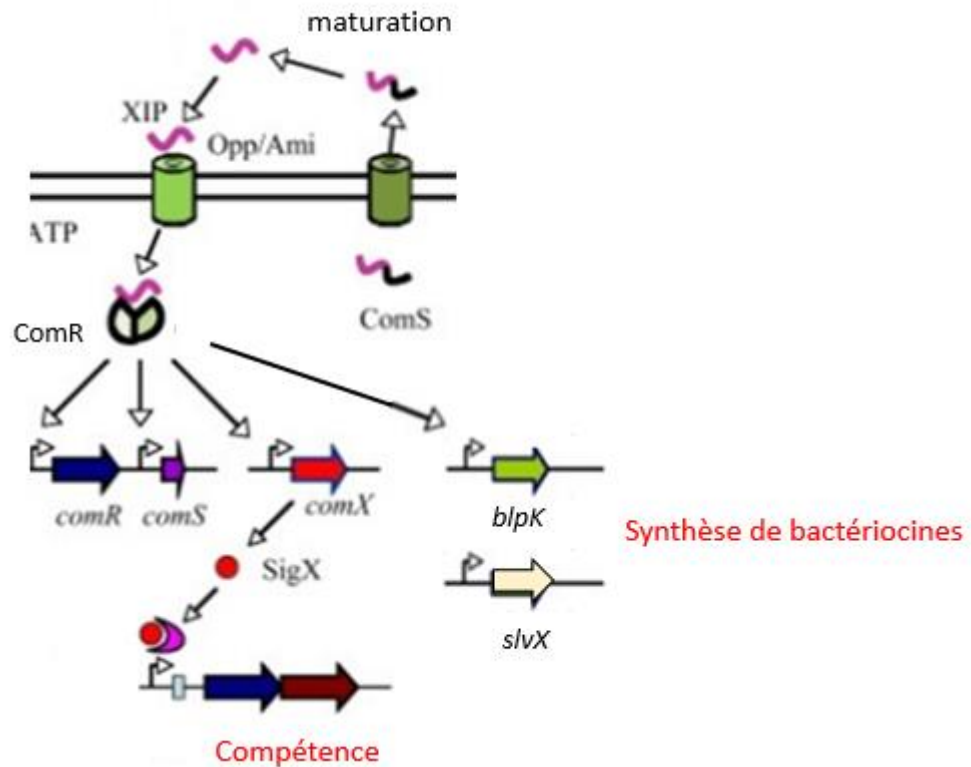


Figure 30 : Schéma ComRS chez *S. mutans*.

Le système direct ComRS va sécréter un précurseur ComS pour donner un peptide mature que l'on nomme XIP. Ce peptide va être réinternalisé pour se fixer sur le régulateur transcriptionnel ComR. Le complexe ComR/XIP va ainsi pouvoir réguler les gènes *ComR* et *ComS* pour sa propre auto régulation positive mais aussi *ComX* qui code pour SigX le maître régulateur de la compétence chez les Streptocoques et enfin différents gènes comme *blpK* et *slvX* qui sont impliqués dans la synthèse de bactériocines.

L'analyse des séquences de la forme mature de ComS, appelée XIP (pour ComX/SigX Inducing Peptide), a permis de classer les systèmes ComRS en 3 types différents (Figure 31). Les peptides de type I, caractérisés par la présence d'un motif (V/L) P (F/Y), sont observés chez les 3 espèces du groupe *salivarius*, i.e. *S. salivarius* (LPYFTGCL), *S. thermophilus* (LPYFAGCL) et *S. vestibularis* (VPPFMIYY). Le type II quant à lui correspond aux peptides possédant un motif « WW » avec dans certains cas des résidus acides (D ou E) et/ou basiques (K, R ou H). C'est le système ComRS qui se trouve entre autres dans les espèces de *Streptococcus* les plus

pathogènes (*S. pyogenes* M1GAS (SAVDWWRL) et *S. mutans* UA159 (MGLDWWSL)). Parce qu'il ne rentrait pas dans les deux groupes précédents, le peptide XIP de *S. suis* (WGTWVEE) qui contient deux résidus aromatiques non adjacents, a été classé dans un type III.

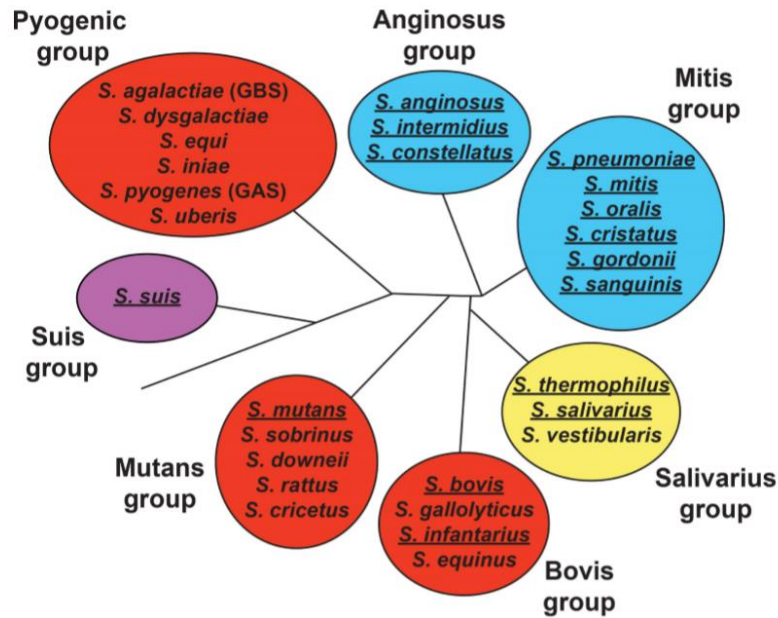


Figure 31 : Classification des voies de compétence chez les Streptocoques (SHANKER *et al.* 2016)
 Le système ComRS de type I se trouve dans le groupe Salivarius (en jaune). Le système ComRS de type II est présente dans les groupes de Bovis, Mutans et Pyogenes (en rouge). Le système ComRS de type III est caractéristique du groupe Suis (en violet). Les groupes Anginosus et Mitis (en bleu) présentent quant à eux le système ComCDE.

Il a été montré que, suivant les espèces, la taille minimale active du XIP est comprise entre 7 et 9 résidus (FONTAINE, DANDOY, *et al.* 2010; MASHBURN-WARREN *et al.* 2010; DESAI *et al.* 2012; KHAN *et al.* 2012; MORRISON *et al.* 2013; MARKS *et al.* 2014; ZACCARIA *et al.* 2014). Chez *S. thermophilus*, la production de ComS dans des conditions artificielles a montré que la peptidase Eep pouvait maturer ComS en un peptide XIP de différentes tailles, avec une forme majoritaire de 11 résidus. Néanmoins, on ne sait pas si les différentes longueurs de ComS cohabitent ou non dans les souches sauvages (FONTAINE *et al.* 2013; GARDAN *et al.* 2013).

L'analyse structurale du système ComRS de *S. thermophilus* (TALAGAS *et al.* 2016) a montré que la structure de la forme apo de la protéine (Code PDB : 5JUF), est un monomère caractérisé par la séquestration de l'hélice $\alpha 3$ du domaine HTH par le domaine TRP (Figure 32).

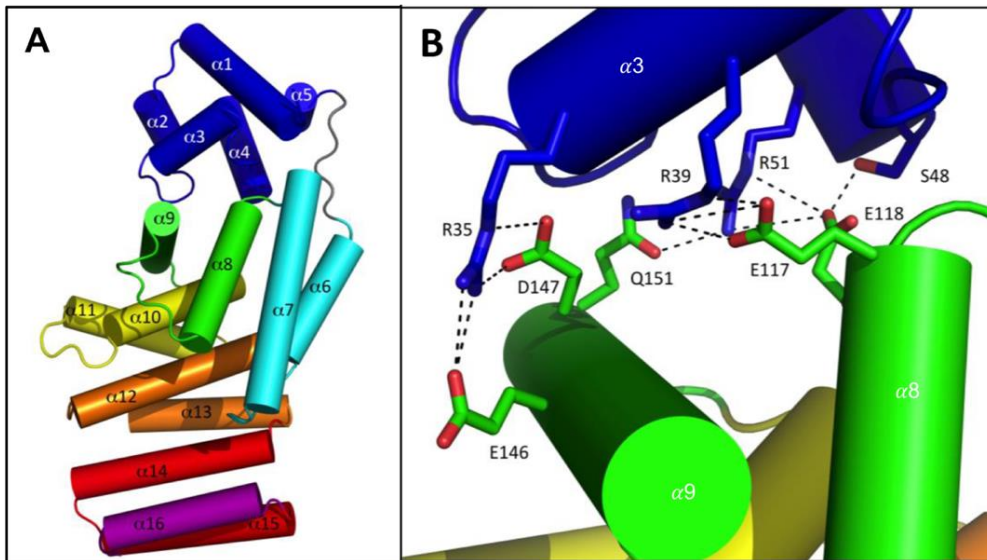


Figure 32 : Structure de la forme apo de ComR.

A- Monomère de ComRStH apo (Code PDB : 5JUF) représenté en cartoon et coloré en gradient de couleur du bleu (domaine HTH N-terminal) au violet (hélice CAP C-terminale). Les hélices sont numérotées. B- Zoom sur les ponts salins entre l'hélice $\alpha 3$ du domaine HTH et les hélices $\alpha 8$ et $\alpha 9$ du domaine TPR

La structure du complexe ternaire ComR/XIP/ADN (Code PDB : 5JUB) montre la formation d'un dimère asymétrique caractérisé par un linker flexible reliant les domaines HTH et TPR (Figure 33). Les domaines HTH, libérés, interagissent en dimère avec l'ADN cible, *via* leurs hélices $\alpha 3$ qui se positionnent alors dans le grand sillon des deux demi-sites de la ComR-box.

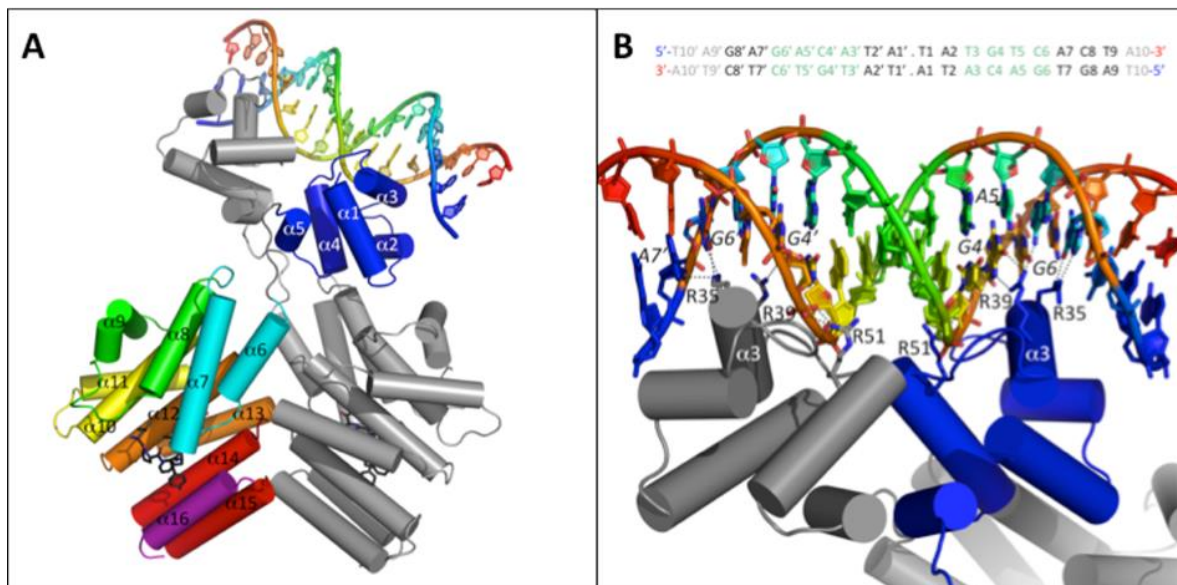


Figure 33 : Structure du complexe ternaire ComRStH/XIPStH/ADN (Code PDB : 5JUB) (TALAGAS *et al.* 2016)

A- Une sous unité est colorée en gradient de couleur du bleu au violet. L'autre est en gris. Chaque brin d'ADN est également coloré du bleu (5') au rouge (3'). B- Zoom sur l'interaction entre le dimère de domaines HTH et la séquence d'ADN indiquée au-dessus.

Une analyse par SEC-MALS ayant montré que le complexe binaire ComR-XIP était dimérique en solution, un mécanisme d'activation avait alors été proposé (Figure 34), suggérant que la fixation du peptide faisait passer la superhélice TPR de ComR d'une conformation ouverte à une conformation plus compacte, induisant le relargage du domaine HTH et la dimérisation des domaines TPR (Figure 35). Cependant, le complexe ComR/XIP n'a jamais été cristallisé chez *S. thermophilus* et l'on ne sait pas si une conformation monomérique avec le domaine HTH libre peut exister ou non.

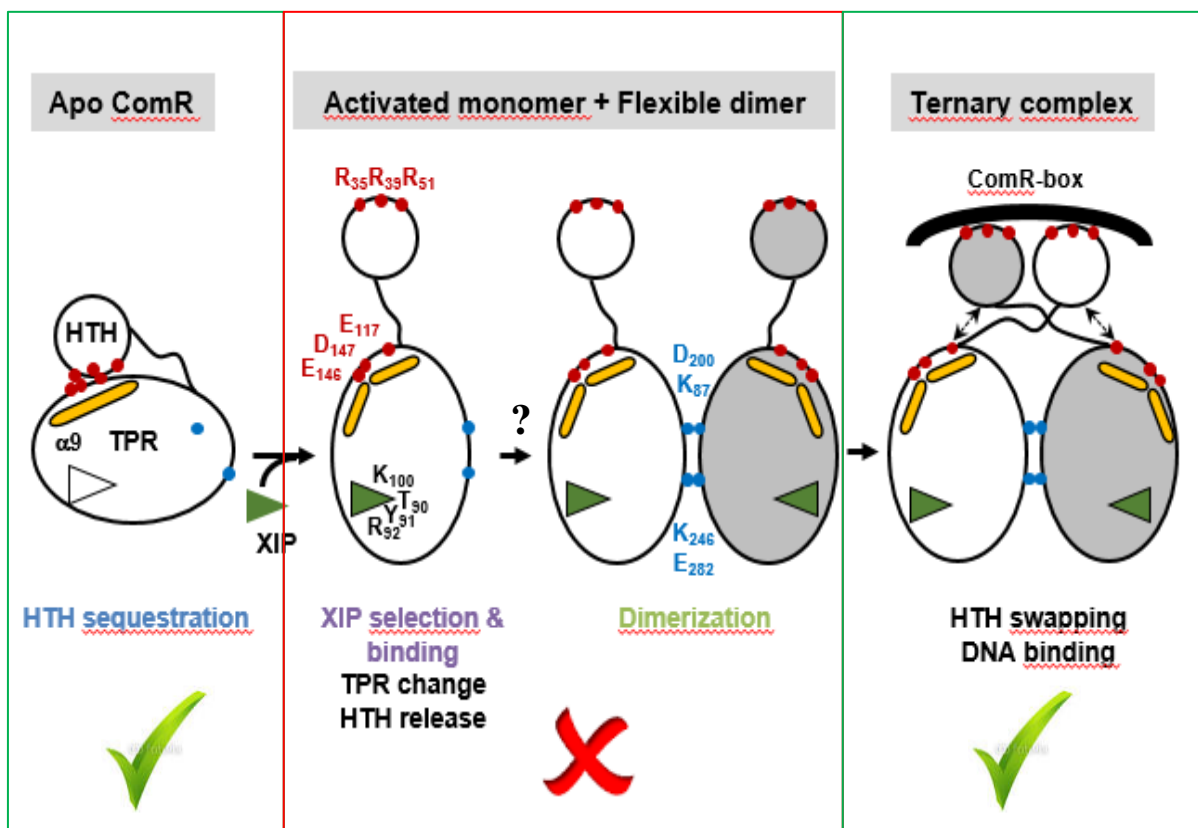


Figure 34 : Mécanisme d'activation proposé pour le système ComRS de *S. thermophilus*. L'hélice $\alpha 9$ est représentée en jaune. Les points rouges représentent les résidus impliqués dans la séquestration du HTH et les points bleus ceux impliqués dans la dimérisation des domaines TPR.

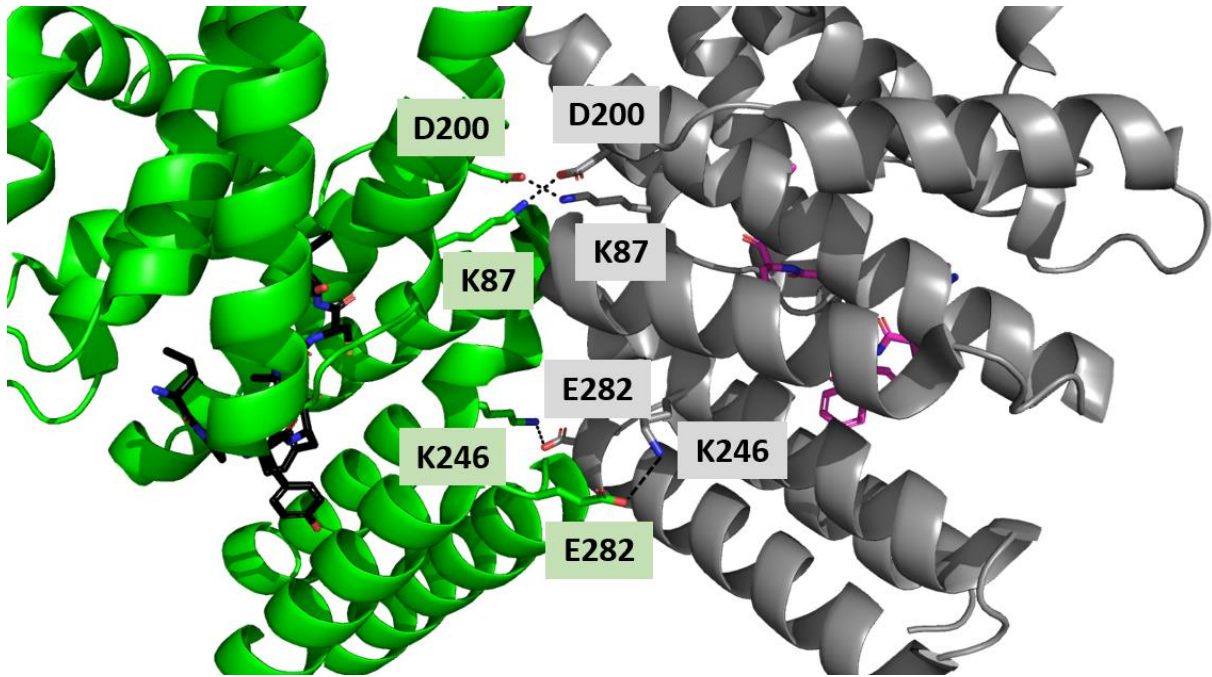


Figure 35 : Zoom sur la zone de dimérisation dans le complexe ComR_{Sth}/XIP/ADN (Code PDB : 5JUF). 4 résidus de chaque chaîne, représentés en sticks et colorés par éléments, permettent de former 4 ponts salins qui stabilisent la conformation dimérique de la protéine. Les domaines TPR de ComR sont représentés en ruban colorés par chaîne, en vert et gris. Les peptides sont représentés en stick noir et rose

Afin d'aller plus loin dans la compréhension du rôle de chaque résidu formant le peptide XIP, une étude fonctionnelle basée sur un Ala-scan de la séquence de l'octapeptide XIP_{Sth} (LPYFAGCL) de *S. thermophilus* a été réalisée. Cette dernière montre que les variants XIP-Y3A (LPAFAGCL), XIP-F4A (LPYAAGCL) et XIP-L8A (LPYFAGCA) n'activent pas la protéine ComR_{Sth} (FONTAINE *et al.* 2013) (Figure 36).

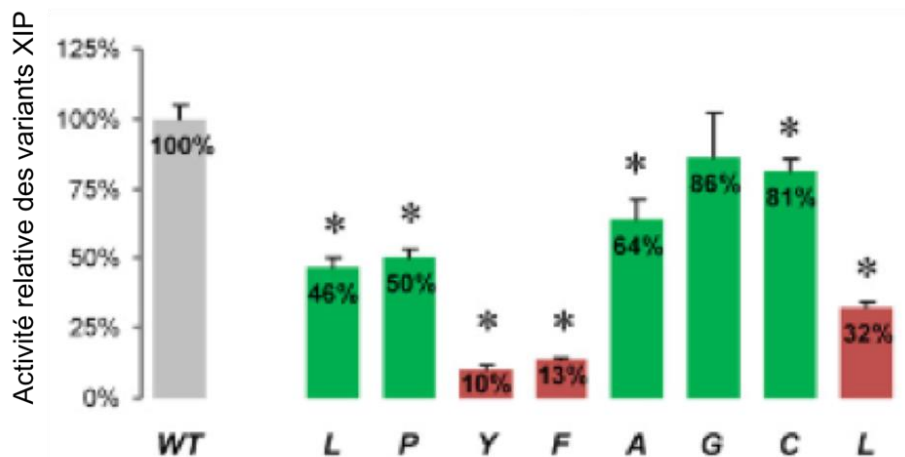


Figure 36 : Activité relative sur ComR_{Sth} de 8 variants XIP. Chacun porte la mutation en alanine d'un résidu du peptide sauvage XIP_{Sth} (LPYFAGCL) (à l'exception du résidu XIP-A5 qui a été substitué par une glycine) (FONTAINE *et al.* 2013)

Des mesures d'interaction par ITC ont ensuite montré que l'inactivité des variants F4A et L8A étaient due à une forte baisse d'affinité, avec des K_D de l'ordre du micromolaire. Par contre, le variant Y3A se fixe sur la protéine ComR_{Sth} avec un K_D du même ordre de grandeur que le peptide natif ($K_D = 32.6 \pm 5.4$ nM pour XIP_{Sth}-Y3A contre 10.0 ± 1 nM pour le peptide natif) (Figure 37) (TALAGAS *et al.* 2016). Ce résultat suggère donc que le résidu XIP_{Sth}-Y3A, joue un rôle essentiel dans le mécanisme d'activation mais pas dans la fixation du peptide.

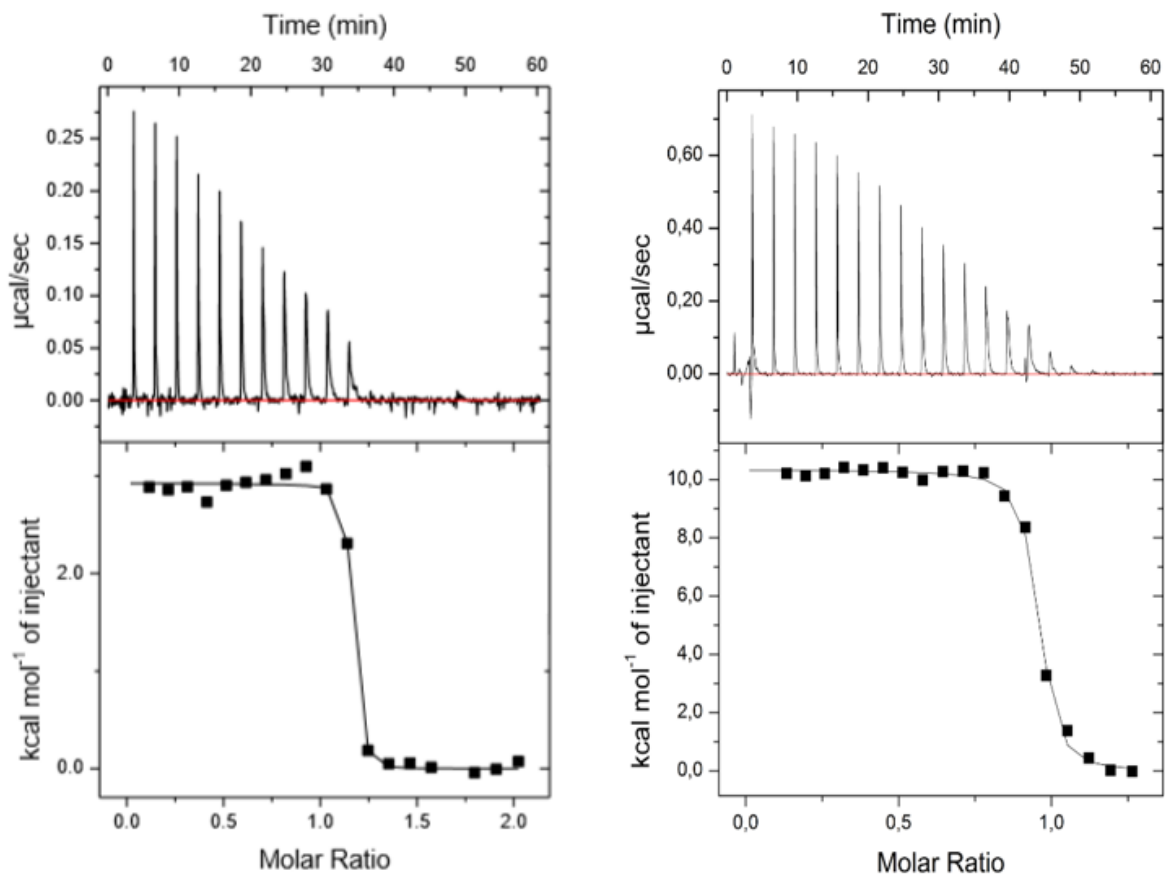


Figure 37 : Comparaison des affinités des peptides XIP_{Sth} et XIP_{Sth}-Y3A.

Isothermes des interaction ComR_{Sth} / XIP_{Sth} (LPYFAGCL, panneau de gauche) et ComR_{Sth} / XIP_{Sth}-Y3A (LPAFAGCL, panneau de droite). La protéine est à 40µM et le peptide est présent 10 fois en excès dans les deux mesures.

Enfin, une étude de cross-talk menée sur les systèmes ComRS de type II (*S. mutans*, *S. pyogenes* et *S. bovis*) et de type III (*S. suis*) a permis de mettre en évidence des différences de permissivité d'une espèce à l'autre (SHANKER *et al.* 2016) (Figure 38). Le récepteur ComR de *S. suis* est activé uniquement par son peptide XIP (WGTWVEE) tout comme *S. mutans* par son peptide natif XIP (GLDWWSL). Ces protéines ne semblent avoir aucune permissivité alors que ComR de *S. bovis*, dont le peptide natif a pour séquence (LTAWWGL) est également activé par le

peptide de *S. mutans* (GLDWWSL) et, à moindre mesure par ceux de *S. pyogenes* (SAVDWWRL) et *S. suis* (WGTWVEE), ce qui montre une forte permissivité.

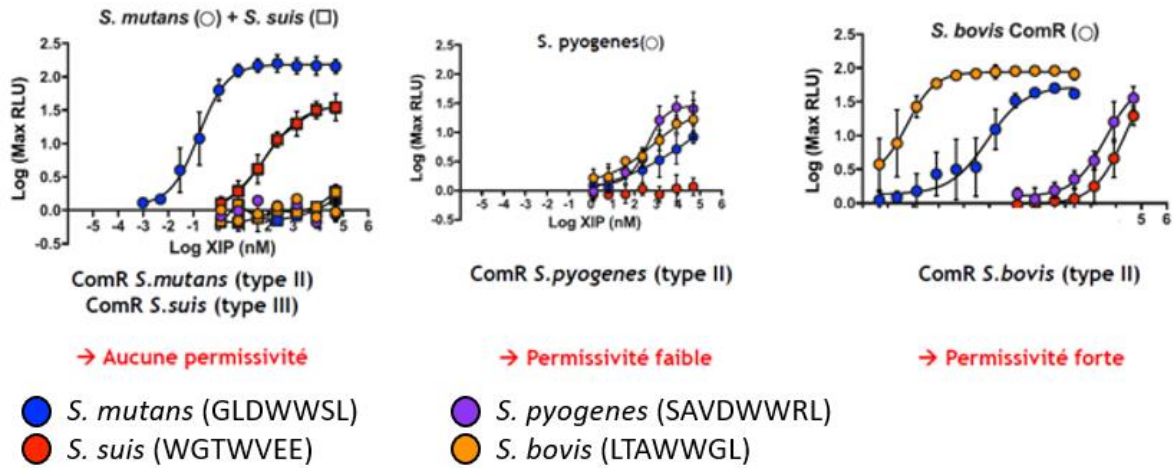


Figure 38 : Analyse de la permissivité des systèmes ComRS de type II

Profil d'activité de ComR en présence de différents peptides XIP chez 4 espèces différentes.

Un gène codant pour la luciférase a été mis sous la régulation de la protéine ComR. Des peptides XIP de différentes tailles provenant d'espèces identiques et hétérologues ont été ajoutés à des concentrations croissantes pour chaque protéine. La mesure de l'activité de luminescence pour chaque essai est normalisée sur la valeur de l'activité du peptide XIP naturel (100% d'activité).

5. But de la thèse

Afin d'apporter de nouveaux éléments à la recherche d'inhibiteurs du QS direct des bactéries à Gram positif, nous avons souhaité compléter l'étude du système ComRS qui régule la compétence chez certains streptocoques (Figure 31).

Les données structurales sur le mode de fixation du peptide ayant été obtenues avec ComR_{Sth} du groupe I (TALAGAS *et al.* 2016) (Figure 39), nous avons donc voulu utiliser ces données pour identifier les déterminants structuraux de la spécificité d'action du peptide XIP chez les récepteurs ComR de type I.

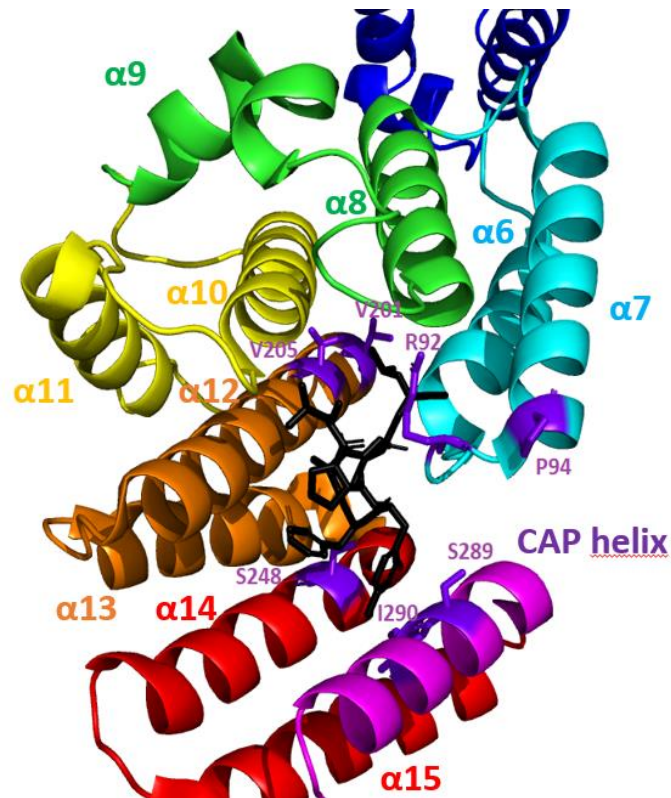


Figure 39 : Poche de fixation du peptide XIP_{Sth} sur son récepteur ComR_{Sth}.

La figure montre un zoom sur le peptide XIP, en sticks noirs, fixé au domaine TPR de ComR_{Sth} représenté en cartoon et gradient de couleurs (Code PDB : 5JUB). Les principaux résidus interagissant avec le peptide sont montrés en sticks.

La comparaison des séquences des ComR de type I montre que la protéine de *S. thermophilus* est très proche de celle de *S. salivarius* avec 92% d'identité de séquence. Les deux protéines ComR_{Sth} et ComR_{Ssa} reconnaissent d'ailleurs des peptides quasi identiques : XIP_{Sth} =

LPYFAGCL et XIP_{Ssa} = LPYFTGCL. La ressemblance est plus éloignée avec ComR_{Sve} de *S. vestibularis* qui ne présente que 84% d'identité de séquence avec ComR_{Sth} et est activé par un peptide plus éloigné, XIP_{Sve} = VPFMIYY. De manière intéressante, il a été montré que ComR_{Sth} était très faiblement activé par ce peptide XIP_{Sve} alors qu'il était plus sensible au peptide de type II de *S. mutans*, XIP_{Smu} = MGLDWWSL (FONTAINE *et al.* 2013).

La première partie de mon travail de thèse a donc consisté en l'analyse comparative des structures des régulateurs ComR de *S. thermophilus* et *S. vestibularis*. Ces travaux, associés à une étude par mutagenèse dirigée réalisée par nos collaborateurs (Equipe de Pascal Hols, LIBST, UCL, Louvain La Neuve, Belgique), font l'objet d'une publication, en reviewing à PNAS, qui est incluse dans la partie Résultats de cette thèse.

Dans un second temps, j'ai initié l'analyse structurale de deux paralogues de ComR qui régulent le couplage de la compétence et de la prédation chez *S. salivarius*. Les résultats préliminaires obtenus au cours de ma thèse suggèrent que l'un de ces paralogues semble présenter un mécanisme moléculaire différent de celui observé chez ComR. Des cristaux ayant été obtenus, la résolution de cette nouvelle structure devrait prochainement permettre d'aller encore plus loin dans la compréhension de la régulation de la compétence chez les streptocoques.

II. Résultats

1. Etude de la spécificité du système ComR/XIP

1.1. Comparaison des séquences de ComR_{Sth} et ComR_{Sve}

La comparaison des séquences de ComR chez *S. thermophilus* LMD9 et *S. vestibularis* F0396 révèle que la grande majorité des résidus ayant été identifiés comme jouant un rôle clef dans le mécanisme d'activation de ComR_{Sth} (TALAGAS *et al.* 2016) sont conservés chez ComR_{Sve} (Figure 40).

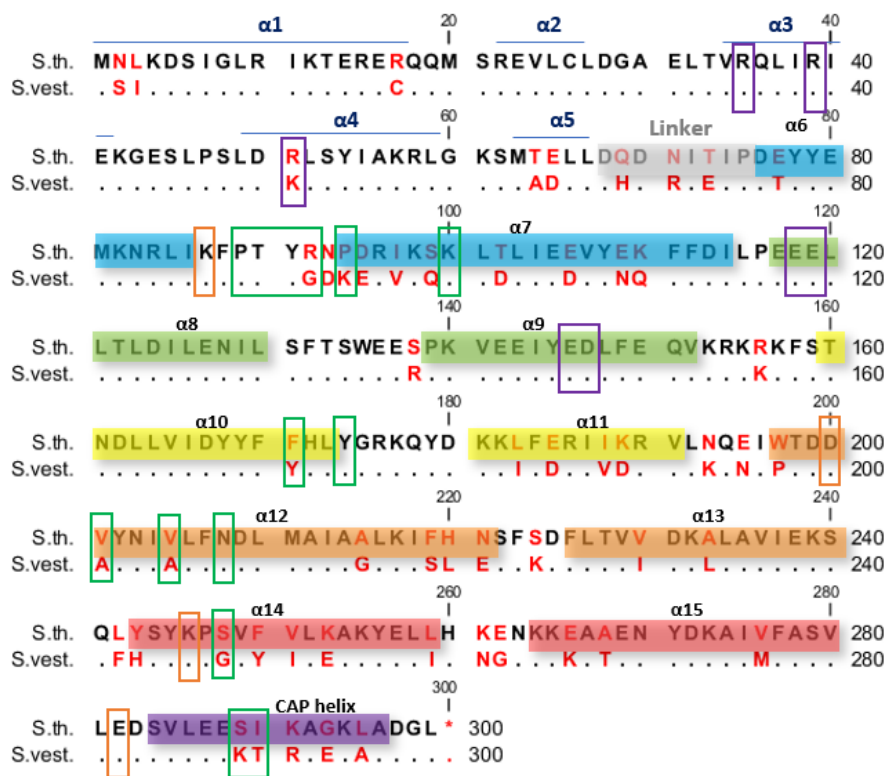


Figure 40 : Alignement de séquence entre ComR_{Sth} et ComR_{Sve}.

La séquence de ComR_{Sth} est montrée en entier avec, en rouge, les résidus non conservés chez ComR_{Sve}. La position des hélices $\alpha 1$ - $\alpha 5$ du domaine HTH sont indiquées par un trait bleu au-dessus de la séquence. Les résidus des hélices du domaine TPR sont colorés dans un gradient de couleur du cyan au violet. Seuls les résidus substitués sont indiqués dans la séquence de ComR_{Sve}. Les résidus impliqués dans la séquestration du domaine HTH sont encadrés en violet, ceux impliqués dans les ponts salins stabilisant le dimère sont encadrés en orange et enfin, ceux en interaction avec le peptide sont encadrés en vert.

Parmi les résidus importants, on observe la substitution conservatrice R51K dans l'hélice α_4 , qui ne devrait pas empêcher la formation du pont salin responsable de la séquestration du domaine HTH. Les résidus impliqués dans la formation des ponts salins stabilisant la dimérisation des domaines TPR sont quant à eux strictement conservés. Ceci suggère que le mécanisme d'activation décrit pour ComR_{Sth} est très probablement conservé chez ComR_{Sve}. L'interaction entre le carboxylate C-terminal du peptide et les résidus T90 et K100 est également conservée. Par contre, on observe, parmi les résidus interagissant avec le reste du peptide, des substitutions pouvant expliquer l'inefficacité de XIP_{Sve} sur ComR_{Sth} (FONTAINE *et al.* 2013). Il s'agit pour la plupart de résidus tapissant la poche de fixation et faisant des interactions apolaires de type van der Waals avec le peptide XIP fixé (Tableau 2).

	ComR <i>S. thermophilus</i>	ComR <i>S. vestibularis</i>
Résidus impliqués dans la séquestration du HTH	R35, R39 et R51 E117, E118, E146, D147	R35, R39 et K51 E117, E118, E146, D147
Résidus formant des ponts salins nécessaire à la dimérisation	K87 et D200 K246 et E282	K87 et D200 K246 et E282
Résidus interagissant avec XIP	P89, T90, Y91, R92, P94, K100, F171, Y174, V201, V205, N208, S248, S289, I290	P89, T90, Y91, G92, K94, K100, Y171, Y174 A201, A205, N208, G248, K289, T290

Tableau 2 : Conservation des résidus importants.

Correspondance chez ComR_{Sve} des résidus identifiés comme importants d'un point de vue fonctionnel chez ComR_{Sth}. Les substitutions sont indiquées en rouge sauf celles qui sont conservatrices et indiquées en vert.

Les substitutions aux positions 92 et 94 sont situées dans la boucle α_6 - α_7 qui, en présence du peptide, entre en interaction avec l'hélice α_{16} (CAP) pour maintenir une conformation « fermée » active du domaine TPR. De plus, chez ComR_{Sth}, R92 interagit directement avec le peptide (Figure 39) et il a été montré que la mutation R92A entraîne une diminution significative de l'affinité pour XIP_{Sth} (TALAGAS *et al.* 2016). Les autres substitutions semblent plus conservatives (ex : F171Y, V201A, V205A) ou pouvant compenser les différences de séquence entre les peptides XIP_{Sth} (LPYFAGCL) et XIP_{Sve} (VPFFMIYY).

1.2. Comparaison structurale de ComR_{Sth} et ComR_{Sve}

Afin de pouvoir comparer avec les données structurales préalablement obtenues au laboratoire sur ComR_{Sth}, j'ai entrepris la résolution structurale de ComR_{Sve}. Dans un premier temps, j'ai résolu la structure cristalline de la protéine sous sa forme apo. Les premiers essais de cristallisation, dont la gamme de concentration en protéine avait été basée sur ce qui avait été fait pour ComR *S. thermophilus*, ont permis d'obtenir une multitude de cristaux de formes différentes et de tailles convenables (Figure 41).

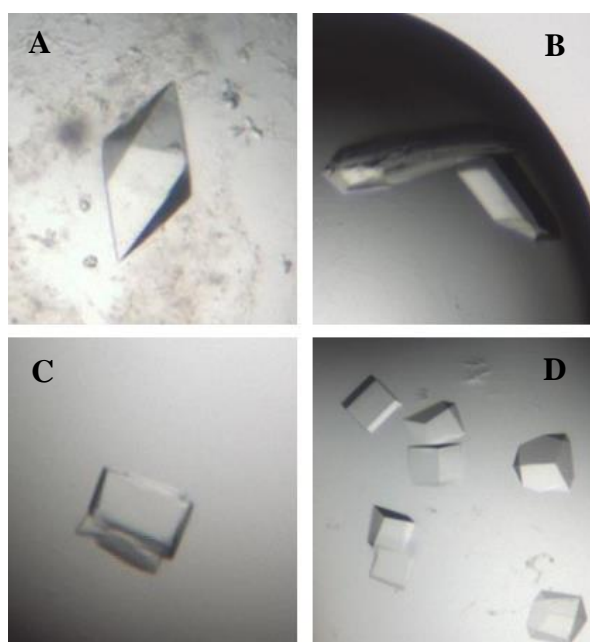


Figure 41: Différents cristaux obtenus lors des essais de cristallisation sur la forme apo de ComR_{Sve}
A- Cristal en forme de bipyramide obtenu en 0.2M Ammonium Sulfate, 0.1M bis-Tris pH5.5, 25% PEG3350 ayant diffracté à 2Å de résolution dans le groupe d'espace I4122. B- Cristal en forme de bâtonnet obtenu en 0.15M DL-Malic pH7, 20% PEG3350 ayant diffracté à 2Å de résolution dans le groupe d'espace C121. C et D- Exemples de cristaux ayant diffractés dans le groupe d'espace I4122 à une résolution moins bonne.

Parmi tous les jeux de données collectés durant diverses campagnes aux synchrotrons ESRF et SOLEIL, deux d'entre eux ont été intégrés à environ 2Å de résolution et ont pu être utilisés pour résoudre la structure de ComR_{Sve} par remplacement moléculaire à partir de la structure apo de ComR_{Sth} (Code PDB : 5JUF). Vu le fort pourcentage d'identité de séquence de plus de 80% entre les protéines ComR *S. thermophilus* et *S. vestibularis*, la séquence utilisée pour le modèle fût celle de *S. thermophilus* et elle fut changée par celle de *S. vestibularis* après le remplacement moléculaire.

Le premier jeu de données est résolu dans le groupe d'espace $I4_122$ ($a = 84\text{\AA}$; $b = 84\text{\AA}$; $c = 188\text{\AA}$; $\alpha = 90^\circ$; $\beta = 90^\circ$; $\gamma = 90^\circ$) avec 1 molécule par unité asymétrique alors que le second appartient au groupe d'espace $C2$ ($a = 110.1\text{\AA}$; $b = 38.3\text{\AA}$; $c = 87.62\text{\AA}$; $\alpha = 90^\circ$; $\beta = 113.2^\circ$; $\gamma = 90^\circ$) avec également 1 molécule par unité asymétrique. Après affinement, j'ai pu constater que, malgré la différence d'empilement cristallin, les deux structures étaient équivalentes (rmsd= 0.68 Å sur toute la séquence) et seule la structure résolue dans le groupe d'espace $C2$ a été déposée à la Protein Data Bank sous l'identifiant 6HU8.

La structure de ComR_{Sve} ainsi obtenue confirme à nouveau que la forme apo de ComR est caractérisée par la séquestration du domaine HTH par le domaine TPR, comme observé avec ComR_{Sth} (Code PDB : 5JUF) et ComR de *S. suis* (Code PDB : 5FD4).

J'ai également effectué de nombreux essais de co-cristallisation dans le but d'obtenir le complexe binaire ComR_{Sve} / XIP_{Sve} et le complexe ternaire ComR_{Sve} / XIP_{Sve} / ADN. L'ADN utilisé pour les différents essais de cristallisation est la séquence pseudo palindromique du promoteur PComX.

Dans mes essais en présence d'ADN, j'ai ainsi obtenu une petite dizaine de cristaux qui diffractaient dans le groupe d'espace $C2$ avec les mêmes paramètres de maille que la forme apo. Après avoir tenté de changer le ratio entre la protéine, le peptide et/ou l'ADN ainsi que la longueur de l'oligonucléotide utilisé (entre 18 et 22 pb) dans les échantillons, j'ai finalement réussi à obtenir, en présence d'ADN 20pb, un cristal unique (Figure 42), impossible à optimiser ni à reproduire, qui a diffracté à 3.4Å de résolution dans le groupe d'espace $P2_12_12_1$ ($a = 60.2\text{\AA}$; $b = 84.0\text{\AA}$; $c = 152.7\text{\AA}$; $\alpha = 90^\circ$; $\beta = 90^\circ$; $\gamma = 90^\circ$) avec 2 molécules par unité asymétrique.



Figure 42 : Unique cristal ayant permis de résoudre la structure du complexe binaire ComR_{Sve} / XIP_{Sve}. Le cristal a été obtenu dans 0.1M MgCl₂, 0.1M tri-Sodium Citrate pH 5.5, 12% PEG4000 et 0.1M NaCl. La goutte contenait la protéine à 10 mg/ml, le peptide XIP 10 fois en excès et l'ADN PcomX 20pb 3 fois en excès.

Un jeu de données de diffraction complet a été collecté à 3.4 Å de résolution sur la ligne PX1 du synchrotron SOLEIL à partir de cet unique cristal. Lors de la procédure de phasage par remplacement moléculaire, en utilisant la structure du complexe ternaire ComR^{Sve} / XIP^{Sve} / ADN (code PDB : 5JUB) comme modèle, n'a pas permis de trouver une solution incluant l'ADN. Il m'a également fallu supprimer les domaines HTH du modèle pour arriver à placer un dimère de domaines TPR (résidus 72 à 300). Au cours de l'affinement, j'ai réussi à reconstruire manuellement un des deux domaines HTH. Malgré l'absence de densité au niveau du linker flexible, ce domaine HTH a été assigné à la chaîne A. Au final, l'analyse de l'empilement cristallin montre qu'il n'y a pas de place pour l'ADN pourtant présent dans la goutte de cristallisation. Par contre, il y aurait la place pour le second domaine HTH mais l'analyse de la carte de densité électronique ne permet pas de le positionner (Figure 43).

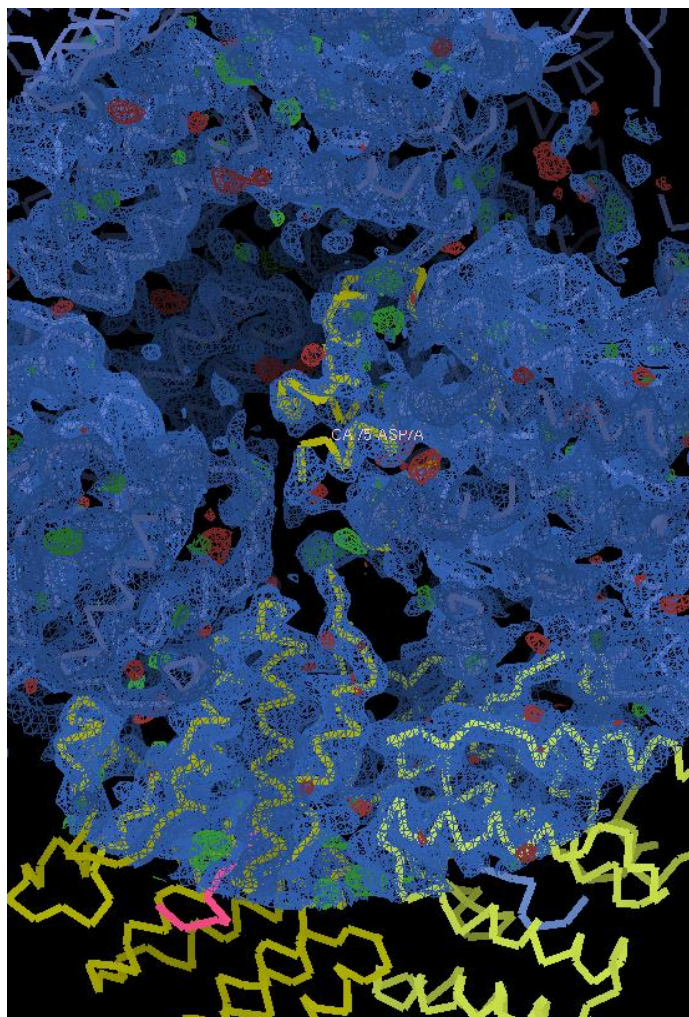


Figure 43 : Carte de densité électronique du complexe binaire ComR_{Sve} / XIP_{Sve}
La carte $2Fo-Fc$ (contourée à 1σ) est représentée en bleu et la carte $Fo-Fc$ (contourée à 3σ) est représentée en rouge et vert.

La structure du complexe binaire (Code PDB : 6HUA), qui n'avait pas été résolue dans le système de *S. thermophilus*, confirme le modèle initialement proposé suggérant que la fixation du peptide entraîne la dimérisation des domaines TPR et le relargage des domaines HTH, qui flottent librement avant de se fixer sous forme dimérique sur la séquence ADN cible.

1.3 Analyse fonctionnelle par mutagenèse dirigée

Comme indiqué plus haut, il avait déjà été montré que XIP_{Sve} n'active pas ComR_{Sth} (FONTAINE *et al.* 2013). Nos collaborateurs (Equipe de Pascal Hols, LIBST, UCL, Louvain La Neuve, Belgique) ont démontré que l'absence de cross-talk entre les systèmes de *S. thermophilus* et *S. vestibularis* est réciproque (Figure 44).

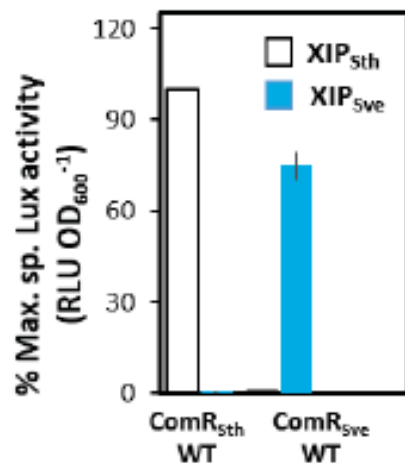


Figure 44 : Absence de cross-talk entre ComR_{Sth} et ComR_{Sve}.

Un test d'activité utilisant la luciférase comme gène rapporteur montre que les récepteurs ComR_{Sth} et ComR_{Sve} ne sont pas activés par le peptide XIP hétérologue.

Ils ont ensuite réalisé une étude par mutagenèse dirigée ciblant chez ComR_{Sth} 7 positions substituées chez ComR_{Sve} (Tableau 2). De nombreux tests d'activité réalisés sur des protéines mutantes et des peptides variants ont permis d'identifier les résidus responsables de la spécificité de reconnaissance du peptide par ces deux systèmes ComRS de type I. En particulier, cette étude a montré qu'en substituant dans ComR_{Sth} 5 positions non conservées par les résidus de ComR_{Sve}, il est possible d'inverser la spécificité de la protéine envers les peptides XIP_{Sth} et XIP_{Sve}. De plus, la seule substitution du résidu 6 de chacun des peptides XIP_{Sth} et XIP_{Sve} par le résidu observé chez l'orthologue résulte en un variant reconnu par les deux protéines ComR_{Sth} et ComR_{Sve}.

Afin de bien comprendre le mode d'activation des peptides, il est important de distinguer ceux qui se fixent sans activer de ceux qui ne se fixent pas. Les tests d'activité doivent donc être complétés par des mesures d'interaction. Pour l'étude du système ComRS de *S. thermophilus*, ces analyses avaient été réalisées par titration calorimétrique isotherme (ITC) et avaient révélé une très forte affinité entre ComR_{Sth} et son peptide XIP_{Sth}, avec un K_D de l'ordre de 10 nM (TALAGAS *et al.* 2016).

La caractérisation par ITC du système de *S. vestibularis* s'est avérée plus délicate du fait de l'obligation de reprendre le peptide XIP_{Sve}, très hydrophobe, dans 100% DMSO, ce qui complique énormément la mise en œuvre de ces mesures. Des mesures préliminaires m'ont néanmoins permis de constater que l'affinité de la protéine pour son peptide est beaucoup plus faible que pour le système de *S. thermophilus*, avec un K_D de 7 +/- 0.9 μ M (Figure 45).

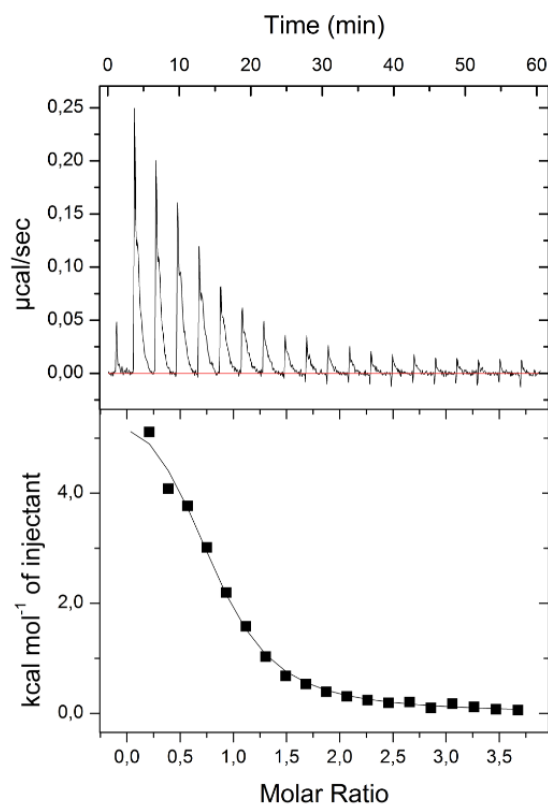


Figure 45 : Isotherme de l'interaction ComR_{Sve} / XIP_{Sve}.

L'expérience a été réalisée avec des concentrations de 50 μ M en protéine et 500 μ M en peptide, à 20°C.

Pour cette étude, les mesures d'interaction ont finalement été réalisées par nos collaborateurs à l'aide de peptides fluorescents marqués en N-terminal avec de l'isothiocyanate de fluorescéine (FITC) (Figure 46).

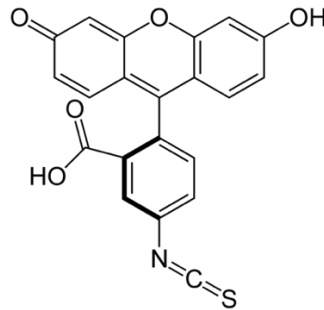


Figure 46 : Structure de l'isothiocyanate de fluorescéine (FITC)

Les mesures d'interaction par fluorescence ont finalement permis de montrer que l'affinité du couple ComR_{Sve} / XIP_{Sve} ($K_d = 120.2 \pm 22.5$ nM) est effectivement plus faible que celle du couple ComR_{Sth} / XIP_{Sth} ($K_d = 2.0 \pm 0.7$ nM) et cette technique a été utilisée pour caractériser tous les couples ComR/XIP étudiés. J'ai complété cette analyse par des mesures de calorimétrie différentielle à balayage (DSC) pour caractériser la stabilité des protéines seules et en présence de différents peptides.

1.3. Manuscrit de l'article

La comparaison de ces résultats avec les données précédemment obtenues sur le système de *S. thermophilus* (TALAGAS *et al.* 2016), nous a permis non seulement de comprendre l'impact de chaque substitution sur la sélectivité des récepteurs ComR de type I mais également de proposer un mécanisme moléculaire d'activation par le peptide très détaillé, ouvrant ainsi la voie à la mise au point de peptides inhibiteurs.

Cette étude est présentée dans le manuscrit joint.

1

2 **Main Manuscript for**

3

4 **Molecular dissection of pheromone selectivity in the competence signaling**
5 **system ComRS of streptococci**

6

7 Laura Ledesma-Garcia^{1#}, Jordhan Thuillier^{2#}, Armando Guzman-Espinola¹, Imke Ensink¹,
8 Inès Li de la Sierra-Gallay², Noureddine Lazar², Magali Aumont-Nicaise², Johann Mignolet^{1§},
9 Patrice Soumillon¹, Sylvie Nessler^{2*}, Pascal Hols^{1*}

10

11 ¹ Louvain Institute of Biomolecular Science and technology (LIBST), Biochemistry and
12 Genetics of Microorganisms, Université catholique de Louvain, Croix du Sud, 4-5 (L7.07.06),
13 B-1348 Louvain-La-Neuve, Belgium

14 ² Institute of Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université
15 Paris-Saclay, 91198 Gif-sur-Yvette cedex, France

16

17 # Co-first authors

18

19 * Co-corresponding authors: Sylvie Nessler and Pascal Hols

20 **Email:** Sylvie.Nessler@i2bc.paris-saclay.fr and Pascal.Hols@uclouvain.be

21

22 [§] Current address: Syngulon, rue du Bois Saint-Jean 15/1, B-4102 Seraing, Belgium

23

24 **Classification**

25 Biological Sciences, Biochemistry

26 **Keywords**

27 Cell-to-cell communication, quorum sensing, crosstalk, (R)RNPP, XIP pheromone.

28 **Author Contributions**

29 LLG, PS, SN, and PH conceived and designed the study. LLG, JT, AGE, IE, MAN carried out
30 the laboratory work. LLG, JT, JNL, ILSG, NL, MAN, JM, PS, SN, and PH analyzed and
31 interpreted the data. LLG, JT, JM, PS, SN, and PH wrote or revised the manuscript. All authors
32 read and approved the final manuscript.

33

34 **This PDF file includes:**

35 Main Text

36 Figures 1 to 6

37 SI Appendix

38

39 **Abstract**

40 Competence allows bacteria to internalize exogenous DNA fragments for the acquisition of
41 new phenotypes such as antibiotic resistance or virulence traits. In most streptococci,
42 competence is regulated by ComRS signaling, a system based on the mature ComS pheromone
43 (XIP), which is internalized to activate the (R)RNPP-type ComR sensor by triggering
44 dimerization and DNA binding. Cross-talk analyses demonstrated major differences of
45 selectivity between ComRS systems and raised questions concerning the mechanism of
46 pheromone-sensor recognition and co-evolution. Here, we decipher the molecular determinants
47 of selectivity of the closely related ComRS systems from *Streptococcus thermophilus* and
48 *Streptococcus vestibularis*. Despite high similarity, we show that the divergence in ComR-XIP
49 interaction does not allow reciprocal activation. We perform the structural analysis of the
50 ComRS system from *S. vestibularis*. Comparison with its ortholog from *S. thermophilus* reveals
51 an activation mechanism based on a toggle switch involving the recruitment of a key loop by
52 the XIP C-terminus. Together with a broad mutational analysis, we identify essential residues
53 directly involved in peptide binding. Notably, we generate a ComR mutant that displays a fully
54 reversed selectivity toward the heterologous pheromone with only five point mutations, as well
55 as other ComR variants featuring XIP bi-specificity and/or neo-functionalization for hybrid XIP
56 peptides. We also reveal that a single XIP mutation relaxes the strictness of ComR activation,
57 suggesting fast adaptability of molecular communication phenotypes. Overall, this study is
58 paving the way toward the rational design or directed evolution of artificial ComRS systems
59 for a range of biotechnological and biomedical applications.

60

61 **Significance statement**

62 In streptococci, competence for DNA transformation is an adaptive response contributing to
63 genome plasticity and virulence. This process is mainly controlled by the signaling system
64 ComRS, activated through an interaction between the ComS-derived pheromone XIP and the
65 cytoplasmic sensor ComR. ComRS systems evolved divergently and display different
66 preferences for cognate and non-cognate XIP peptides. In this work, we identify the molecular
67 basis of this selectivity for two similar systems that do not cross-talk to each other. Structural
68 and mutational analyses reveal five key residues involved in specificity and activation
69 mechanism, allowing us to easily manipulate ComRS towards relaxed, reverted or new
70 selectivity. This work is a milestone for understanding ComR-XIP co-evolution and pheromone
71 communication in streptococci.

72

73 **Introduction**

74 Competence for natural DNA transformation is a mechanism of horizontal gene transfer that
75 has been discovered for the first time in *Streptococcus pneumoniae* (1) but exists in many
76 bacteria (2). This transient physiological state, which only affects part of the bacterial
77 population, is induced by various environmental stimuli and allows bacteria to internalize
78 exogenous DNA fragments, which could then be integrated into the genome (3-5). Competence
79 provides a high genetic plasticity and facilitates the acquisition of new phenotypic features.
80 Concerning pathogens, competence has been shown to play a role in the evolution of escape
81 mechanisms against the host immune system or in the acquisition of antibiotic resistance genes
82 (6). Competence activation also contributes to the virulence of many streptococci by improving
83 bacterial fitness in the host through bacteriocin production and biofilm formation (7-9).
84 Competence is regulated at the population level by cell-to-cell communication systems based
85 on secreted signaling molecules (10). Phylogenetic analyses of streptococcal genome sequences
86 and functional studies revealed the existence of two paralogous competence regulating systems,
87 i.e. the well-characterized ComCDE three-component system (11-13) and the more recently
88 identified ComRS signaling system (10, 14-18). In contrary to the response regulator ComE,
89 which is indirectly activated by the ComC-derived competence-stimulating peptide through the
90 transmembrane histidine kinase receptor ComD, the sensor ComR is a cytoplasmic competence
91 regulator directly activated by the reinternalized ComS-derived peptide pheromone. ComR
92 belongs to the family of regulators called (R)RNPP for (Rgg), Rap, NprR, PlcR and PrgX (17,
93 19, 20), characterized by a C-terminal TPR-type peptide-binding domain (21) and a N-terminal
94 HTH-type DNA-binding domain (22). The ComR regulon includes the alternative sigma factor
95 SigX/ComX, which positively regulates genes required for DNA transformation. The activating
96 pheromone, a small linear peptide (7 to 11 aa) derived from the C-terminal part of the ComS
97 precursor, has thus been named XIP for *sigX/comX* Inducing Peptide (14, 16, 17, 23-25).

98 Although the ComRS system is widespread in streptococci groups, sequence analysis has
99 shown an evolutionary divergence of both transcriptional regulator and peptide, suggesting a
100 pheromone-sensor co-evolution (10). In the case of ComR, this polymorphism is more
101 pronounced in the peptide-binding receptor domain than in the conserved DNA-binding domain
102 (10). Furthermore, on the basis of observed differences in XIP sequences, three general classes
103 of ComRS systems have been proposed among the 7 phylogenetic groups of the *Streptococcus*
104 genus (10, 17). Type I peptides are characterized by the presence of a hydrophobic
105 (V/L)P(F/Y)F motif and are exclusively found in streptococci from the salivarius group.
106 Peptides from type II, which are found in the groups bovis, mutans, agalactiae and pyogenes,
107 contain a WW motif and in some cases, a basic and/or acidic residue. Because, in most *S. suis*
108 strains, the XIP peptides do not contain two contiguous aromatic residues but display a
109 WG(T/K)W motif, they have been classified as type III (10). A cross-talk analysis performed
110 on ComRS systems from type II and III showed that they could be classified as strict (e.g. *S.*
111 *mutans*, *S. suis*), intermediate (e.g. *S. agalactiae*) or promiscuous (e.g. *S. bovis*) based on their
112 ability to recognize heterologous XIP peptides (26). The sequence diversity in XIP peptides
113 and variations in ComR selectivity raise questions related to the mechanism of peptide
114 pheromone recognition and the evolutionary driver of this divergence. Interestingly, among
115 type I ComRS systems, cross-activation has been observed between ComR_{Sth} from *S.*
116 *thermophilus* LMD-9 and ComR_{Ssa} from *S. salivarius* SK126, which recognize highly similar
117 peptide sequences, i.e. XIP_{Sth} LPYFAGCL and XIP_{Ssa} LPYFTGCL, respectively (14). In turn,
118 ComR_{Sth} was not activated by the more distant type I XIP_{Sve} peptide VPPFFMIYY from *S.*
119 *vestibularis* F0396 (15). These two ComRS systems thus represent an interesting model to
120 decipher the molecular basis of XIP selectivity.

121 Our previous structure-function analysis of the ComRS system from *S. thermophilus* LMD-9
122 revealed the first clues on the activation mechanism of type I ComR (27). We showed that

123 peptide binding induces ComR dimerization and release of the HTH domain, which is
124 sequestered in the apo form of the regulator, thus preventing inappropriate activation of
125 competence in absence of peptide. To go further in the understanding of the molecular
126 mechanism of peptide selectivity by ComR regulators, we present here a detailed structural and
127 functional comparison of the ComRS systems from *S. thermophilus* and *S. vestibularis*. In
128 particular, we solved the structure of ComR_{Sve} from *S. vestibularis* F0396, alone and in complex
129 with its cognate peptide XIP_{Sve}. We engineered ComR chimeric proteins and ComR_{Sth} point
130 mutants to perform activation assays using wild-type and hybrid peptides. Together, these
131 results allowed us to identify residues directly involved in peptide selectivity and helped
132 decipher the XIP-induced conformational change required for activation.

133

134 **Results**

135

136 *Absence of crosstalk between ComR_{Sth} and ComR_{Sve}*

137 Because natural competence had not yet been demonstrated under laboratory conditions for *S.*
138 *vestibularis*, and to confirm the functionality of its ComRS system, we tested its
139 transformability in response to XIP. We showed that *S. vestibularis* F0396 is unable to be
140 transformed in absence of its cognate XIP_{Sve} peptide but becomes competent when the
141 pheromone is added, as observed in most strains of the salivarius group (**Table S1**) (28). We
142 then performed cross-talk experiments with a bioluminescent reporter strain of *S. thermophilus*
143 LMD-9 (*P_{coms-luxAB}*) that is deficient for XIP production and contains a replacement of
144 ComR_{Sth} by ComR_{Sve} from *S. vestibularis* F0396. The HTH DNA-binding domains of the two
145 proteins (helices $\alpha 1$ to $\alpha 6$) display 93% sequence identity and recognize the same
146 (pseudo)palindromic ComR-box (**Fig. 1**). As already observed with ComR_{Sth} (15), we showed
147 that ComR_{Sve} is strongly activated by addition of its cognate peptide XIP_{Sve} but not by the
148 heterologous XIP_{Sth}, (**Fig. 2**). This demonstrated the absence of cross-activation between these
149 two type I ComRS systems, despite 79% sequence identity between their TPR peptide-binding
150 domains (**Fig. 1**) (15).

151 Interaction experiments using fluorescence polarization assays demonstrated that this absence
152 of cross-talk is directly correlated with the incapacity (ComR_{Sve}) or weak ability (ComR_{Sth}) to
153 bind the XIP peptide of the other species (**Fig. 3A**). The data revealed that the affinity of ComR
154 for its cognate peptide is lower in the *S. vestibularis* system compared to *S. thermophilus* with
155 K_d values of 120.2±22.5 nM and 2.0 ±0.7 nM, respectively.

156 Differential Scanning Calorimetry (DSC) experiments showed that both ComR proteins are
157 strongly stabilized upon binding of their cognate peptide, with a melting temperature (T_m)
158 shifted from 35-40°C in absence of peptide to about 75°C in presence of peptide (**Fig. 3B**). This
159 is most probably due to the formation of a similar dimeric complex. However, the thermal

160 denaturation curve of the ComR_{Sth}·XIP_{Sth} complex showed a minor peak at a T_m of 35°C,
161 presumably reflecting an equilibrium between monomeric and dimeric forms of the protein,
162 whereas the ComR_{Sve}·XIP_{Sve} complex displayed two small additional peaks revealing the
163 presence of less stable intermediate forms. Interestingly, addition of DNA stabilized the more
164 stable ComR_{Sve}·XIP_{Sve} complex (**Fig. S1A**).

165 Electrophoretic Mobility Shift Assays (EMSA) experiments confirmed that ComR_{Sve}, like
166 ComR_{Sth}, binds to DNA only in the presence of its cognate peptide (**Fig. 3C and Fig. S1B**) and
167 forms multimeric complexes in addition to the well-defined ComRS-DNA complex, as already
168 observed for ComR_{Sth} (27).

169 Together, these results demonstrate that type I ComRS systems from *S. thermophilus* and *S.*
170 *vestibularis* behave similarly but are not cross-talking to each other.

171 *Conserved XIP-binding mode between ComR_{Sth} and ComR_{Sve}*

172 To go further in the comparison of the peptide-induced activation mechanism of the two
173 proteins, a structural analysis of ComR_{Sve} was undertaken. We solved the crystal structure of
174 the apo form of ComR_{Sve} (PDB ID 6HU8) showing that it is highly similar to apo ComR_{Sth}
175 (PDB ID 5JUF), with a rmsd of 0.8Å over 298 aligned Cα atoms. Crystal packing analysis did
176 not identify strong intermolecular contacts, thus confirming that ComR_{Sve} is monomeric in the
177 absence of bound peptide. Remarkably, the sequestering HTH-TPR interactions observed in
178 apo ComR_{Sth} (27) are conserved (**Fig. 4A**).

179 We also solved the crystal structure of ComR_{Sve} in complex with its cognate peptide XIP_{Sve}
180 (V₁PFFMIYY₈). In the refined structure of the dimeric ComR_{Sve}·XIP_{Sve} binary complex (PDB
181 ID 6HUA), only one DNA-binding domain is visible. The second HTH domain as well as the
182 linker region of both subunits are disordered, confirming the previous model of ComR
183 regulation mechanism (27). This demonstrates in particular that HTH release does not require
184 the presence of DNA and that TPR dimerization can also occur in the absence of DNA. The

185 TPR domain of ComR_{Sve}·XIP_{Sve} binary complex forms a tight dimer highly similar to that
186 observed in the ternary complex ComR_{Sth}·XIP_{Sth}·DNA (PDB ID 5JUB), with a rmsd of 0.92 Å
187 over 451 aligned Cα atoms (**Fig. 4B**). However, we can notice that the E282-K246 salt bridge
188 stabilizing the ComR_{Sth} dimer is replaced by a H-bond between E282 and Y250 in the
189 ComR_{Sve}·XIP_{Sve} complex (**Fig S2**). This may explain why less stable intermediate forms of
190 ComR_{Sve}·XIP_{Sve} were observed in the DSC experiments when no DNA stabilizes the dimer
191 upon binding of the HTH domains (**Fig. 3B and Fig. S1A**).

192 Both ComR_{Sve}·XIP_{Sve} and ComR_{Sth}·XIP_{Sth}·DNA complexes display a conserved peptide-
193 binding mode despite all sequence variations (**Fig. S3**). The XIP octapeptide is bound with its
194 C-terminal extremity deeply buried in the closed conformation of the pocket formed by the TPR
195 superhelix while its N-terminal moiety forms a single-turn 3_{10} -helix pointing toward the
196 solvent. XIP main chain atoms are superimposable, except for the first residue. In XIP_{Sve}, the
197 N-terminal V1 residue is rotated toward the solvent to accommodate the XIP A/M-5
198 substitution and allow the side-chain of XIP_{Sve}-M5 to bind in the hydrophobic pocket formed
199 by residues F/Y171 and Y-174 from helix α_{10} , where the XIP_{Sth}-L1 residue binds (**Fig. 4C**).

200 Most polar contacts are conserved, in particular the non-specific interaction between the peptide
201 backbone and ComR residue N208, corresponding to the asparagine conserved in most TPR
202 domains (29). The essential interactions between the C-terminal carboxylate of the peptide and
203 the two residues K100 from helix α_7 and T90 from the highly conserved P₈₉T₉₀Y₉₁ motif of
204 loop α_6 - α_7 (named L6) are also conserved (**Fig. 4C**) (27). Interestingly, the ComR R/G-92
205 substitution (X₁/X₂ substitution, X₁ for Sth and X₂ for Sve through the whole manuscript)
206 compensates the XIP G/I-6 substitution (**Fig. 4C**). Similarly, ComR substitutions V/A-205 in
207 helix α_{12} and S/G-248 in helix α_{14} compensate for the XIP substitutions L/Y-8 and C/Y-7,
208 respectively. Finally, the conservative XIP Y/F-3 substitution maintains hydrophobic
209 interactions with the capping helix α_{16} (named CAP), and the impact of the ComR substitutions

210 S/K-289, I/T-290 and G/E-293 remains unclear (**Fig. 4C**). Taken together, these observations
211 suggest that only few ComR substitutions ensure peptide selectivity and prevent cross-talk
212 between the ComRS systems from *S. thermophilus* and *S. vestibularis*.

213 *Reversion of ComR_{Sth} selectivity by swapping of multiple α -helices*

214 In parallel to this structural analysis, a large mutagenesis program was undertaken to understand
215 ComR-XIP specificity. As a first approach, we investigated the role of the loop L6 (residues
216 89-94) and helices $\alpha 7$ (residues 95-115), $\alpha 12$ (residues 197-221), $\alpha 14$ (residues 242-262) and
217 CAP (residues 284-296) in the peptide recognition. We engineered ComR_{Sth} chimeric proteins
218 by exchanging these TPR elements by the corresponding sequences of ComR_{Sve}. The
219 transcriptional response of these chimeras to XIP_{Sth} and XIP_{Sve} peptides was then analyzed by
220 *in vivo* bioluminescent assays (**Fig. 2A**). Single swapping of each individual element had only
221 a minor impact on the activation capacity of XIP_{Sve}. Furthermore, the chimera carrying these
222 single helix substitutions retained a complete or partial activation capacity in presence of XIP_{Sth}.
223 By contrast, a dual substitution of $\alpha 7(+L6)$ with either $\alpha 12$, $\alpha 14$ or CAP differentially resulted
224 in a change of specificity. In particular, the dual swapping of $\alpha 7(+L6)$ with $\alpha 14$ completely
225 reversed the recognition strictness towards the heterologous XIP_{Sve} peptide. The same pattern
226 was observed with the chimeric protein carrying the simultaneous substitution of the four
227 elements, which displayed a quasi-similar behavior than ComR_{Sve}. Taken together, these results
228 confirm that the TPR elements identified in the crystal structure as essential for specific XIP
229 recognition play a major role in ComR selectivity and activation.

230 *Reversion of ComR_{Sth} selectivity by point mutations*

231 In order to gain insights into the exact roles played by substituted residues either in direct
232 contact with the bound XIP or located in the vicinity of its binding site, we prepared several
233 mutant proteins carrying point substitutions in the abovementioned TPR elements. The
234 engineered ComR_{Sth} mutants R92G-P94K (L6*), V205A-V201A ($\alpha 12^*$), S248G ($\alpha 14^*$) and

235 S289K-I290T (CAP*) globally displayed similar XIP selectivity profiles compared to the
236 corresponding ComR_{Sth} chimeric proteins (**Fig. 2B**). Furthermore, the combination of all these
237 seven point mutations had the same effect as the chimeric protein carrying the simultaneous
238 swapping of all corresponding TPR elements. Intriguingly, the tetra-mutant combining R92G-
239 P94K and V201A-V205A (L6*+ α 12*) has dual specificity, contrasting with the corresponding
240 helix-swapped chimera that is not activated by XIP_{Sth}. This suggests that additional
241 substitution(s) in α 12 are required to restore the strict profile observed with the complete
242 exchange of helix α 12. However, increased selectivity toward XIP_{Sve} was partially restored
243 upon addition of the fifth α 14* mutation S248G compensating the C/Y-7 XIP substitution, and
244 fully recovered upon further addition of the CAP* mutations S289K-I290T, which shows that
245 they play a major role in the recruitment of the CAP helix by the XIP_{Sve} residue F3. In addition,
246 we verified using EMSA experiments that the multiple mutants of ComR_{Sth}, which are activated
247 by XIP_{Sve}, can bind DNA in presence of this heterologous peptide (**Fig. 3C and Fig. S1B**). By
248 careful examination of the structure, residues 94 and 201 were identified as less essential in the
249 protein-peptide interaction. Indeed, reversion of these two mutations had no impact on the
250 inversion of peptide selectivity. Hence, we showed that a ComR_{Sth} mutant protein carrying only
251 five substitutions (i.e. R92G-V205A-S248G-S289K-I290T) displayed a full inversion of
252 peptide selectivity (**Fig. 2B**). This emphasizes the fact that stabilization of the active closed
253 conformation of the TPR domain requires not only the recruitment of loop L6 through
254 interactions between residues XIP-6 and ComR-92 but also recruitment of CAP through
255 interactions of the peptide with helices α 12 and α 14.

256 *Tight interconnection between XIP affinity and activation effect*

257 In order to compare the involvement of the abovementioned residues in *in vitro* peptide binding
258 with *in vivo* transcriptional activation, the affinity of ComR_{Sth} point mutants for XIP_{Sth} and
259 XIP_{Sve} was measured by fluorescence polarization (**Fig. 3A**). Comparison with the fluorescence

260 data obtained with the wt protein demonstrated that the R92G-P94K (L6*) mutations alone
261 decreased the affinity of ComR_{Sth} for its cognate peptide and slightly improved the interaction
262 with XIP_{Sve}, whereas V205A-V201A (α 12*) or S289K-I290T (CAP*) mutations did not
263 significantly affect the peptide affinity of ComR_{Sth} (**Fig. S1C**). Combination of L6* and α 12*
264 mutations produced a promiscuous receptor with similar and low affinity for both peptides,
265 mainly due to L6* mutations. Surprisingly, the S248G (α 14*) mutation showed an essential
266 role for XIP_{Sve} binding, with an estimated Kd value of 246.4 \pm 136.9 nM, similar to the Kd of
267 the natural receptor ComR_{Sve} (120.2 \pm 22.5 nM). This result contrasts with the absence of *in vivo*
268 activation observed with the α 14* mutation in presence of XIP_{Sve} (**Fig. 2B**), which could be
269 explained by a lack of activating conformational change even if the pocket of this mutant
270 protein accommodates the non-cognate peptide. Interestingly, the α 14* mutation only weakly
271 affected the affinity for XIP_{Sth} (Kd = 8.2 \pm 1.4 nM) and required its combination with L6* and
272 α 12* mutations to achieve inversion of selectivity, with a Kd value of 66.2 \pm 23.0 nM for XIP_{Sve}.
273 Further addition of the CAP* mutations had almost no effect on the affinity for XIP_{Sve} (Kd
274 value of 70.3 \pm 42.4 nM) (**Fig. S1C**), suggesting that they do not play an essential role in peptide
275 binding but might be important for the stabilization of the closed conformation of the TPR
276 domain. These results nicely correlate with the *in vivo* activity assays (**Fig. 2A and 2B**) and
277 further emphasize the role of the R/G-92 and S/G248 ComR substitutions compensating the
278 G/I-6 and C/Y-7 XIP substitutions. Taken together, they show that affinity and activating effect
279 of the peptide are tightly interconnected with selectivity.

280 *Compensating mutations between XIP and ComR*

281 In order to validate the essential role played by the abovementioned compensating substitutions,
282 we evaluated the impact of various XIP variants on ComR selectivity. In a first step, we
283 investigated the compensating substitution XIP G/I-6 versus ComR R/G-92. For this purpose,
284 we tested the activity of the hybrid variants XIP_{Sth}-v1 (LPYFAICL) and XIP_{Sve}-v1

285 (VPFFMGYY) on both wt proteins and on the ComR_{Sth} point mutants (**Fig. 2C**), as well as the
286 chimera with swapped helices (**Fig. S4A**). Notably, both wt proteins ComR_{Sth} and ComR_{Sve}
287 were efficiently activated by these two XIP variants, showing that a single XIP substitution at
288 position 6 could result in cross-activating peptides. In particular, whereas XIP_{Sve} was totally
289 inefficient with ComR_{Sth}, introducing a G residue at position 6 is sufficient for proper binding
290 of the peptide variant XIP_{Sve-v1} (VPFFMGYY) and stabilization of the active closed
291 conformation of the protein. However, activation of ComR_{Sth} by XIP_{Sth-v1} (LPYFAICL)
292 demonstrates that the I6 residue impairing the binding of XIP_{Sve} has not the same effect when
293 introduced in XIP_{Sth}. This suggests that the steric clash between XIP_{Sve}-I6 and ComR_{Sth}-R92
294 that was expected from the structure analysis (**Fig. 4C**) can be avoided in the context of the
295 XIP_{Sth} sequence. Reciprocally, the requirement of XIP-I6 for ComR_{Sve} activation is only needed
296 in the XIP_{Sth} context, further emphasizing the complexity of the peptide recognition
297 mechanism. Finally, the ComR_{Sth} R92G-P94K (L6*) mutant was efficiently activated by the
298 XIP_{Sth-v1} (LPYFAICL) containing the compensating XIP G/I-6 substitution but was only
299 poorly or not activated by the peptides containing a G residue at position 6, i.e. XIP_{Sth}
300 (LPYFAGCL) and XIP_{Sve-v1} (VPFFMGYY). Moreover, the swap of helix $\alpha 7(+L6)$ had a
301 dominant-negative effect on XIP_{Sve-v1} activation when combined with swapped $\alpha 12$, $\alpha 14$ or
302 CAP, reinforcing the importance of this complementarity for activation (**Fig. S4A**).

303 Together, these results thus confirm the importance of the compensation between XIP G/I-6
304 and ComR R/G-92 for activation. They also support the key role of residue XIP-6 in controlling
305 the selectivity of peptide recognition.

306 In a second step, we addressed the importance of compensating substitutions XIP L/Y-8 and
307 C/Y-7 versus ComR V/A-205 and S/G-248 in peptide selectivity by introducing the double
308 substitution L/Y-8 and C/Y-7 in XIP_{Sth} (XIP_{Sth-v2} LPYFAGYY), which is characteristic of the
309 C-terminus of XIP_{Sve} (**Fig. 2C**). The modified peptide XIP_{Sth-v2} impaired activation of both

310 ComR_{Sth} and ComR_{Sve}. However, activation of ComR_{Sth} by XIP_{Sth-v2} was partially restored by
311 the mutations V201A-V205A (α 12*) or S248G (α 14*), supporting the complementarity
312 between compensating substitutions. Intriguingly, both wt ComR are activated by XIP_{Sve-v1}
313 (VPFFMGYY) but not by XIP_{Sth-v2} (LPYFAGYY), suggesting that interactions involving the
314 N-terminus of the bound peptide (i.e. residues LPYFA versus VPFFM) could also play an
315 important role in the activating conformational change.

316 DSC experiments (**Fig. S4B**) showed that binding of XIP_{Sth-v1} had the same stabilizing effect
317 than the wt peptide on ComR_{Sth} but only resulted in intermediate forms of the complex with
318 ComR_{Sve}. In turn, none of the two proteins were shifted to the highly stable form upon binding
319 of XIP_{Sve-v1}, which only resulted in less stable complexes. Finally, the scans obtained in
320 presence of the inactive XIP_{Sth-v2} variant were highly similar to the apo form of ComR_{Sth} and
321 ComR_{Sve}. Interestingly, EMSA experiments demonstrated that the less stable complexes
322 formed by ComR_{Sth} and ComR_{Sve} with either XIP_{Sth-v1} or XIP_{Sve-v1} shifted DNA in a similar
323 manner than the cognate complexes (**Fig. S4C**). These results thus demonstrate that non-
324 cognate peptides can bind to ComR regulators and form alternative active complexes.

325 Finally, we can conclude that these results allowed us to identify key XIP and ComR residues
326 directly implicated in the activating conformational change and the main compensatory
327 substitutions that lead to divergence in peptide selectivity between the two type I ComRS
328 systems.

329 *Integrative view between ComR selectivity and activation*

330 Resolution of a new crystal structure of ComR_{Sth} apo form (PDB ID 6QER) demonstrated that
331 the superhelix of the peptide-free TPR domain can adopt distinct conformations (**Fig. S5A and**
332 **S5B**), thus demonstrating the structural plasticity of the monomeric open conformation of
333 ComR. Modelling analysis showed that, in both ComR_{Sve} and ComR_{Sth} proteins, the closed
334 conformation of the peptide-binding pocket, which is not accessible to the peptide from outside

335 of the protein (**Fig. 5A**), can however easily accommodate all non-cognate (active and inactive)
336 XIP peptides used in this study. In particular, a simple shift of the R92 side chain allows
337 modelling of XIP_{Sve} in the pocket of ComR_{Sth} without any clash with XIP-I6 (**Fig. S6**).
338 Together, these results suggest that the pheromone first binds in the flexible open conformation
339 of the TPR pocket and that the activation mechanism mostly relies on the ability of the peptide
340 to stabilize the closed conformation of the TPR domain.

341 At the bottom of the open pocket, helices $\alpha 14$ and $\alpha 12$ most probably form the first binding
342 scaffold for the peptide (**Fig. 4C**). This would explain the essential role in peptide selectivity
343 played by residues 205 (from $\alpha 12$) and 248 (from $\alpha 14$), which accommodate the XIP
344 substitutions L/Y-8 and C/Y-7, respectively. Formation of an essential salt bridge between the
345 C-terminal carboxyl group of the peptide and the conserved $\alpha 7$ residue K100 (27) insures the
346 recruitment of loop L6 *via* an additional H-bond with residue T90 from the conserved P₈₉T₉₀Y₉₁
347 motif. Proper positioning of the carboxylate requires the binding of residue XIP-8 in the
348 hydrophobic pocket formed by residues from helices $\alpha 6$, $\alpha 8$, $\alpha 10$ and $\alpha 12$ (**Fig. 4C and Fig.**
349 **5B**). Our mutational analysis further highlighted the essential role played by the compensating
350 XIP G/I-6 and ComR R/G-92 substitutions in loop L6 recruitment and peptide selectivity.

351 Comparison of the structures showed that, in both ComRS systems, the recruitment of loop L6
352 results in a small shift of helix $\alpha 7$, and in particular of its conserved residues Y91 and R96 (**Fig.**
353 **5C**). At the other end of the TPR domain, the CAP helix is recruited by a hydrophobic
354 interaction between residues 289-290 and XIP-3 and shifted toward $\alpha 7$, allowing the formation
355 of (i) a salt bridge between residues R96 and D283 and (ii) a hydrophobic stacking interaction
356 between Y91 and L286, thus ensuring stabilization of the closed conformation (**Fig. 5C**). These
357 movements further allow stabilization of the TPR dimer, in particular through reorientation of
358 the conserved residue E282 from the loop preceding the CAP helix, which is directly involved
359 in the interaction with the neighboring subunit (**Fig. S2**). In addition, they also result in the

360 release of the sequestered HTH domains. In particular, we observed in both ComRS systems
361 that the recruitment of loop L6 is associated with a shift of helix $\alpha 6$. The latter contains residue
362 K82, which forms a conserved salt bridge with E119 from $\alpha 8$ in the absence of bound peptide
363 (**Fig. 5D**). This strong interaction allows proper positioning of $\alpha 8$ and formation of the
364 sequestering interaction between E117 and R39 from helix $\alpha 3$ (27). Upon peptide binding, the
365 break of the K82-E119 interaction might play an essential role in the disruption of the
366 interaction between HTH and TPR domains.

367 The proposed mechanism thus explains how peptide binding induces the break of helix $\alpha 9$
368 shown to be essential for the release of the sequestered HTH domain and the dimerization
369 required for DNA binding.

370

371 **Discussion**

372 This structural and functional comparison between two type-I ComRS systems that do not
373 cross-talk to each other shed light on the molecular determinants responsible of pheromone
374 selectivity. In addition, modified sensors and pheromones were generated that allowed
375 communication between these two signaling systems due to either high promiscuity or
376 exchange in specificity. This work also highlighted the role of key residues in ComR-XIP
377 complementarity and helped gain a better understanding of the peptide-induced activation
378 mechanism of ComRS systems.

379 Two crystal structures of ComR_{Sve} were solved in this study, its apo form and the binary
380 complex with XIP_{Sve}. They demonstrate that the overall activation mechanism, initially
381 proposed for the protein from *S. thermophilus* (27), is conserved in *S. vestibularis* (**Fig. 6**). In
382 particular, we confirmed our previous hypothesis suggesting that, in the absence of bound
383 DNA, the ComR·XIP complex dimerizes while the sequestered HTH domains are released and
384 freely float around the TPR dimer through highly flexible linkers. Comparison of the
385 ComR_{Sve}·XIP_{Sve} binary complex with the previously published ternary complex
386 ComR_{Sth}·XIP_{Sth}·DNA (27), together with an extensive mutagenesis analysis based on the
387 transfer of ComR_{Sve} features in ComR_{Sth} mutants, allowed us to go further in the molecular
388 understanding of the specific peptide-induced activation of type I ComR regulators (**Fig. 5**).
389 This mechanism relies on an initial pheromone selection by the TPR domain in an open
390 conformation, in which helix $\alpha 14$ plays a predominant role for peptide discrimination. Then,
391 we proposed that the recruitment of the first TPR motif ($\alpha 6$ -L6- $\alpha 7$) by the C-terminus of the
392 XIP peptide acts as a “toggle switch” for the conformational change of the whole TPR domain
393 from an open to a closed conformation (**Fig. 6**). We suggest that this shift allows (i) the
394 stabilization of the CAP helix (*via* $\alpha 7$ -CAP and XIP-CAP interactions) in a configuration that
395 induces dimerization and (ii) the reorientation of key residues resulting in HTH release (*via* an

396 $\alpha 6$ - $\alpha 8$ relay and a break of $\alpha 9$) (**Fig. 6**). The details of the molecular model proposed for the
397 peptide-induced activation of ComR have never been reached for any other sensor of the
398 (R)RNPP family.

399 A previous modelling and mutagenesis analysis performed on type II and type III ComR
400 proteins by using the crystal structure of apo ComR_{Ssu} from *S. suis* (PDB ID 5FD4) as a guide
401 showed that the overall architecture as well as HTH sequestration are conserved in this type III
402 regulator. This study also suggested that a conserved face of the peptide-binding pocket could
403 contribute to XIP binding whereas a more variable face could be involved in pheromone
404 selectivity (26). Finally, type II ComR were classified according to their permissiveness in
405 peptide recognition. Our functional analysis allowed to go further in the understanding of
406 ComR selectivity.

407 In particular, we showed that a single mutation at position 6 in either of the two XIP sequences
408 induced cross-talk between these two cell-to-cell communication systems (**Fig. 2C**). These
409 results demonstrate that XIP has a high intrinsic genetic plasticity to evolve from a strict
410 pheromone toward a promiscuous inducing peptide for type I ComR sensors. Reciprocally, we
411 also generated the XIP_{Sth-v2} variant (LPYFAGYY), which cannot activate wt ComR_{Sth} and
412 ComR_{Sve}, but instead activates the ComR_{Sth} mutants $\alpha 12^*$ and $\alpha 14^*$ (**Fig. 2C**), therefore
413 highlighting that the mutated residues compensate the XIP C/Y-7 and L/Y-8 substitutions.
414 Globally, we demonstrated that five amino acid substitutions are mainly responsible for the
415 distinct peptide specificity of the two closely related XIP sensors. This conversion of ComR_{Sth}
416 is associated with progressive trade-offs between initial and new function, whereas intermediate
417 mutants (e.g. L6*+ $\alpha 12^*$) can be described as generalist receptors that are responsive to both
418 XIPs (**Fig. 2B**). In nature, a progressive co-evolution of the pheromone/receptor couple should
419 be the real scenario and it is not clear whether the evolutionary trajectory would involve more
420 or less generalist systems. However, such generalist receptors were previously reported for type

421 II ComRS systems of the bovis group of streptococci (26). The receptor and pheromone alleles
422 might co-evolve sequentially in an “evolutionary ratchet”, dictated either by a mutation in XIP
423 or ComR. Based on our results showing fast XIP evolvability, one possible scenario would be
424 an initial single mutation in the XIP sequence. This modification in the signaling peptide could
425 only slightly reduce its activation capacity (e.g. XIP_{Sth-v1} and XIP_{Sve-v1}), and be followed by
426 the selection of modifications in the receptor in order to optimize protein interactions (e.g. XIP-
427 6 and ComR-92 positions). It is interesting to note that sequence alignment of type I ComR
428 proteins revealed a higher mutation rate in α -helices that are in close contact with the XIP
429 peptide (i.e. α 7(+L6), α 14 and CAP) as well as a strong conservation in each species of the
430 substituted residues involved in XIP selectivity (**Fig. S7**). Although the advantages of highly
431 selective versus promiscuous ComRS systems in complex microbial populations are not clearly
432 understood yet (26), our results indicate that the ComRS system is easily evolvable into each
433 direction. The capacity to easily modulate XIP specificity will help us to design competition
434 experiments to address the biological relevance of this question.

435 Finally, the dissection of the molecular regulatory mechanism of ComR and understanding of
436 the structural determinants of its peptide pheromone selectivity exposed here also paves the
437 way toward the rational design and/or directed evolution of artificial transcriptional control and
438 chemical communication based on ComRS systems with a large potential of applications in the
439 fields of synthetic biology and pathogen biocontrol (e.g. quorum enhancers/quenchers).

440

441 **Materials and Methods**

442 *Bacterial strains, plasmids, oligonucleotides and growth conditions*

443 See SI Materials and Methods. Bacterial strains, plasmids and oligonucleotides used in this
444 study are listed in **Tables S2-S4**.

445

446 *Transformation assays, DNA techniques, genetic constructions, protein purification, and in*
447 *silico analyses*

448 See SI Materials and Methods.

449

450 *XIP peptide preparation*

451 Synthetic XIP octapeptides and FITC N-labeled nonapeptides were supplied by Peptide 2.0
452 (Chantilly, VA, USA) or GeneScript (Piscataway, NJ, USA), and resuspended in 100% DMSO
453 [v/v].

454

455 *Luciferase assays*

456 Luciferase assays were performed as previously described (15, 30). Growth (OD₆₀₀) and
457 luciferase (Lux) activity (expressed in relative light unit; RLU) of the cultures were monitored
458 after addition of synthetic XIP peptide (1 μM) during 24 h in a multi-wells plate reader (Hidex
459 Sense, Hidex, Turku, Finland).

460

461 *Electrophoretic mobility shifts assays*

462 A fixed concentration of purified ComR protein was mixed with a 40-bp dsDNA fragment (20
463 ng) carrying the ComR box of P_{comS} coupled to the Cy3 fluorophore in absence or presence of
464 increasing amounts of XIP peptide as previously described (15). ComR_{Sth}-L6*+α12*+α14* (1
465 μM) and ComR_{Sth}-L6*+α12*+α14*+CAP* (1 μM) were mixed with 2-fold serial dilutions of

466 XIP at an initial concentration of 2 μ M. ComR_{Sth}-L6*+ α 12* (4 μ M) was mixed with 2-fold
467 serial dilutions of XIP at an initial concentration of 8 μ M. Increased concentrations of ComR_{Sth}-
468 L6*+ α 12* and XIP were used to better visualize the complexes as this mutant has a lower
469 affinity for both wt XIPs. ComR_{Sth}·XIP_{Sth} and ComR_{Sve}·XIP_{Sve} WT controls were respectively
470 assayed in the two above-cited conditions. Negative controls were performed with 1 μ M ComR
471 in absence or presence of 2 μ M XIP. Peptide variants were tested at a fixed concentration of 1
472 μ M ComR WT. The samples (20 μ l) prepared in binding buffer (20 mM Tris-HCl, pH 8.0, 150
473 mM NaCl, 1 mM EDTA, 10% glycerol [v/v], 1 mg.ml⁻¹ BSA) were incubated at 37°C for 10
474 min prior analysis on a native 4-20% gradient gel (iD PAGE Gel; Eurogentec) in Tris-MOPS
475 buffer (50 mM Tris-base, 50 mM MOPS, 1 mM EDTA, pH 7.7). DNA complexes were detected
476 by fluorescence on the Ettan DIGE Imager with bandpass excitation and emission filters of
477 540/25 and 595/25 nm, respectively (GE Healthcare, Waukesha, WI).

478

479 *Fluorescence polarization assays*

480 Increasing concentrations of purified ComR were mixed with a FITC N-labelled version (9 aa)
481 of XIP_{Sth} or XIP_{Sve} peptides (ILPYFAGCL and AVPFMIYY, respectively) at a fixed
482 concentration of 30 nM. The samples were prepared in a final volume of 100 μ l in binding
483 buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % glycerol [v/v]) and
484 incubated for 10 min at 30°C in black well 96-well plates (Greiner, Alphen a/d Rijn, The
485 Netherlands). Anisotropic measurements were performed in a multi-wells plate reader (Hidex
486 Sense, Hidex, Turku, Finland) in polarization mode with 485/10 nm and 535/20 nm excitation
487 and emission filters, respectively. In order to calculate K_d values, curves that reached saturation
488 were fitted by the one ligand binding site equation using the SigmaPlot software.

489

490

491 *Differential Scanning Calorimetry*

492 Purified ComR samples (15 μM) were analyzed alone and in presence of XIP peptides (150
493 μM) in the protein SEC buffer in presence of 2% DMSO [v/v] by using a MicroCal PEAQ-
494 DSC Automated system (Malvern). Samples of 350 μl were heated from 20°C to 90°C using a
495 1°C/min rate. The thermal behavior of samples was recorded and analyzed using the MicroCal
496 PEAQ-DSC software and the thermal transition midpoint (T_m) representing the stability of the
497 sample was determined at to the top of the DSC peak. The effect of DNA was tested by addition
498 of a 20-bp dsDNA fragment carrying the ComR box of P_{comX} (45 μM). We also verified that
499 the peptide alone did not give any signal.

500

501 *Crystal structure determination*

502 Crystallization trials were performed at 18°C using a Cartesian robot and commercial kits.
503 Initial hits were reproduced and optimized manually using the hanging drop method and
504 homemade solutions. The apo form of ComR_{Sve} crystallized at a concentration of 20 $\text{mg}\cdot\text{ml}^{-1}$ in
505 20% PEG 3350 [w/v], 150 mM DL-malic acid, pH 7.0. The complex ComR_{Sve}·XIP_{Sve}
506 crystallized in 12% PEG 4000 [w/v], 100 mM MgCl₂, 100 mM NaCl, 100 mM tri-sodium
507 citrate, pH 5,5, at a protein concentration of 6.6 $\text{mg}\cdot\text{ml}^{-1}$ in presence of 1.5 mM peptide. Crystals
508 were flash frozen in the crystallization solution supplemented with 25-30% glycerol [v/v] and
509 conserved in liquid nitrogen prior X-ray diffraction assays.

510 Diffraction data were collected at the French synchrotron SOLEIL on beamline Proxima-2 for
511 the apo form of ComR_{Sve} and on beamline Proxima-1 for the ComR_{Sve}·XIP_{Sve} complex. The
512 crystals diffracted at 1.9 Å and 3.4 Å resolution, respectively. The data were processed using
513 XDS (31) and the crystallographic phase problem was solved by molecular replacement using
514 PHASER (32). The structure of apo ComR_{Sth} (PDB ID 5JUF) and the TPR domain from the
515 ComR_{Sth}·XIP_{Sth}·DNA crystal structure (PDB ID 5JUB) were used as initial models for the apo

516 form of ComR_{Sve} and its complex with XIP_{Sve}, respectively. The resulting models of ComR_{Sve}
517 (PDB ID 6HU8) and ComR_{Sve}·XIP_{Sve} (PDB ID 6HUA) were refined using PHENIX wizard
518 (33) and manually optimized using COOT (34). The statistics of the diffraction data sets and of
519 the refined models are given in **Table S5**.

520 Crystal packing analysis and structure comparison were performed using PISA (35) and
521 PDBeFold (36), respectively. Peptide-protein interactions were analyzed and illustrated by
522 LIGPLOT+ (37). The figures were prepared by using the graphic software PyMol (38).

523

524 **Conflict of interest statement**

525 The authors declare that the submitted work was not carried out in the presence of any personal,
526 professional or financial relationships that could potentially be construed as a conflict of
527 interest.

528

529 **Funding**

530 This work was supported by the Belgian National Fund for Scientific Research (FNRS), the
531 Interuniversity Attraction Poles (IAP) Programme of the Belgian Science Policy Office
532 (BELSPO), and the Concerted Research Actions (ARC) from Federation Wallonia-Brussels.
533 JT was supported by a PhD fellowship from the French Ministère de l'Enseignement supérieur,
534 de la Recherche et de l'Innovation. PH is Senior Research Associate at FNRS.

535

536 **Acknowledgements**

537 The crystallization assays and the calorimetry experiments were performed on the IBiSA
538 (Biology Health and Agronomy Infrastructures) labeled platforms from the Institute for
539 Integrative Biology of the Cell (I2BC) (<http://www.i2bc.paris-saclay.fr/spip.php?article281>),
540 supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-

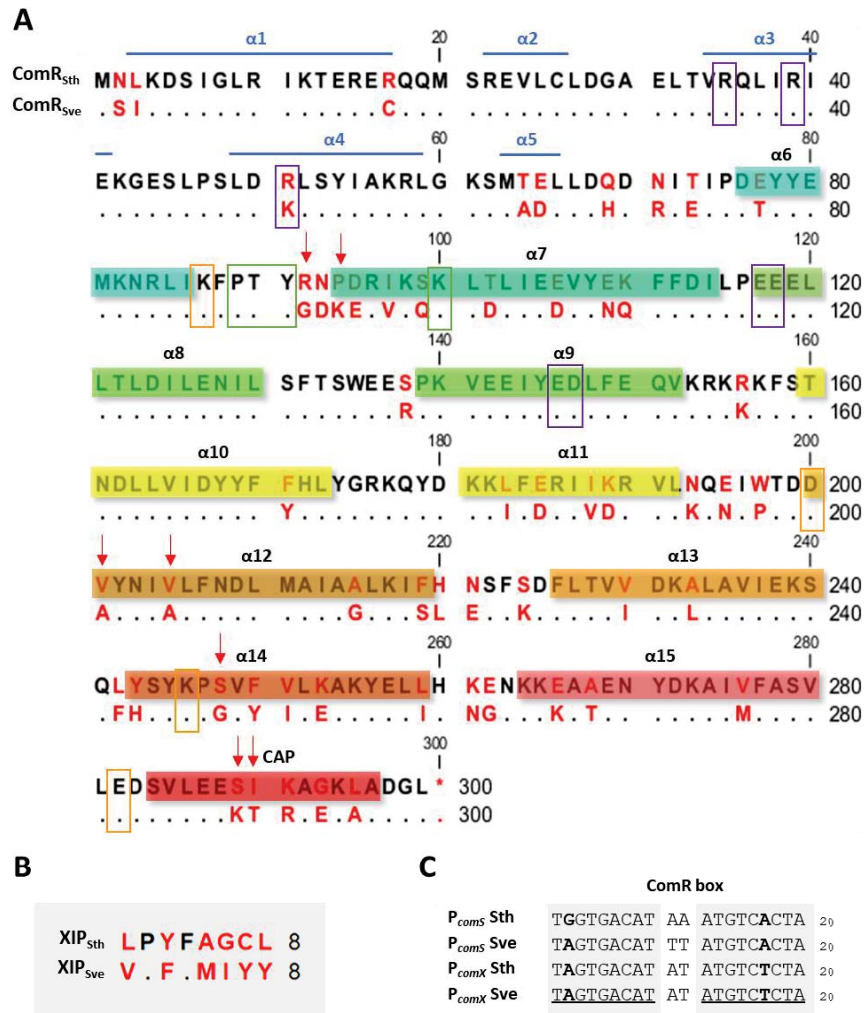
541 INBS-05. The X-ray diffraction experiments were performed on beamlines Proxima-1 and
542 Proxima-2 at SOLEIL Synchrotron, France, as well as on beamlines ID30b, 1D29 and ID23 at
543 the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to the
544 local contacts at both sites for providing assistance as well as to their synchrotron staff for
545 smoothly running the facility.
546

547 **References**

- 548 1. Griffith F (1928) The Significance of Pneumococcal Types. *J Hyg (Lond)* 27(2):113-159.
- 549 2. Ning N & Wang H (2018) [Mechanism and influencing factors of natural transformation
550 in bacteria]. *Sheng Wu Gong Cheng Xue Bao* 34(8):1297-1305.
- 551 3. Lorenz MG & Wackernagel W (1994) Bacterial gene transfer by natural genetic
552 transformation in the environment. *Microbiol Rev* 58(3):563-602.
- 553 4. Johnston C, Martin B, Fichant G, Polard P, & Claverys JP (2014) Bacterial transformation:
554 distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* 12(3):181-196.
- 555 5. Blokesch M (2016) Natural competence for transformation. *Curr Biol* 26(21):R1126-
556 R1130.
- 557 6. Lattar SM, *et al.* (2018) A Mechanism of Unidirectional Transformation, Leading to
558 Antibiotic Resistance, Occurs within Nasopharyngeal Pneumococcal Biofilm Consortia. *MBio*
559 9(3).
- 560 7. Bikash CR, Hamry SR, & Tal-Gan Y (2018) Structure-Activity Relationships of the
561 Competence Stimulating Peptide in *Streptococcus mutans* Reveal Motifs Critical for Membrane
562 Protease SepM Recognition and ComD Receptor Activation. *ACS Infect Dis* 4(9):1385-1394.
- 563 8. Reck M, Tomasch J, & Wagner-Dobler I (2015) The Alternative Sigma Factor SigX
564 Controls Bacteriocin Synthesis and Competence, the Two Quorum Sensing Regulated Traits in
565 *Streptococcus mutans*. *PLoS Genet* 11(7):e1005353.
- 566 9. Wholey WY, Kochan TJ, Storck DN, & Dawid S (2016) Coordinated Bacteriocin
567 Expression and Competence in *Streptococcus pneumoniae* Contributes to Genetic Adaptation
568 through Neighbor Predation. *PLoS Pathog* 12(2):e1005413.
- 569 10. Fontaine L, Wahl A, Flechard M, Mignolet J, & Hols P (2015) Regulation of competence
570 for natural transformation in streptococci. *Infect Genet Evol* 33:343-360.
- 571 11. Claverys JP & Havarstein LS (2002) Extracellular-peptide control of competence for
572 genetic transformation in *Streptococcus pneumoniae*. *Front Biosci* 7:d1798-1814.
- 573 12. Martin B, *et al.* (2010) Expression and maintenance of ComD-ComE, the two-component
574 signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol*
575 *Microbiol* 75(6):1513-1528.
- 576 13. Cheng Q, Campbell EA, Naughton AM, Johnson S, & Masure HR (1997) The com locus
577 controls genetic transformation in *Streptococcus pneumoniae*. *Mol Microbiol* 23(4):683-692.

- 578 14. Fontaine L, *et al.* (2010) A novel pheromone quorum-sensing system controls the
579 development of natural competence in *Streptococcus thermophilus* and *Streptococcus*
580 *salivarius*. *J Bacteriol* 192(5):1444-1454.
- 581 15. Fontaine L, *et al.* (2013) Mechanism of competence activation by the ComRS signalling
582 system in streptococci. *Mol Microbiol* 87(6):1113-1132.
- 583 16. Gardan R, *et al.* (2013) Extracellular life cycle of ComS, the competence-stimulating
584 peptide of *Streptococcus thermophilus*. *J Bacteriol* 195(8):1845-1855.
- 585 17. Mashburn-Warren L, Morrison DA, & Federle MJ (2010) A novel double-tryptophan
586 peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Mol*
587 *Microbiol* 78(3):589-606.
- 588 18. Mignolet J, *et al.* (2018) Circuitry Rewiring Directly Couples Competence to Predation in
589 the Gut Dweller *Streptococcus salivarius*. *Cell Rep* 22(7):1627-1638.
- 590 19. Declerck N, *et al.* (2007) Structure of PlcR: Insights into virulence regulation and evolution
591 of quorum sensing in Gram-positive bacteria. *Proc Natl Acad Sci U S A* 104(47):18490-18495.
- 592 20. Neiditch MB, Capodagli GC, Prehna G, & Federle MJ (2017) Genetic and Structural
593 Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria. *Annu Rev Genet*
594 51:311-333.
- 595 21. D'Andrea LD & Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem Sci*
596 28(12):655-662.
- 597 22. Wintjens R & Rooman M (1996) Structural classification of HTH DNA-binding domains
598 and protein-DNA interaction modes. *J Mol Biol* 262(2):294-313.
- 599 23. Wenderska IB, *et al.* (2012) A novel function for the competence inducing peptide, XIP,
600 as a cell death effector of *Streptococcus mutans*. *FEMS Microbiol Lett* 336(2):104-112.
- 601 24. Wynendaele E, *et al.* (2013) Quorumpeps database: chemical space, microbial origin and
602 functionality of quorum sensing peptides. *Nucleic Acids Res* 41(Database issue):D655-659.
- 603 25. Khan R, *et al.* (2012) Extracellular identification of a processed type II ComR/ComS
604 pheromone of *Streptococcus mutans*. *J Bacteriol* 194(15):3781-3788.
- 605 26. Shanker E, *et al.* (2016) Pheromone recognition and selectivity by ComR proteins among
606 *Streptococcus* species. *PLoS Pathog* 12(10):e1005979.
- 607 27. Talagas A, *et al.* (2016) Structural Insights into Streptococcal Competence Regulation by
608 the Cell-to-Cell Communication System ComRS. *PLoS Pathog* 12(12):e1005980.
- 609 28. Fontaine L, *et al.* (2010) Development of a versatile procedure based on natural
610 transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*.
611 *Appl Environ Microbiol* 76(23):7870-7877.

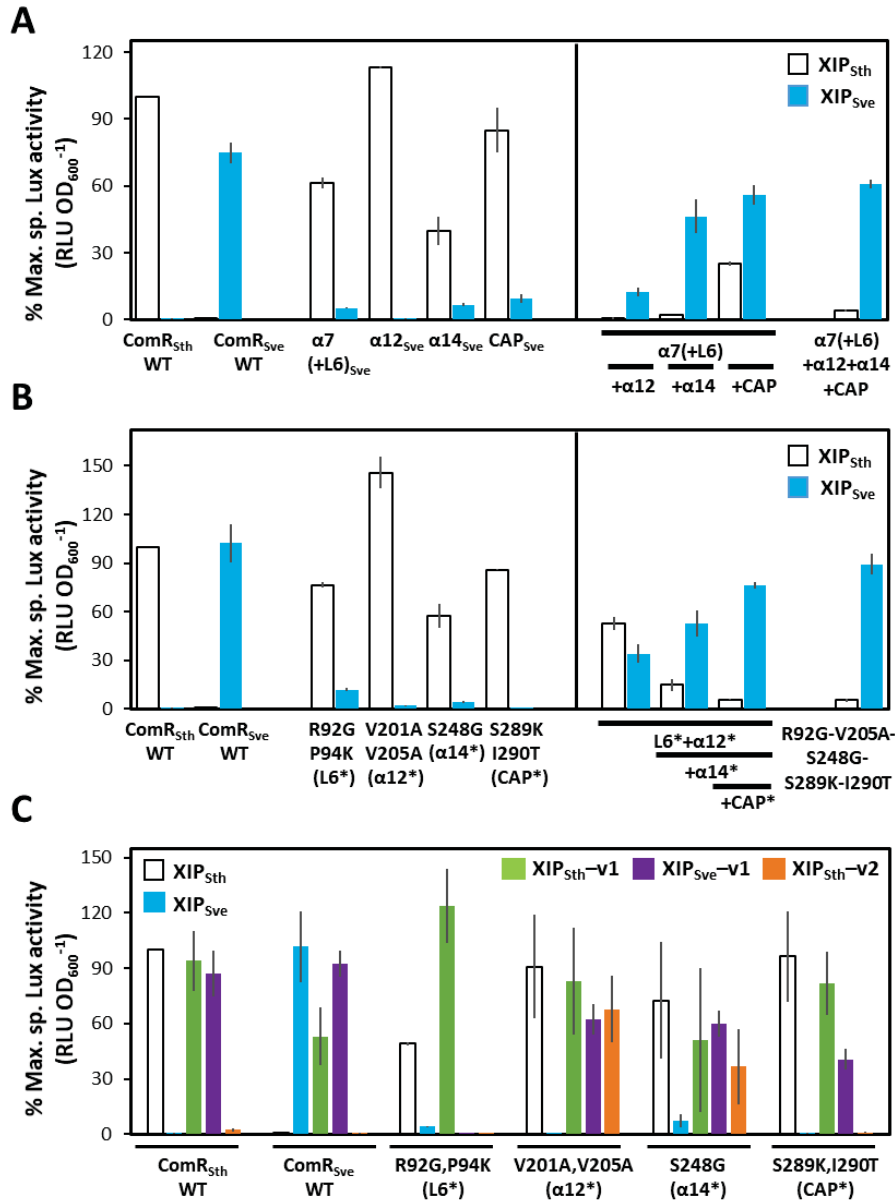
- 612 29. Zeytuni N & Zarivach R (2012) Structural and functional discussion of the tetra-trico-
613 peptide repeat, a protein interaction module. *Structure* 20(3):397-405.
- 614 30. Boutry C, *et al.* (2012) Adaptor protein MecA is a negative regulator of the expression of
615 late competence genes in *Streptococcus thermophilus*. *J Bacteriol* 194(7):1777-1788.
- 616 31. Kabsch W (2010) Xds. *Acta crystallographica. Section D, Biological crystallography*
617 66(Pt 2):125-132.
- 618 32. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Cryst* 40:658-674.
- 619 33. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for
620 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
- 621 34. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta*
622 *Cryst D-Biol. Cryst* 60:2126-2132.
- 623 35. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline
624 state. *J Mol Biol* 372(3):774-797.
- 625 36. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast
626 protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12
627 Pt 1):2256-2268. Epub .
- 628 37. Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction
629 diagrams for drug discovery. *J Chem Inf Model* 51(10):2778-2786.
- 630 38. DeLano WL (2002) The PyMOL Molecular Graphics System. *on World Wide Web*
631 <http://www.pymol.org/>.
- 632
- 633



635

636 **Fig.1: Sequence alignment of ComR, XIP and ComR-box from *S. thermophilus* LMD9**
 637 **and *S. vestibularis* F0396.**

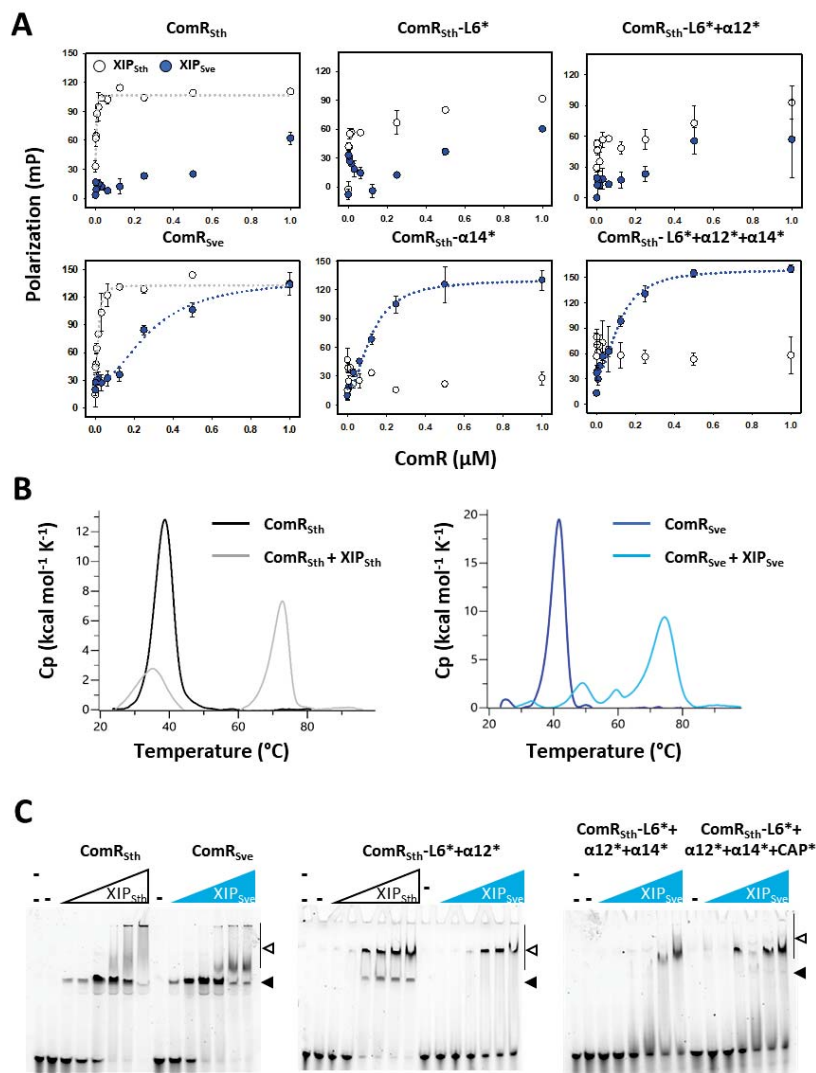
638 **A – ComR.** The α -helices are labelled in blue in the HTH domain and colored by spectrum
 639 according to the structure representation in the TPR domain (**Fig. 4**). Non conserved residues
 640 are indicated in red. Functionally important residues are highlighted: residues implicated in
 641 HTH-sequestration are squared in purple (R35, R39 and R/K-51 in the HTH, and E117, E118,
 642 E146, D147 in the TPR-domain); residues involved in dimer stabilization are squared in orange
 643 (K87, D200, K246, E282); the conserved PTY motif and residue K100 interacting with the XIP
 644 C-terminus are squared in green. Red arrows indicate positions of substitutions that have been
 645 introduced by mutation in ComR_{Sth} (R/G-92, P/K-94, V/A-201, V/A-205, S/G-248, S/K-289,
 646 I/T-290). **B – XIP.** Non conserved residues are indicated in red. **C – ComR-box of P_{comS} and**
 647 **P_{comX}.** The palindromic sequences are underlined. Differences in the nucleotide sequence are
 648 indicated in bold.



649

650 **Fig. 2: In vivo selectivity assays.**

651 Maximum specific luciferase activity (RLU OD₆₀₀⁻¹), expressed in percentage, emitted by the
 652 *S. thermophilus* P_{ComS} reporter strain. The strain producing A - ComR_{Sth} WT, ComR_{Sve} WT, and
 653 chimeric proteins (ComR_{Sth} + α helices from ComR_{Sve}) grown in presence of 1 μM XIP_{Sth}
 654 (LPYFAGCL, open bars) or XIP_{Sve} (VPFFMIYY, cyan bars), B - ComR_{Sth} point mutants grown
 655 in presence of 1 μM XIP_{Sth} (open bars) or XIP_{Sve} (cyan bars), and C - ComR_{Sth} WT, ComR_{Sve}
 656 WT, and ComR_{Sth} point mutants grown in presence of XIP_{Sth} (open bars), XIP_{Sve} (cyan bars),
 657 XIP_{Sth}-v1 (LPYFAICL, green bars), XIP_{Sve}-v1 (VPFFMGYY, purple bars), or XIP_{Sth}-v2
 658 (LPYFAGYY, orange bars). Bars represent the average of at least three independent repeats ±
 659 standard deviation. Loop α6-α7 is noted L6. Point mutants are noted as L6*, α12*, α14* and
 660 CAP*.
 661

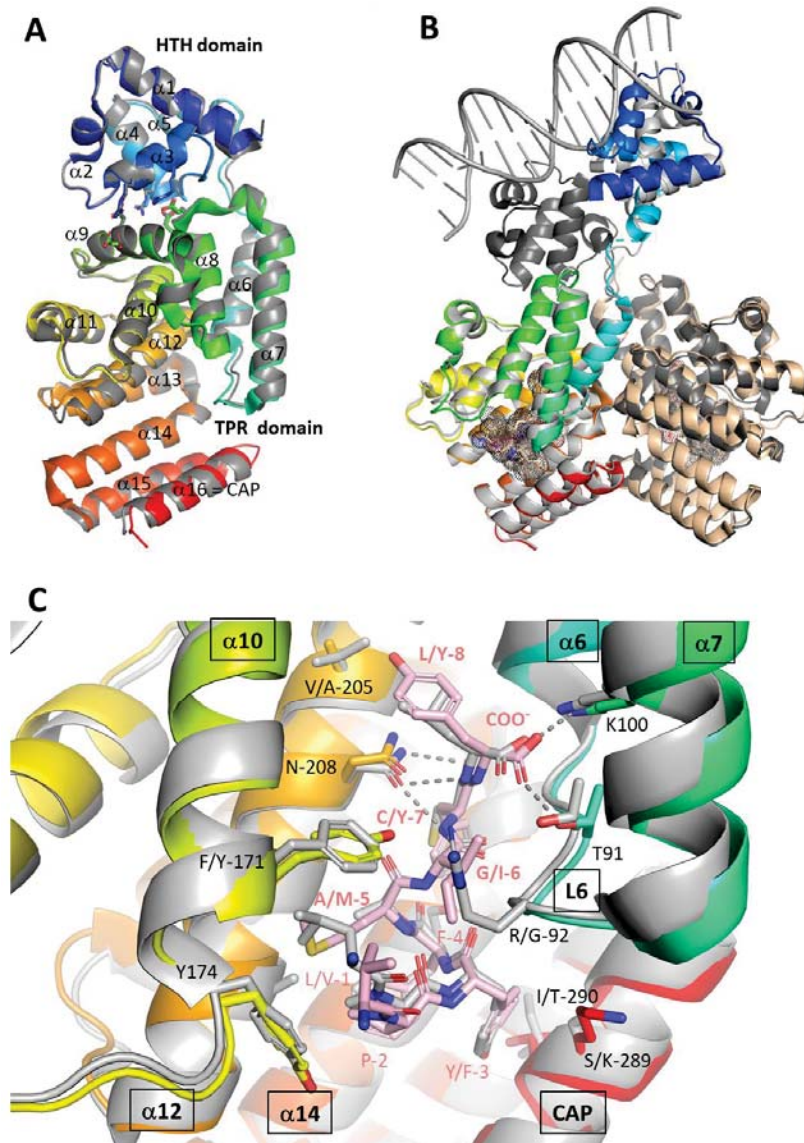


662

663 **Fig. 3: *In vitro* interaction measurements.**

664 **A - Fluorescence polarization assays.** Titration was performed in presence of a fixed
 665 concentration (30 nM) of FITC N-labelled XIP_{Sth} or XIP_{Sve} peptide (open or blue dots,
 666 respectively) and serial dilutions of ComR starting from 1 μM. Experimental values represent
 667 the averages ± standard deviation of at least three replicates. Curves reaching the saturation
 668 were fitted by the one ligand binding site equation to deduce the apparent K_d value. **B – DSC**
 669 **analysis of ComRs WT.** Comparison of the variation of heat capacity (C_p) during thermal
 670 denaturation of 15 μM ComR alone and in complex with 150 μM of the cognate XIP peptide.
 671 The thermal transition midpoint (T_m) representing the stability of the sample was determined
 672 at the top of the DSC peak. We verified that the peptide alone did not give any signal. **C –**
 673 **EMSA.** Labelled P_{comS} 40-bp DNA fragments (20 ng) were incubated with a fixed
 674 concentration of ComRs WT or mutants (1 μM, except for ComR_{Sth}-L6*+α12* used at 4 μM)
 675 in absence of peptide (-) or in presence of increasing concentrations of peptide (from 0 to 2 μM,
 676 except for ComR_{Sth}-L6*+α12*, which was incubated with 0 to 8 μM XIP). The well-defined
 677 ComR-XIP-DNA complex and multimeric complexes are respectively indicated by closed and
 678 open arrow heads. One representative experiment of at least two independent replicates is
 679 shown.

680

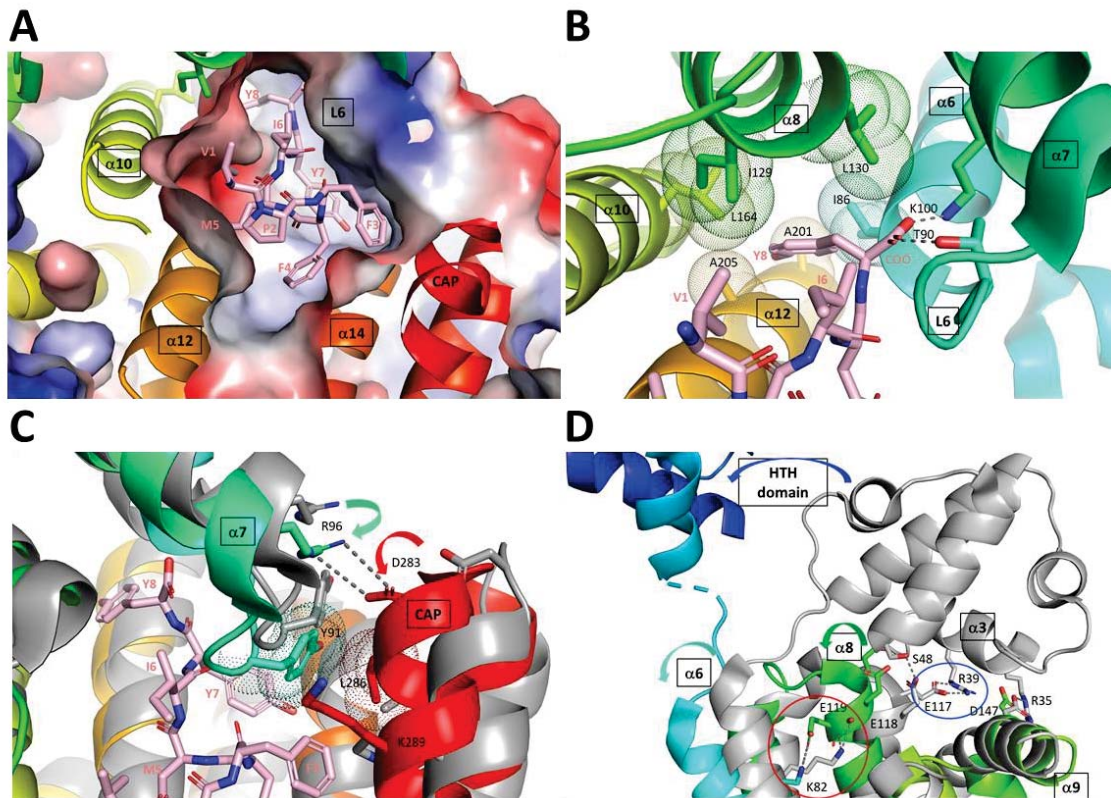


681

682 **Fig. 4: Structural comparison of ComR_{Sth} and ComR_{Sve}**

683 **A – Superimposition of the apo forms.** The superimposed crystal structures of the proteins are shown as cartoon. ComR_{Sve} (PDB ID 6HU8) is colored by spectrum from blue (N-ter) to red (C-ter). ComR_{Sth} (PDB ID 5JUF) is shown in gray. The α-helices are labelled as well as the HTH and TPR domains. **B – Superimposition of the ComR_{Sve}·XIP_{Sve} and ComR_{Sth}·XIP_{Sth}·DNA complexes.** The proteins are shown in cartoon representations and the bound peptides are shown in sticks and highlighted by dots. The ternary complex from *S. thermophilus* (PDB ID 5JUB) is colored in gray. One subunit of the binary complex from *S. vestibularis* (PDB ID 6HUA) is colored by spectrum and the second subunit, which HTH domain is not visible, is shown in beige. **C – Close view of the superimposed peptide-binding sites.** ComR_{Sve} is colored by spectrum and ComR_{Sth} is in gray. The bound peptides are shown in sticks, colored by atom types (blue for N and red for O). The C atoms of XIP_{Sve} (VPFFMIYY) are shown in pink and those of XIP_{Sth} (LPYFAGCL) in gray. Their residues are labelled in pink. Most relevant ComR residues are highlighted in sticks and labelled in black. The α-helices are labelled.

696
697

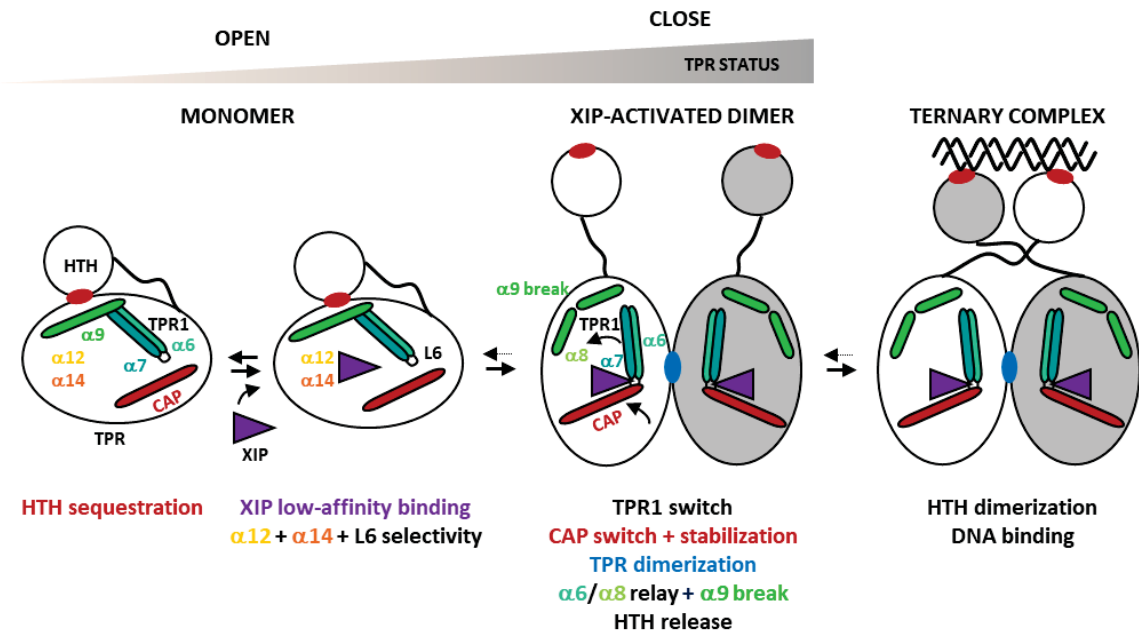


699

700

701 **Figure 5: Structural insights into ComR activation mechanism.**

702 The protein from the ComR_{Sve}·XIP_{Sve} binary complex (PDB ID 6HUA) is represented as
 703 cartoon and colored by spectrum as in Fig. 4. Loop L6 and important helices are labelled. The
 704 bound XIP_{Sve} (VPFFMIYY) is represented in sticks as in Fig. 4C. **A –XIP-binding pocket.**
 705 The closed peptide-bound pocket of ComR^{Sth} is shown as surface colored by electrostatic
 706 potential (basic regions in blue, acidic region in red and non-polar residues in white). **B –**
 707 **Recruitment of loop L6.** Residues from helices $\alpha 6$, $\alpha 8$, $\alpha 10$ and $\alpha 12$ forming the hydrophobic
 708 pocket where residue XIP_{Sve}-Y8 binds are shown in sticks and dots. The interactions between
 709 the C-terminal carboxylate (COO⁻) group of the peptide and residues K100 and T90 are shown
 710 as black dashed lines. **C - Stabilization of the TPR closed conformation.** The complex
 711 ComR_{Sve}·XIP_{Sve} is superimposed with the apo form of ComR_{Sve} (PDB ID 6HU8) shown as
 712 cartoon colored in gray. The shifts of helices $\alpha 7$ and CAP are highlighted by green and red
 713 arrows, respectively. The stabilizing salt bridge between R96 and D283 is shown as black
 714 dashed lines. Residues Y91 and L286 forming a hydrophobic stacking interaction are shown as
 715 sticks and dots. **D – Release of the HTH domain.** The superimposed structures of the apo and
 716 peptide-bound forms of ComR_{Sve} are colored as in panel C, shown from the opposite face
 717 compared to panels A, B and C. The shift of $\alpha 6$ (cyan arrow) is transmitted to $\alpha 8$ (green arrow)
 718 through the conserved salt bridge between residues K82 and E119 (red circle). The resulting
 719 break of the salt bridge between E117 and residue R35 from helix $\alpha 3$ (blue circle) releases the
 720 HTH domain, which is shifted away (blue arrow).
 721



723

724 **Figure 6: Schematic drawing of the proposed XIP-driven activation mechanism of ComR.**
 725 The HTH and TPR domains of the protein are shown with the flexible linker. Helices $\alpha 12$, $\alpha 14$,
 726 and L6 implicated in peptide binding are labelled. Helices $\alpha 6$, $\alpha 7$, $\alpha 8$, $\alpha 9$ and CAP, which are
 727 directly implicated in the activating conformational change, are shown as sticks and labelled.
 728 Helices are colored as in Fig. 4. Red and blue ellipses represent residues involved in HTH
 729 sequestration and TPR dimerization, respectively.
 730

PNAS

www.pnas.org

731

732 Supplementary Information for

733 **Molecular dissection of pheromone selectivity in the competence signaling**
734 **system ComRS of streptococci**

735

736 Laura Ledesma-Garcia, Jordhan Thuillier, Armando Guzman-Espinola, Imke Ensinck, Inès Li
737 de la Sierra-Gallay, Nouredine Lazar, Magali Aumont-Nicaise, Johann Mignolet, Patrice
738 Soumillion, Sylvie Nessler*, Pascal Hols*

739

740 * Co-corresponding authors: Sylvie Nessler and Pascal Hols

741 **Email:** Sylvie.Nessler@i2bc.paris-saclay.fr and Pascal.Hols@uclouvain.be

742

743 **This PDF file includes:**

744 **SI Materials and Methods**

745 **Figures S1 to S7**

746 **Tables S1 to S5**

747 **SI References**

748

749 **SI Materials and Methods**

750

751 *Bacterial strains, plasmids, oligonucleotides and growth conditions*

752 Bacterial strains, plasmids and oligonucleotides used in this study are listed in **Tables S2-S4**.
753 *Escherichia coli* was grown in LB medium with shaking at 37°C (1). *S. thermophilus* LMD-9
754 and *S. vestibularis* F0396 were grown at 37°C in M17 broth (Difco Laboratories Inc., MI) or in
755 CDM (2). All synthetic media contained 1% glucose [w/v] (M17G and CDMG broth,
756 respectively). When required, ampicillin (200 µg.ml⁻¹ for *E. coli*), streptomycin (50 µg.ml⁻¹ for
757 *E. coli*), chloramphenicol (5 µg.ml⁻¹ for *S. thermophilus*) or erythromycin (2.5 µg.ml⁻¹ for *S.*
758 *thermophilus* and *S. vestibularis*) was added to the media. Plates inoculated with *S.*
759 *thermophilus* or *S. vestibularis* cells were incubated anaerobically (BBL GasPak systems,
760 Becton Dickinson, NJ) at 37°C.

761

762 *Transformation frequency analysis*

763 Transformation assays in *S. thermophilus* and *S. vestibularis* were performed by addition of
764 XIP peptide as described previously (3). Transformation frequency is calculated as the ratio of
765 transformants (erythromycin-resistant colony-forming units [CFU]) to the total CFU count per
766 1 µg.ml⁻¹ of multimeric circular DNA (pGhost9-core plasmid).

767

768 *DNA techniques and electrotransformation*

769 *E. coli* strain TOP10 (Invitrogen, CA) was used as a host in cloning experiments.
770 Electrotransformations were performed as described previously (4) and all DNA manipulations
771 were performed according to standard procedures (1). PCRs were achieved with Phusion high-
772 fidelity DNA polymerase (Finnzymes, Espoo, Finland). The primers used in this study were
773 purchased from Eurogentec (Seraing, Belgium).

774

775 *Construction of comR-strep expression plasmids*

776 For construction of plasmid pBADcomR_{Sve}-strep, the gene *comR*_{Sve}-strep was amplified by
777 PCR from chromosomal DNA of *S. vestibularis* F0396 with primers carrying restriction sites
778 at the 5'-ends (*Bsp*HI and *Eco*RI in forward and reverse primers, respectively). The reverse
779 primer also encoded a glycine/alanine flexible linker and the StreptagII affinity tag. The
780 digested PCR product was cloned into the *Nco*I/*Eco*RI-restricted pBADHisA vector. For the
781 construction of plasmids expressing the *comR*_{Sth}-strep point mutants, insertion of the desired
782 mutations in the template vector pBADcomR_{Sth}-strep (5) was obtained using inverse PCR with
783 partially overlapping primers. The resulting PCR products were re-circularized by the Gibson
784 method (6) and the plasmid used as template was eliminated by *Dpn*I digestion. Plasmids
785 carrying distant point mutations were obtained by performing successive mutagenesis rounds.
786 All *comR* variants were validated by DNA sequencing.

787

788 *Construction of S. thermophilus mutant strains by natural transformation*

789 The different *comR* mutants were expressed from the *comR* native locus in a ComS-deficient
790 LMD-9 strain carrying a P_{comS}-luxAB reporter fusion at the *blp* locus (5). Replacement of the
791 *comRS* gene by a linear DNA fragment carrying the *comR* mutant and a P₃₂-cat cassette was
792 performed by double homologous recombination using natural transformation.

793 The chimeric DNA fragments used for *comR*_{Sth}/*comR*_{Sve} helix swapping were obtained *in vitro*
794 by joining three PCR fragments using overlapping PCR. *S. vestibularis* F0396 chromosomal
795 DNA was used as template to amplify the region to be exchanged, and overlapping sequences
796 were included in the internal primers. In the case of point mutants, two PCR fragments were
797 joined by overlapping PCR using internal primers carrying the mutation to incorporate. In all
798 cases, the upstream PCR fragment included a ~1-kb region located upstream of *comR*, and a

799 downstream region carrying the *comS::P₃₂-cat* cassette and ~1 kb-region located downstream
800 of *cat*, amplified from a ComS-deficient strain (5). The complementarity of the flanks of the
801 PCR products allowed their junction, and finally the full-length product was amplified with the
802 external primers. Natural transformation of the purified overlapping PCR products was
803 performed as previously described (3, 7). Transformants were selected on M17G broth
804 containing chloramphenicol. Integration of the antibiotic resistance cassette at the right location
805 was checked by PCR and the candidates were verified by sequencing.
806 Strains carrying distant mutations were obtained by performing successive mutagenesis rounds
807 (strains LL16-LL19)

808

809 *Protein purification*

810 Purification of ComR-StreptagII proteins was performed as described previously with some
811 modifications (5). One litre of culture at an initial OD₆₀₀ of 0.05 was incubated at 42 °C with
812 shaking until an OD₆₀₀ of 0.5. Protein production was induced by adding 0.05% of L-arabinose
813 and the culture was incubated at 30 °C for 4 h with continuous shaking. Cells were then
814 harvested by centrifugation (5,000 × *g* for 15 min), and pellets were resuspended in cold buffer
815 W (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) supplemented with 0.5 mg ml⁻¹
816 of lysozyme, incubated for 30 min at 4°C, and sonicated at 4°C. The soluble fractions were
817 collected after centrifugation (15,000 × *g* for 60 min at 4°C) and purification of the recombinant
818 ComR-StreptagII proteins was performed by one-step affinity chromatography with Strep-
819 Tactin Superflow column (IBA BioTAGnology, Göttingen, Germany) according to the
820 manufacturer's instructions. Purified proteins were aliquoted and stored at -80 °C in buffer W
821 with 10% glycerol [v/v]. For crystallization purposes, samples eluted from the StrepTag HP
822 5ml column (GE-Healthcare) were loaded on a HiLoad 16/600 Superdex 75 prep grade column
823 (GE-Healthcare) equilibrated in buffer SEC (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM
824 EDTA, 10% glycerol [v/v]). Purified proteins were concentrated by ultra-filtration using
825 Vivaspinn-20 (Sartorius) with a molecular cut-off of 10 kDa, aliquoted, flash frozen and
826 conserved at -20°C. Protein purity was analyzed by SDS-PAGE and concentration was
827 measured using a Nanodrop apparatus (ThermoFisher Scientific).

828

829 *In silico modelling*

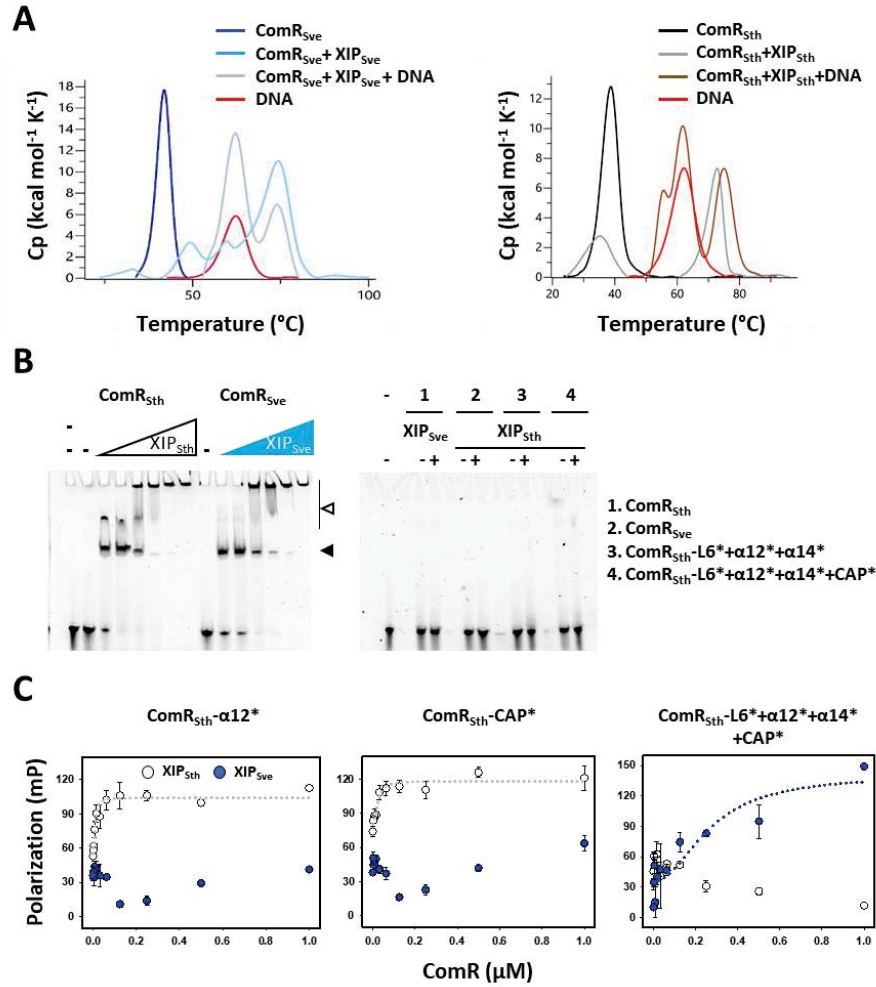
830 We used the ComR_{Sth}·XIP_{Sth}·DNA structure (PDB ID 5JUB) (8) and the ComR_{Sve}·XIP_{Sve} (PDB
831 ID 6HUA) as starting models to manually introduce substitutions in the sequence of the bound
832 peptide and produce initial models of ComR_{Sth} and ComR_{Sves} in complex with non-cognate
833 peptides. These models were optimized by energy minimization using the NOMAD-Ref
834 Gromacs server software (9). As a control, energy minimization was also performed on the
835 reference crystal structures. The absence of clashes in the minimized complex was probed by
836 using Molprobit validation tool in COOT (10).

837

838 *Multiple sequence alignments*

839 Multiple alignments of ComR orthologues and ComR-binding sites were generated with Clustal
840 Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (11). The formatting of alignments was
841 performed with CLC Main Workbench 7 (<https://www.qiagenbioinformatics.com/>).

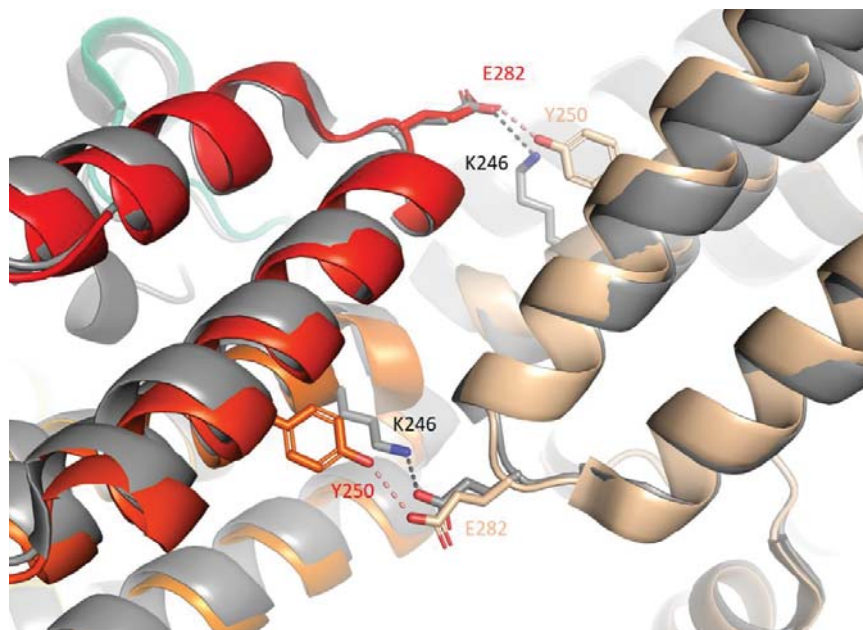
842



843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

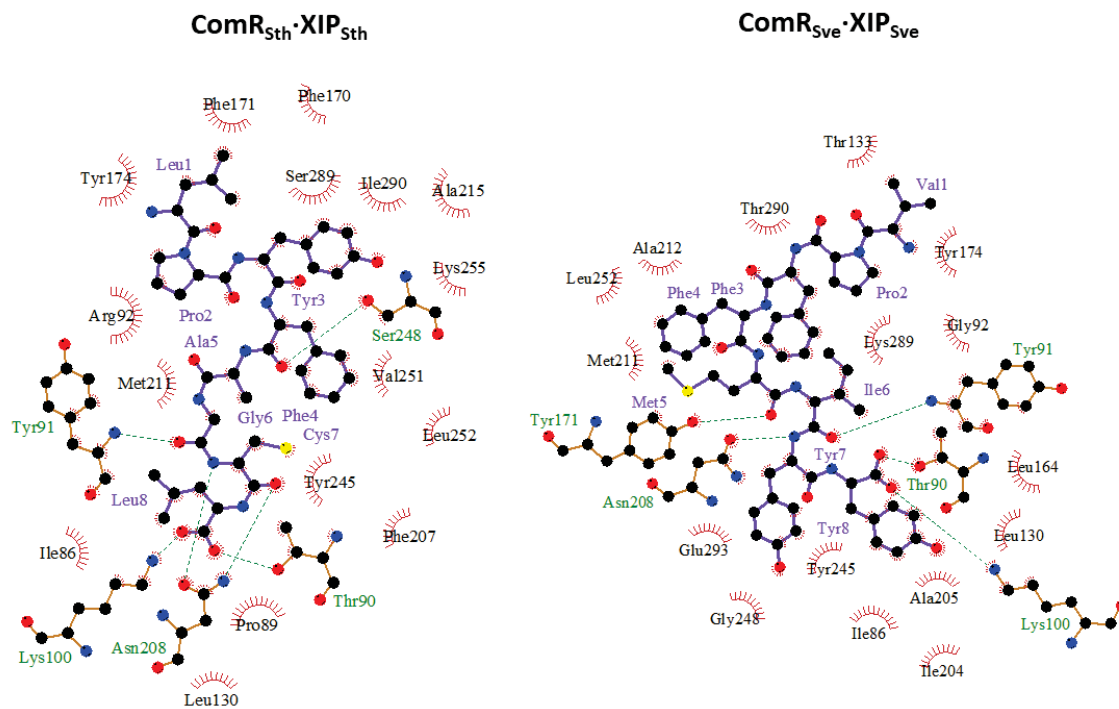
Figure S1: Additional *in vitro* interaction assays with ComRs WT and mutants.

A – DSC assays. Variation of heat capacity (Cp) during thermal denaturation of 15 μM ComR alone and in complex with XIP at 150 μM colored as in Fig. 3B compared with a 20-bp DNA fragment carrying the ComR box of *P_{comX}* alone at 45 μM and the ternary complex ComR·XIP·DNA at the same concentrations. **B – EMSA.** (Left panel) ComRs WT controls for the ComR_{Sth}-L6*+α12* mutant shown in Fig. 3C. The assays were performed at 4 μM protein in absence of peptide (-) or in presence of increasing concentrations of peptide (0 to 8 μM XIP) by using the labelled 40-bp *P_{comS}* DNA fragment (20 ng). The well-defined ComR·XIP·DNA complex and multimeric complexes are respectively indicated by closed and open arrow heads. (Right panel) Controls showing the absence of XIP cross-activation for ComRs WT and ComR_{Sth} mutants (L6*+α12*+α14* and L6*+α12*+α14*+CAP) shown in Fig. 3C. The assays were performed at 1 μM protein in absence (-) or presence of 2 μM XIP (+) by using the labelled 40-bp *P_{comS}* DNA fragment (20 ng). One representative experiment of at least two independent replicates is shown. **C – Fluorescence polarization assays.** ComR_{Sth}-α12* and ComR_{Sth}-CAP* mutants with no change in peptide affinity, and ComR_{Sth}-L6*+α12*+α14*+CAP* showing a similar affinity pattern than the mutant protein ComR_{Sth}-L6*+α12*+α14*. Titration was performed in presence of a fixed concentration (30 nM) of FITC N-labelled XIP_{Sth} or XIP_{Sve} peptide (open or blue dots, respectively) and serial dilutions of ComR starting from 1 μM. Experimental values represent the averages ± standard deviation of at least three replicates.



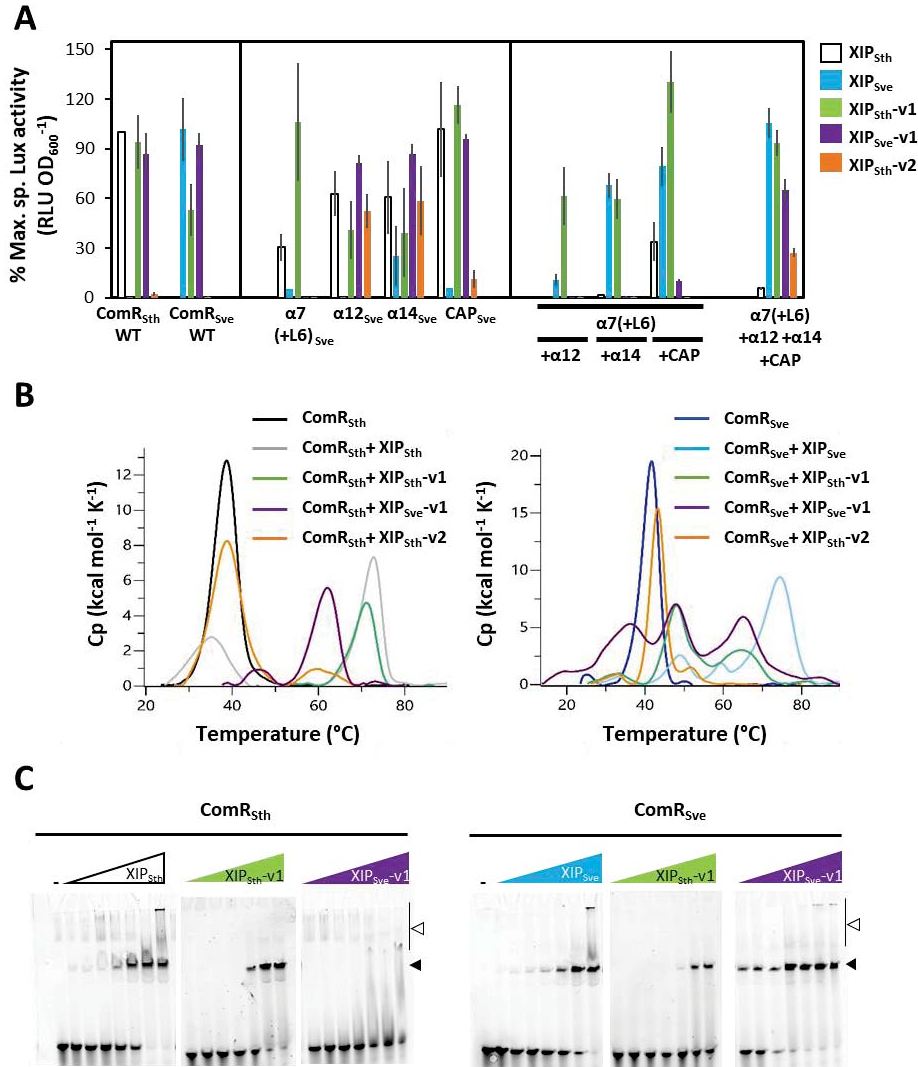
864
865
866
867
868
869
870
871
872
873

Figure S2: Comparison of the dimer interface of ComR_{Sth} and ComR_{Sve}. The proteins are shown in cartoon representations. The ternary complex ComR_{Sth}·XIP_{Sth}·DNA (PDB ID 5JUB) is colored in gray. One subunit of the ComR_{Sve}·XIP_{Sve} binary complex (PDB ID 6HUA) is colored by spectrum and the second subunit is shown in beige. The conserved residue E282 forms a salt bridge with neighboring residue K246 in ComR_{Sth} and H-bonds with Y250 in ComR_{Sve}. These interactions are highlighted by black and red dashed lines, respectively. The contact residues are shown in sticks and labelled.



874
875
876
877
878
879
880
881
882

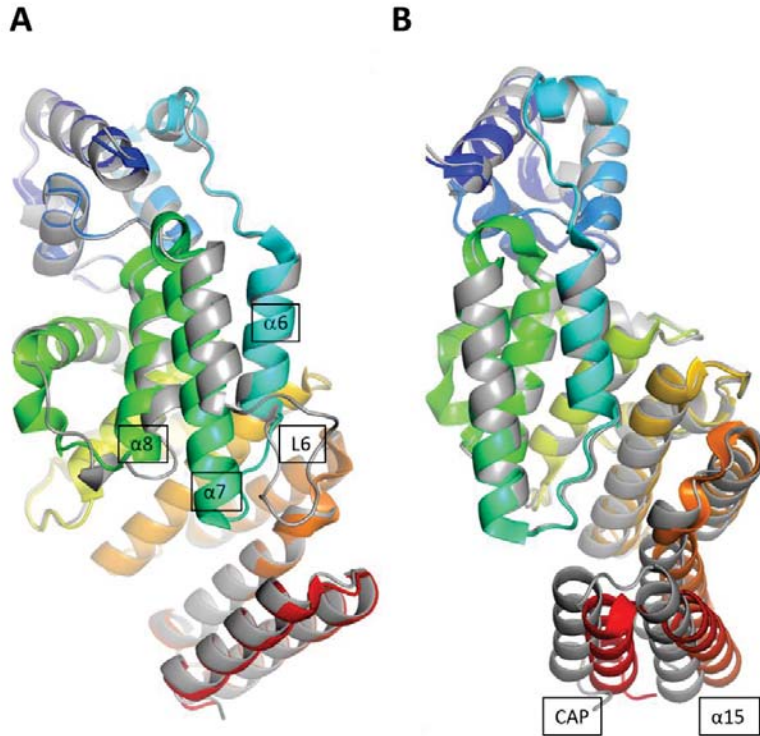
Figure S3: Ligplot diagram of the peptide-binding mode. 2D representation of the peptide-protein interactions in the ComR_{Sth}·XIP_{Sth}·DNA ternary complex (PDB ID 5JUB) on the left side and in the ComR_{Sve}·XIP_{Sve} binary complex (PDB ID 6HUA) on the right. The peptide and protein side chains are shown in ball-and-stick representation, with the peptide bonds colored in purple. H-bonds are shown as green dotted lines and the spoked arcs represent protein residues making non-bonded van der Waals contacts with the peptide.



883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901

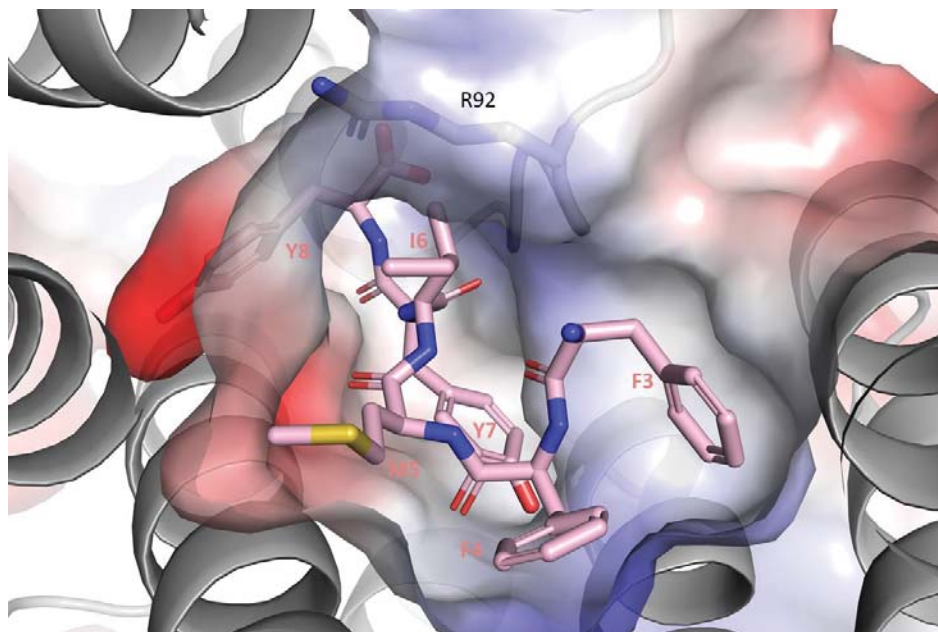
Figure S4: Additional *in vivo* and *in vitro* assays with XIP variants

A - Activity assays with ComR_{Sth} chimeric proteins. Maximum specific luciferase activity (RLU OD₆₀₀⁻¹), expressed in percentage, emitted by the *S. thermophilus* P_{comS} reporter strain. The strains were grown in presence of 1 μM XIP_{Sth} (LPYFAGCL, open bars), XIP_{Sve} (VPFFMIYY, blue bars), XIP_{Sth-v1} (LPYFAICL, green bars), XIP_{Sve-v1} (VPFFMGYY, purple bars), or XIP_{Sth-v2} (LPYFAGYY, orange bars). Bars represent the average of at least three independent repeats ± standard deviation. **B - DSC analysis of ComRs WT with all XIP variants.** Comparison of the variation of heat capacity (Cp) during thermal denaturation of 15 μM ComR alone and in complex with 150 μM of XIP_{Sth}, XIP_{Sve}, XIP_{Sth-v1}, XIP_{Sve-v1} or XIP_{Sth-v2}. **C - EMSA of ComRs WT with XIP_{Sth-v1} and XIP_{Sve-v1}.** Labelled P_{comS} 40-bp DNA fragments (20 ng) were incubated with a fixed concentration of ComRs WT (1 μM) in absence of peptide (-) or in presence of increasing concentrations (from 0 to 2 μM) of XIP_{Sth} (open triangles), XIP_{Sve} (cyan triangles), XIP_{Sth-v1} (green triangles) or XIP_{Sve-v1} (purple triangles). The well-defined ComR·XIP·DNA complex and multimeric complexes are respectively indicated by closed and open arrow heads. One representative experiment of at least two independent replicates is shown.



902
 903
 904
 905
 906
 907
 908
 909
 910
 911

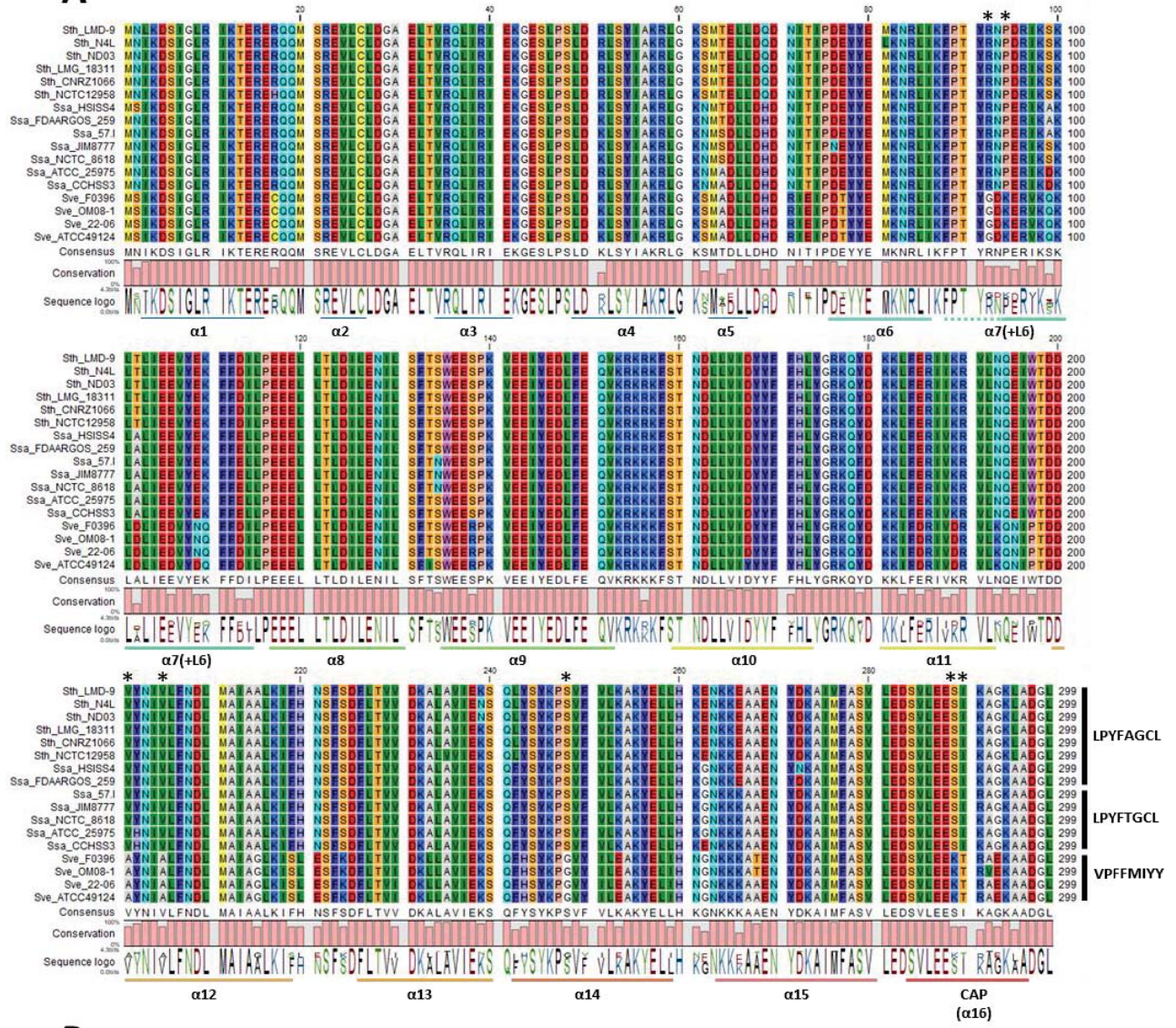
Figure S5: Alternative conformations of apo ComR_{sth}. An alternative crystal structure of apo ComR_{sth} has been solved in space group C2 at 2.5Å resolution (PDB ID 6QER). The protein is monomeric, as observed in the previous structure solved in space group C222 (PDB ID 5JUF), but the asymmetric unit contains 4 independent molecules (chains A, B, C and D) displaying distinct conformations. The structure of 5JUF, shown as cartoon and colored by spectrum, is compared to the two distinct chains of 6QER, superimposed in gray. **A** – Alternative positions of loop L6. **B** – Alternative conformation of the TPR superhelix.



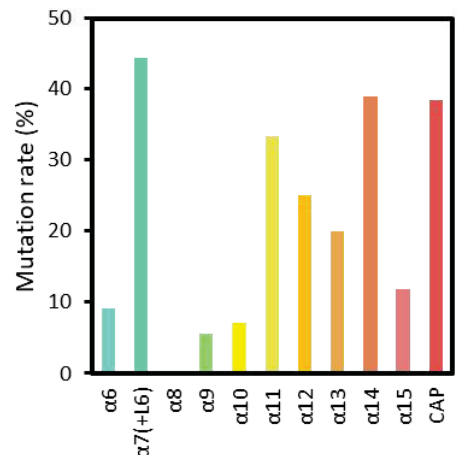
912
913
914
915
916
917
918
919
920
921
922
923

Figure S6: Modelling of the heterologous ComR_{Sth}·XIP_{Sve} complex. Close view of the peptide-binding pocket of ComR_{Sth} from the ComR_{Sth}·XIP_{Sth}·DNA structure (PDB ID 5JUB) (8) shown as electrostatic surface (basic regions in blue, acidic region in red and non-polar residues in white). The sequence of the bound peptide was manually modified into the XIP_{Sve} sequence and the resulting heterologous model was optimized by energy minimization. The resulting minimized complex shows in particular that a small shift of ComR_{Sth} residue R92 avoids clashes with XIP_{Sve}-I6. The bound peptide is shown in sticks colored by atom types with C atoms in pink.

A



B



924
925
926

927 **Figure S7: Conservation of ComR orthologues from the salivarius group of streptococci.**
928 **A – Multiple alignment.** All unique ComR sequences from *S. vestibularis* (4 ComR), *S.*
929 *thermophilus* (6 ComR), and *S. salivarius* (7 ComR) available in public databases were aligned
930 with Clustal Omega and visualized with CLC Main Workbench 7. Residues are color-coded
931 according to the Rasmol color scheme and the conservation between ComRs is summarized
932 below the alignment. The different α -helices are delimited and colored according to figure 4,
933 and asterisks point out the residues mutated during this work. The XIP peptide sequence
934 associated to each ComR is shown at the end of the sequence of the regulator. **B – α -helix**
935 **mutation rate.** Percentage of non-conserved residues ($> 5\%$ at each position) in each α -helix
936 of the TPR domain of all unique ComR orthologues. The calculated percentage for $\alpha 7$ includes
937 loop L6 ($\alpha 7(+L6)$). A higher mutation rate is observed for $\alpha 7(+L6)$, $\alpha 14$, and CAP.
938

939 **Table S1: Transformation rate of *S. thermophilus* LMD-9 and *S.vestibularis* F0396**

Strains	No XIP	+ cognate XIP	
		0.1 μ M	1 μ M
<i>S. thermophilus</i> LMD-9	9.9×10^{-5a}	9.6×10^{-4}	4.6×10^{-4}
<i>S. vestibularis</i> F0396	ND ^b	1.0×10^{-7}	1.6×10^{-5}
	No XIP	+ 1 μ M XIP _{StH}	+ 1 μ M XIP _{Sve}
<i>S. thermophilus</i> LMD-9 Δ comS	ND	2.0×10^{-6}	ND

940 ^a Data presented in the table are one representative experiment of 2 independent experiments showing similar
 941 results.

942 ^b ND, not detected.

943

944 **Table S2: Bacterial strains used in this study**

Strains	Genotype	Characteristic ^a	Reference/source
<i>E. coli</i>			
<i>E. coli</i> TOP10	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS mcrBC</i>) φ80 <i>lacZ</i> Δ <i>M15</i> Δ <i>lacX74</i> <i>recA1 araD139</i> Δ(<i>ara-leu</i>) 7697 <i>galU galK rpsL</i> (Str ^r) <i>endA1</i> <i>nupG</i> λ ⁻		Invitrogen, CA
<i>Streptococcus</i> sp.			
<i>S. vestibularis</i> F0396	WT		BEI Resources ^b
<i>S. thermophilus</i> LMD-9	WT		ATCC ^c
LF121	LMD-9 (<i>blpD-blpX</i>):P _{comS} - <i>luxA</i>		(5)
LF134	LF121 Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	(5)
LL1	LF121 <i>comR</i> :: <i>comR</i> _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL2	LF121 <i>comR</i> :: <i>comR</i> -α7(+L6) _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL3	LF121 <i>comR</i> :: <i>comR</i> -α12 _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL4	LF121 <i>comR</i> :: <i>comR</i> -α14 _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL5	LF121 <i>comR</i> :: <i>comR</i> -CAP _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL6	LF121 <i>comR</i> :: <i>comR</i> -α7(+L6)-α12 _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL7	LF121 <i>comR</i> :: <i>comR</i> -α7(+L6)-α14 _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL8	LF121 <i>comR</i> :: <i>comR</i> -α7(+L6)-CAP _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL9	LF121 <i>comR</i> :: <i>comR</i> -α7(+L6)-α12-α14-CAP _{Sve} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL10	LF121 <i>comR</i> :: <i>comR</i> _{R92G-P94K} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i> ; L6*	Cm ^r	This study
LL11	LF121 <i>comR</i> :: <i>comR</i> _{V201A,V205A} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i> ; α12*	Cm ^r	This study
LL12	LF121 <i>comR</i> :: <i>comR</i> _{S248G} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i> ; α14*	Cm ^r	This study
LL13	LF121 <i>comR</i> :: <i>comR</i> _{S289K-I290T} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i> ; CAP*	Cm ^r	This study
LL14	LF121 <i>comR</i> :: <i>comR</i> _{R92A} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL15	LF121 <i>comR</i> :: <i>comR</i> _{V205A} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL16	LF121 <i>comR</i> :: <i>comR</i> _{R92G-P94K-V201A-V205A} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL17	LF121 <i>comR</i> :: <i>comR</i> _{R92G-P94K-V201A-V205A-S248G} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL18	LF121 <i>comR</i> :: <i>comR</i> _{R92G-P94K-V201A-V205A-S248G-S289K-I290T} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study
LL19	LF121 <i>comR</i> :: <i>comR</i> _{R92G-V205A-S248G-S289K-I290T} , Δ <i>comS</i> ::P ₃₂ - <i>cat</i>	Cm ^r	This study

945

946 ^a Cm^r indicates resistance to chloramphenicol.947 ^c BEI Resources, 10801 University Boulevard, Manassas, VA.948 ^b ATCC, American Type Culture Collection, Rockville, MD.

949

950
951

Table S3: Plasmids used in this study

Plasmid name	Description	Characteristic ^a	Reference
pGhost9-core	pGhost9 vector derivative forming multimers in presence of the site-specific recombinase TnpI of Tn4430	Ery ^r	Lab. collection
pBADHisA	Expression vector for protein purification	Ap ^r	Invitrogen
pBADcomR _{Sth} -strep	pBADHisA derivative containing the translation fusion P _{araBAD} - <i>comR</i> _{Sth} - <i>strep</i>	Ap ^r	(5)
pBADcomR _{Sve} -strep	pBADHisA derivative containing the translation fusion P _{araBAD} - <i>comR</i> _{Sve} - <i>strep</i>	Ap ^r	This study
pBADcomR _{R92G-P94K} -strep	pBADcomR _{Sth} -strep derivative containing the mutant <i>comR</i> _{Sth-R92G-P94K} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{V201A-V205A} -strep	pBADcomR _{Sth} -strep derivative containing the mutant <i>comR</i> _{Sth-V201A-V205A} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{S248G} -strep	pBADcomR _{Sth} -strep derivative containing the mutant <i>comR</i> _{Sth-S248G} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{S289K-I290T} -strep	pBADcomR _{Sth} -strep derivative containing the mutant <i>comR</i> _{Sth-S289K-I290T} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{R92G-P94K-V201A-V205A} -strep	pBADcomR _{Sth} -strep derivative containing the mutant <i>comR</i> _{Sth-R92G-P94K-V201A-V205A} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{R92G-P94K-V201A-V205A-S248G} -strep	pBADcomR _{LMD-9} -strep derivative containing the mutant <i>comR</i> _{Sth-R92G-P94K-V201A-V205A-S248G} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study
pBADcomR _{R92G-P94K-V201A-V205A-S248G-S289K-I290T} -strep	pBADcomR _{LMD-9} -strep derivative containing the mutant <i>comR</i> _{Sth-R92G-P94K-V201A-V205A-S248G-S289K-I290T} - <i>strep</i> allele under the control of P _{araBAD}	Ap ^r	This study

952

953

^a Ery^r and Ap^r indicate resistance to erythromycin and ampicillin, respectively.

954

955

Table S4: Primers used in this study

Oligonucleotide name	Sequence and description (5'-3') ^a	Reference	
Primers used to insert <i>comR_{Sve}</i>			
pBAD-ComRsvest-Fw	GGTCATGAGCATAAAAGACAGC ATTGGAC	Construction of pBADcomR _{Sve} -strep. <i>Bsp</i> HI restriction site	This study
pBAD-ComRsv-Rv + streptag	GGGAATTCTATTCTCGAACTGC GGGTGACTCCAACCAGCACCCAA ACCATCTGCCGCTTCTC	Construction of pBADcomR _{Sve} -strep. <i>Eco</i> RI restriction site	This study
UP.ComRStH+I.ComRSve s.Fw	TTGATATAAAGGAGATTCTCTT GAGCATAAAAGACAGCATTG	Construction of strain LMD-9 with <i>comR_{Sve}</i>	This study
DN.ComRStH.Fw	AAATGGTGGTGACATAAATG		This study
UP.ComRStH.Rv	GAGAATCTCCTTTATATCAAAA C		This study
End.ComRSves+DN.Com RStH.Rev	CATTTATGTCACCACCATTTTA CAAACCATCTGCCGC		This study
Primers used to generate mutations in <i>comR</i>			
RV.StHComR left 2a	TGGAAACTTAATCAAACGATTC	For exchange of $\alpha 7(+L6)_{Sve}$ in <i>comR_{StH}</i>	This study
F.StHComR left 2a+Svst.start	GAATCGTTTGATTAAGTTCCA ACGTACGGAGATAAGGAGAG		This study
RV Svst 2a end+StH right	GTTAATAATTCTTCTTCTGGAAG AATATCAAAGAATTGATTATA CACATC		This study
FW.StH right 2a	CCAGAAGAAGAATTATTA ACTTAG		This study
RV.StH ComR left 7a	AATTCCTGATTAATACTCTC	For exchange of $\alpha 12_{Sve}$ in <i>comR_{StH}</i>	This study
FW StH left 7a+Svst start	GAGTATTAATCAGGAAATTC AACGGATGATGCTTATAATAT TG		This study
RV Svst 7a end+StH right	GTTAAGAAGTCTGAGAAGGACT CAAGAGAAATCTTAACC		This study
FW StH right 7a	TCCTTCTCAGACTTCTTAACAG		This study
Rv.8aStH+Inic 9aSvest	GACACCTGGTTTATAACTATG AAATTGTGATTTTCTATGACA GCTAAGG	For exchange of $\alpha 14_{Sve}$ in <i>comR_{StH}</i>	This study
Fw.Inic 9aSvest	TTTCATAGTTATAAACCAGGTG TC		This study
RVComRStH9aSvtEnd+St hRig	CTGCAGCCTCTTCTTGTTC CCCA TTATGTATAAGCTCATACTTA		This study
Fw.9a-end.StH.Ssa	AACAAGAAAGAGGCTGCAG		This study

Fw.End10aSth+I.CAPSves t	TTCCGTTTTGGAAGACTCCGTT TTGGAAGAAAAGAC	For exchange of CAP _{Sve} in <i>comR</i> _{Sth}	This study
Rv.Post-End10aSth	GAGTCTTCCAAAACGGAA		This study
ComRsthR92G-P94K.Fw	CAACGTACgGtAACaaaGACAGA ATAAAGTCTAAAC	To generate <i>comR</i> _{R92G-P94K} mutant in LL10 strain and pBADcomR _{R92G-} P94K-strep	This study
ComRsthR92G-P94K.Rev	CTGTCtttGTTaCcGTACGTTGGAA ACTTAATC		This study
Fw.ComRsth-R92G	CAACGTACgGtAACCCCTGACAGA ATAAAGTCTAAAC	To generate <i>comR</i> _{R92G} mutant in strains LL14 and LL19	This study
Rv.ComRsth-R92G	CTGTCAGGGTTaCcGTACGTTGG AACTTAATC		This study
ComRsthV201A- V205A.Fw	GATGcTTACAATATTGcTTTATT TAATGATTTGATG	To generate <i>comR</i> _{V201A-V205A} in strains LL11 and LL16, and pBADcomR _{V201A-V205A-strep} and pBADcomR _{R92G-P94K-V201A-V205A-strep}	This study
ComRsthV201A- V205A.Rev	CATTAAATAAAgCAATATTGTA AgCATCATCTGTCCAAATTC		This study
Fw.ComRStH-V205A	GATGTTTACAATATTGcTTTATT TAATGATTTGATG	To generate <i>comR</i> _{V205A} in strains LL15 and LL19	This study
Rv.ComRStH-V205A	CATTAAATAAAgCAATATTGTA ATCATCATCTGTCC		This study
Fw.ComRStH-S248G	CAAGCCTgGTGTTTTTGTACTTA AG	To generate <i>comR</i> _{S248G} in strains LL12, LL17 and LL19, and pBADcomR _{S248G-strep} and pBADcomR _{R92G-P94K-V201A-V205A-S248G-} strep	This study
Rv.ComRStH-S248G	CAAAAACACcAGGCTGTAGCT ATATAATTG		This study
Fw.ComRStH-S289K- I290T	GAGGAAAagActAAGGCAGGAAA ATTGGCAGATG	To generate <i>comR</i> _{S289K-I290T} in strains LL13, LL18 and LL19, and pBADcomR _{S289K-I290T-strep} and pBADcomR _{R92G-P94K-V201A-V205A-S248G-} S289K-I290T-strep	This study
Rv.ComRStH-S289K- I290T	GCCTT _{agTet} TTTCCTCTAAAACC GAGTCTTCC		This study
UpIntComRmut-Fw	TTCTGTTTTAGGAACGATTTTGC TTACAGTTGC	External primers to build LMD-9 derivative strains	(1)
DNIntComRmut_Rv	AAATCATCAATAATAGCAGTAT TGACCTGACTATTTGC		(1)
Fw.extComRSlocus.UP	GCTATTTTTGGAAGAAAAATCA AACAAGC	Confirmation of transformants	This study
Rv.extComRSlocus.DN	TCATGGAATTCACCTCAATTC TTGCTAAC		This study

Primers used in EMSA experiments

Cy3-	ATAGAAATGGTGGTGACATAAA	Probe Cy3-ComR box of P _{comS} from	This study
Fw.ComSboxSth.com.dire	TGTCACTATTTTTTTAG	<i>S. thermophilus</i> LMD-9	
ct			
Rv.ComSboxSth.com.dire	CTAAAAAAAAATAGTGACATTTA		This study
ct	TGTCACCACCATTCTAT		

958

959

^a Mutant codon introduced in primers is underlined and mutated nucleotides are in lower case.

960

Sequence from *comR*_{Sve} in the hybrid overlapping primers is indicated in bold.

961

Restriction sites are underlined.

962

StreptagII is indicated in italics

963

964

965 **Table S5: X-ray diffraction data processing and refinement statistics**
 966

Data processing statistics	Apo ComR _{Sve} (PDB ID 6HU8)	ComR _{Sve} ·XIP _{Sve} (PDB ID 6HUA)
Space group	C 2	P 21 21 21
Unit-cell parameters (Å)	a=110.07 b=38.29; c=87.62	a=60.18 b=84.03; c=152.75
Unit-cell angles (°)	α=90.0; β=113.2; γ=90.0	α=90.0; β=90.0; γ=90.0
Resolution range (Å)*	50.0-1.89 (2.01-1.89)	40.0-3.4 (3.59-3.39)
No. of unique reflections	26 936 (4195)	11 266 (1748)
Completeness (%)	99.1 (96.2)	99.7 (98.5)
Redundancy	6.7 (6.2)	71.0 (70.9)
Mean I/α(I)	12.58 (1.14)	15.16 (1.51)
R _{meas} (%) ^a	9.0 (135.6)	34.9 (338.1)
Refinement statistics		
Resolution range	36.81-1.89 (1.96-1.89)	38.88-3.39 (3.73-3.39)
No. of molecules/a.u.	1	1
R _{work} (%) ^b	20.91 (37.11)	27.1 (33.00)
R _{free} (%) ^c	24.3 (41.14)	34.9 (48.25)
Ramachandran		
Favored (%)	97.6	81.4
Outliers (%)	0.0	3.7
R.M.S.D.		
Bond lengths (Å)	0.008	0.015
Bond angles (°)	1.073	1.495
Chirality	0.042	0.066
Planarity	0.005	0.007
Dihedral	15.492	17.7
Average B, all atoms (Å ²)	45.75	106.55

967 * Numbers in parentheses represent values in the highest resolution shell.

968 ^a $R_{\text{meas}} = \sum_{\text{hkl}} \frac{1}{N-1} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$ where N is the multiplicity of a given reflection,
 969 $I_i(\text{hkl})$ is the integrated intensity of a given reflection and $\langle I(\text{hkl}) \rangle$ is the mean intensity of multiple
 970 corresponding symmetry-related reflections.

971 ^b $R_{\text{work}} = \sum ||\text{Fobs}| - |\text{Fcalc}|| / \sum |\text{Fobs}|$, where |Fobs| and |Fcalc| are the observed and calculated structure factor
 972 amplitudes respectively.

973 ^c R_{free} is the same as R_{work} but calculated with a 20% subset of all reflections that was never used in refinement.

974

975

976 **SI References**

977

- 978 1. Sambrook J, Fritsch EF, & Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*.
979 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- 980 2. Letort C & Juillard V (2001) Development of a minimal chemically-defined medium for
981 the exponential growth of *Streptococcus thermophilus*. *J Appl Microbiol* 91(6):1023-1029.
- 982 3. Fontaine L, *et al.* (2010) Development of a versatile procedure based on natural
983 transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*.
984 *Appl Environ Microbiol* 76(23):7870-7877.
- 985 4. Dower WJ, Miller JF, & Ragsdale CW (1988) High efficiency transformation of *E. coli* by
986 high voltage electroporation. *Nucleic Acids Res* 16(13):6127-6145.
- 987 5. Fontaine L, *et al.* (2013) Mechanism of competence activation by the ComRS signalling
988 system in streptococci. *Mol Microbiol* 87(6):1113-1132.
- 989 6. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred
990 kilobases. *Nat Methods* 6(5):343-345.
- 991 7. Fontaine L, *et al.* (2010) A novel pheromone quorum-sensing system controls the
992 development of natural competence in *Streptococcus thermophilus* and *Streptococcus*
993 *salivarius*. *J Bacteriol* 192(5):1444-1454.
- 994 8. Talagas A, *et al.* (2016) Structural Insights into Streptococcal Competence Regulation by
995 the Cell-to-Cell Communication System ComRS. *PLoS Pathog* 12(12):e1005980.
- 996 9. Lindahl E, Azuara C, Koehl P, & Delarue M (2006) NOMAD-Ref: visualization,
997 deformation and refinement of macromolecular structures based on all-atom normal mode
998 analysis. *Nucleic Acids Res* 34(Web Server issue):W52-56.
- 999 10. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta*
1000 *Cryst D-Biol. Cryst* 60:2126-2132.
- 1001 11. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence
1002 alignments using Clustal Omega. *Mol Syst Biol* 7:539.
- 1003

1.4. Description du mécanisme d'activation chez les ComR de type I

1.4.1. Fixation du peptide

Malgré les substitutions observées dans la poche formée par le domaine TPR des protéines ComR de *S. thermophilus* et *S. vestibularis*, les peptides XIP_{Sth} (LPYFAGCL) et XIP_{Sve} (VPFFMIYY) s'y fixent de façon très similaire, en adoptant une conformation en hélice 3₁₀ (Figure 47 - A). La Proline conservée en position 2 de chaque peptide joue certainement un rôle important dans ce repliement. Le carboxylate C-terminal des peptides assure leur ancrage au fond de la poche de fixation, comme observé pour les autres RNPP, et mis à part le premier résidu situé en N-terminal qui pointe vers le solvant, les squelettes peptidiques des deux XIP se superposent assez bien avec un rmsd de 0.57Å sur l'ensemble du peptide (Figure 47 - B).

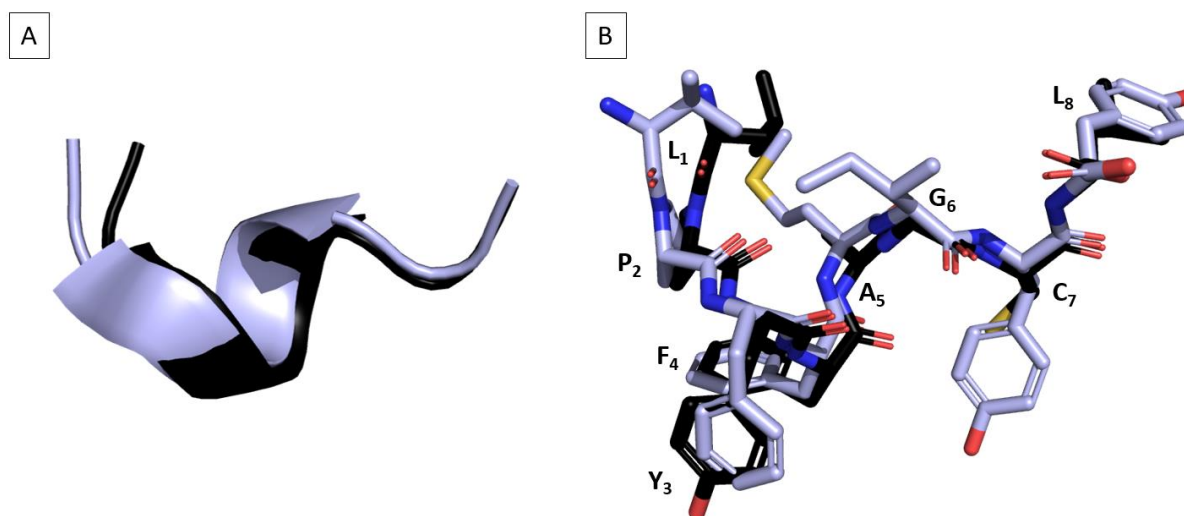


Figure 47 : Superposition des peptides de XIP_{Sve} VPFFMIYY (en bleu) et XIP_{Sth} LPYFAGCL (en noir).

A- En se fixant dans la poche formée par le domaine TPR, les deux peptides adoptent une conformation avec un tour d'hélice 3₁₀. B- Les deux peptides sont représentés en sticks et colorés par éléments. Le peptide XIP_{Sth} est labélisé. Le squelette peptidique des deux peptides XIP se superposent à l'exception du résidu en première position.

La poche de fixation n'est accessible au peptide que dans sa conformation ouverte, observée dans les structure de la forme apo de la protéine. Le peptide s'y lie probablement par le biais de faibles interactions de type van der Waals avec l'hélice α 12, qui fait également une liaison hydrogène non-spécifique entre l'asparagine conservée N208 et le squelette du peptide.

L'hélice $\alpha 14$ agit également comme un support de liaisons non spécifiques de type van der Waals pour le peptide. L'analyse par mutagenèse décrite dans l'article joint a également permis de montrer que les substitutions V/A-205 sur l'hélice $\alpha 12$ et S/G-248 sur l'hélice $\alpha 14$, qui compensent respectivement les substitutions L/Y-8 et C/Y-7 du peptide XIP (Nomenclature = résidu chez *thermophilus*/résidu chez *vestibularis* – numéro du résidu), jouent à ce stade un rôle essentiel dans la spécificité de reconnaissance du peptide. (Figure 48).

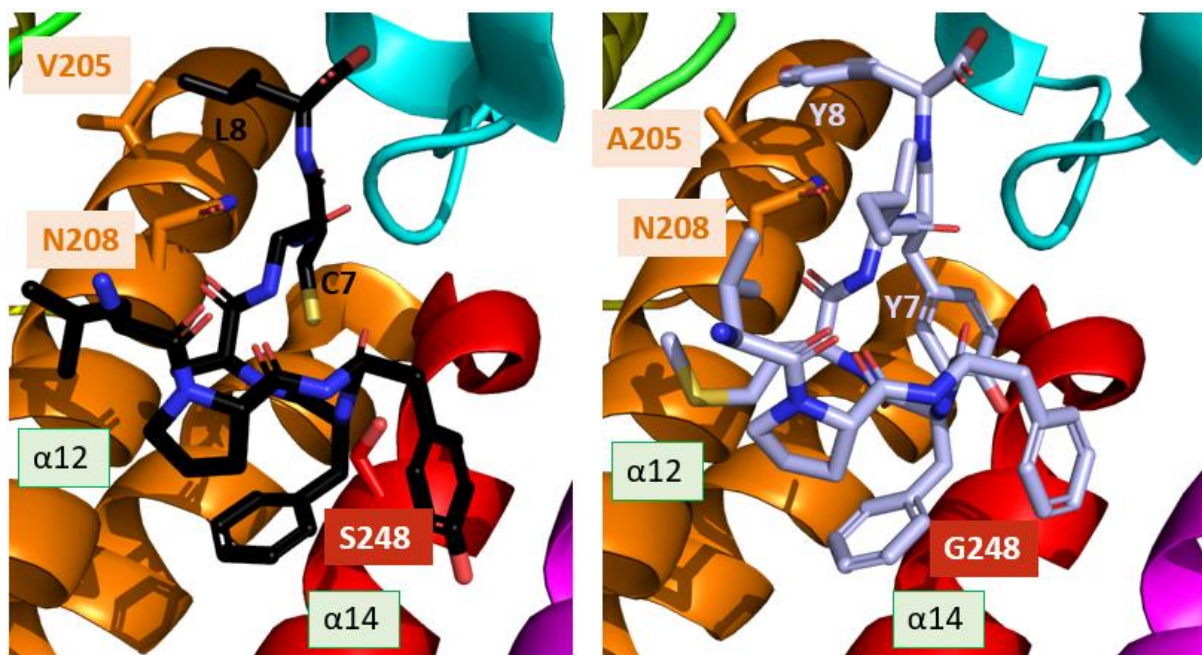


Figure 48 : Comparaison des sites de fixation du peptide XIP sur ComR_{Sth} et ComR_{Sve}

La protéine est représentée en cartoon et colorée dans un gradient de couleur. Les résidus de la protéine et du peptide sont labélisés de la même couleur que dans la structure.

Un point essentiel à l'ancrage du peptide dans la poche est la formation d'un pont salin entre son carboxylate C-terminal et le résidu K100 conservé sur l'hélice $\alpha 7$ D147 (TALAGAS *et al.* 2016) (Figure 49 - A et B). La bonne orientation du carboxylate est assurée par le positionnement de la chaîne latérale XIP-8 dans une poche hydrophobe formée par des résidus portés par les hélices $\alpha 8$, $\alpha 10$ et $\alpha 12$ (Figure 49 - C et D).

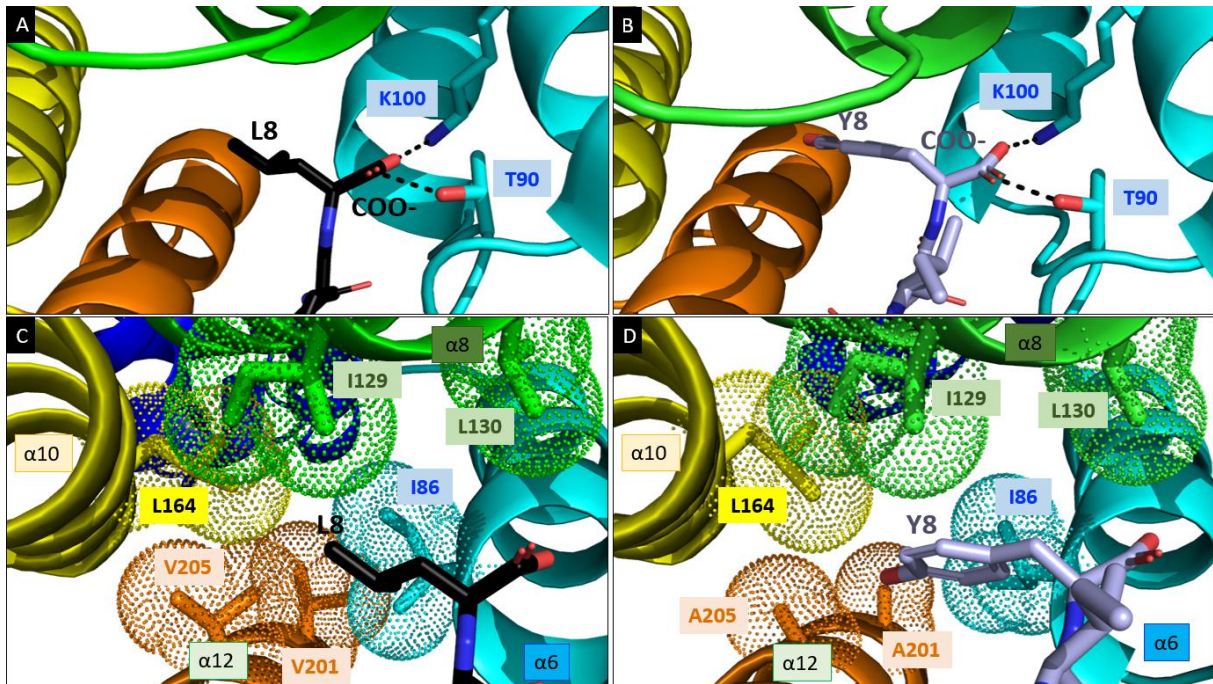


Figure 49 : Comparaison de l'ancrage des peptides XIP dans les protéines ComR de *S. thermophilus* (à gauche) et de *S. vestibularis* (à droite).

La protéine est représentée en cartoon suivant un gradient de couleur alors que le peptide est représenté en stick. Il est de couleur noir pour *S. thermophilus* et en couleur « lightblue » pour *S. vestibularis*. Les différents résidus mentionnés dans le texte, ainsi que les hélices sont également labélisés.

A et B- Zoom sur les ponts salins formés entre le COO- du peptide et les résidus T90 et K100. **C-** Le résidu C-terminal du peptide LPYFAGCL chez *S. thermophilus* vient se fixer dans une poche hydrophobe formée par les résidus avoisinants. Les résidus importants sont représentés en sticks et en dots. **D-** Le résidu C-terminal du peptide VPFMIYY chez *S. vestibularis* vient se fixer de la même manière que chez *S. thermophilus*. Les résidus sont représentés de la même manière.

1.4.2. Stabilisation de la conformation fermée de la poche de fixation

La stabilisation de la conformation fermée de forte affinité pour le peptide nécessite alors le recrutement de la boucle flexible $\alpha 6$ - $\alpha 7$, appelée boucle L6. Nous avons montré que l'interaction entre le résidu G/I-6 du peptide et le résidu R/G-92 de la boucle joue un rôle essentiel dans ce processus (Figure 50). Le changement de conformation entraîne une liaison hydrogène supplémentaire entre le carboxylate C-terminal du peptide et le résidu T90 de la boucle (Figure 49).

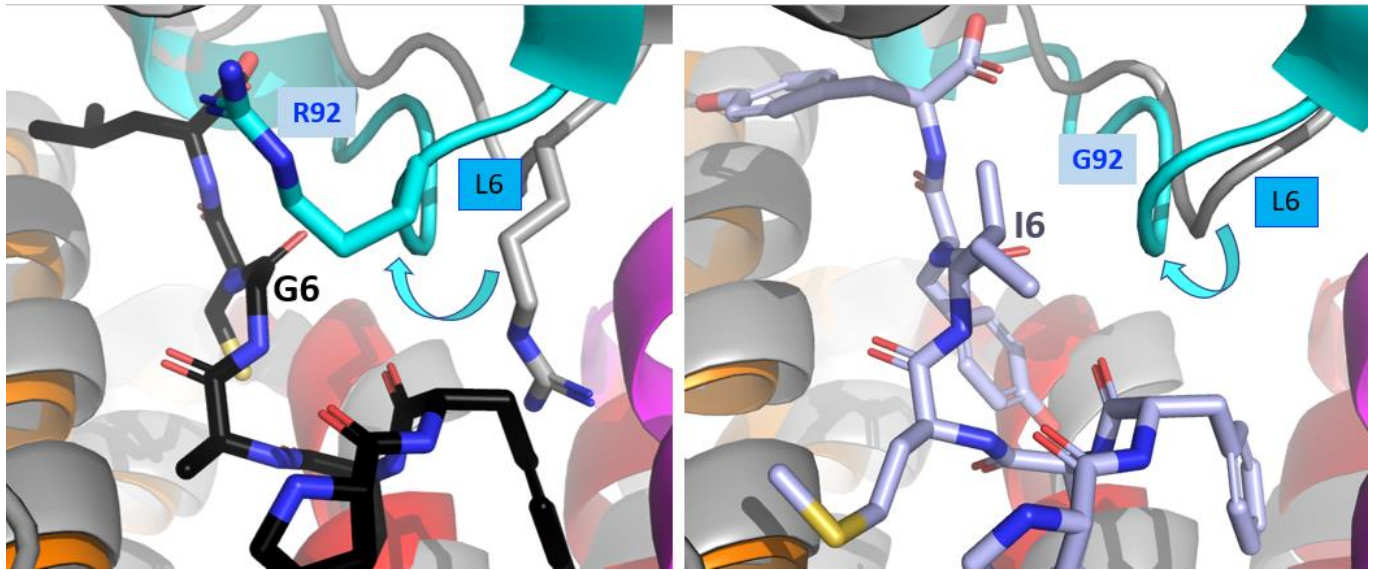


Figure 50 : Superposition des sites de fixation entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* à gauche et *S. vestibularis* à droite. Dans chaque cas, le recrutement de la boucle L6 est matérialisé par une flèche de couleur cyan.

Conjointement au recrutement de la boucle L6 par la partie C-terminale du peptide, le bon positionnement du résidu XIP Y/F-3 va permettre de recruter l'hélice CAP, via des interactions hydrophobes avec les résidus essentiels S/K-289 et I/T-290 (Figure 51 – A et B). Il en résulte un rapprochement des deux hélices CAP et $\alpha 7$, qui va permettre la formation d'un pont salin entre les résidus conservés R96 (de l'hélice $\alpha 7$) et D283 (de l'hélice CAP) (Figure 51 – C et D) ainsi qu'une interaction d'empilement entre les résidus conservés Y91 (de la boucle L6) et L286 (de l'hélice CAP) (Figure 51 – E et F). Ces deux interactions vont permettre de stabiliser le domaine TPR dans une conformation fermée de haute affinité pour le peptide.

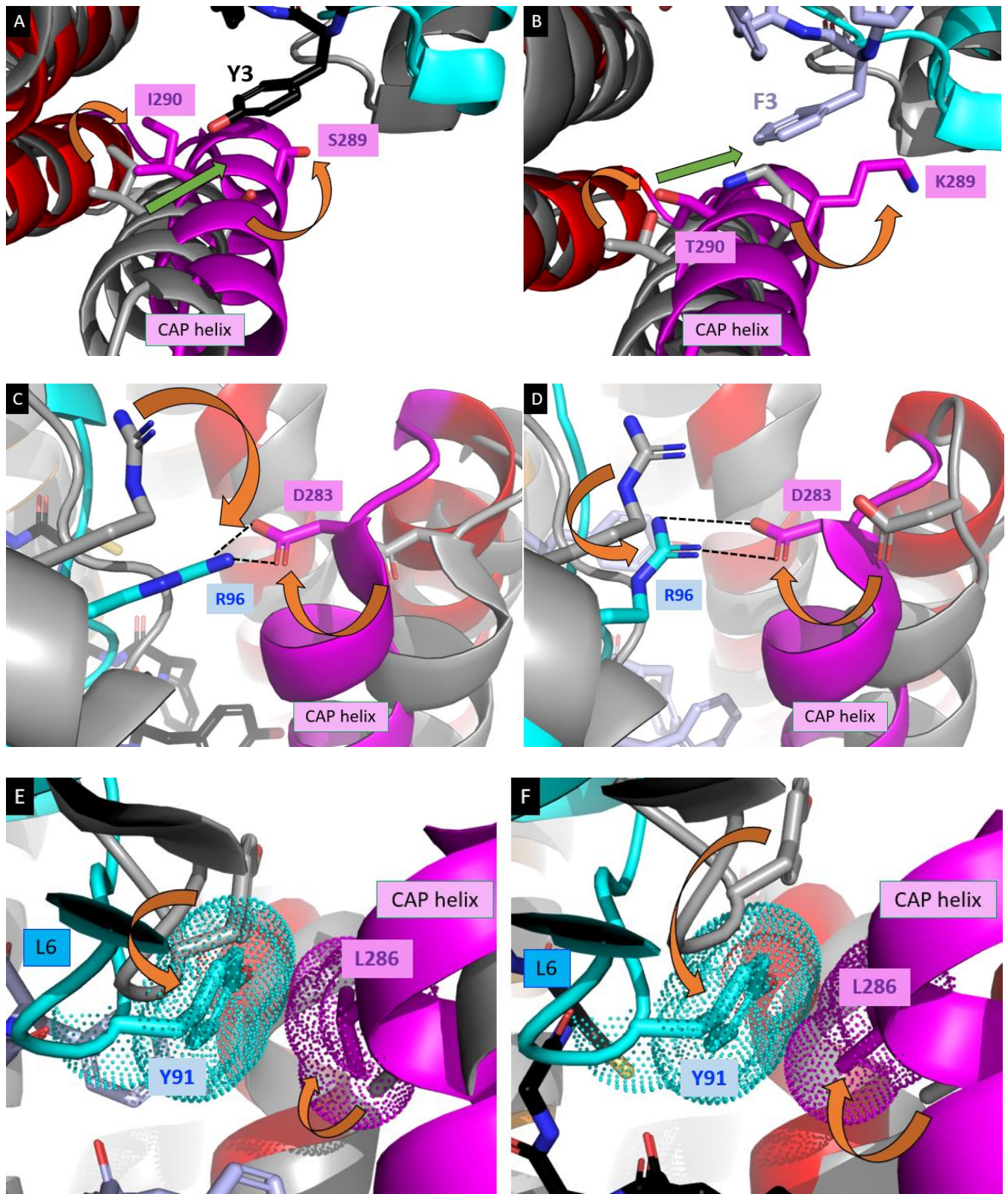


Figure 51 : Stabilisation de la forme active fermée.

Superposition entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* (A, C, E) et *S. vestibularis* (B, D, F). Les protéines sont représentées en cartoon avec les résidus importants labélisés et montrés en sticks. Des flèches orange représentent les mouvements des résidus importants induits par la fixation du peptide. Les flèches vertes dans les figures A et B représentent le mouvement de l'hélice CAP.

1.4.3. Relargage des HTH

Le recrutement de la boucle L6 entraîne un léger décalage des hélices $\alpha 6$ et $\alpha 7$ et par conséquent de l'hélice $\alpha 8$, notamment grâce au pont salin entre les résidus E119 (de l'hélice $\alpha 8$) et K82 (de l'hélice $\alpha 6$) (Figure 52).

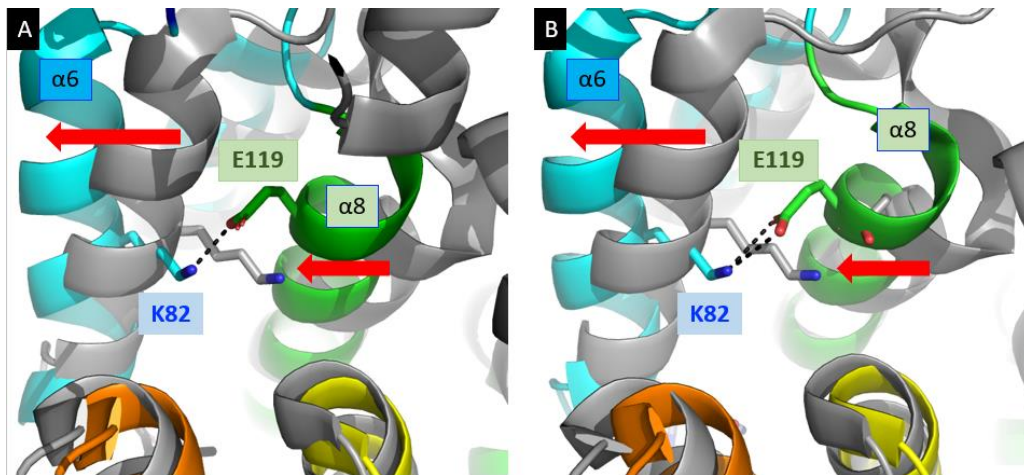


Figure 52 : Shift de l'hélice $\alpha 8$

Superposition entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* (A) et *S. vestibularis* (B). Les protéines sont représentées en cartoon et les résidus importants sont labélisés et montrés en sticks. Les flèches rouge représentent le déplacement de l'hélice $\alpha 6$ et par conséquent celle de l'hélice $\alpha 8$.

Le décalage de l'hélice $\alpha 8$ conduit à la perte d'un pont salin utile à la séquestration du domaine HTH sur le domaine TPR entre le résidu E117 (du domaine TPR) et R39 (du domaine HTH) ainsi que la perte des interactions hydrophobes avec la partie N-terminale de l'hélice $\alpha 9$ (Figure 53).

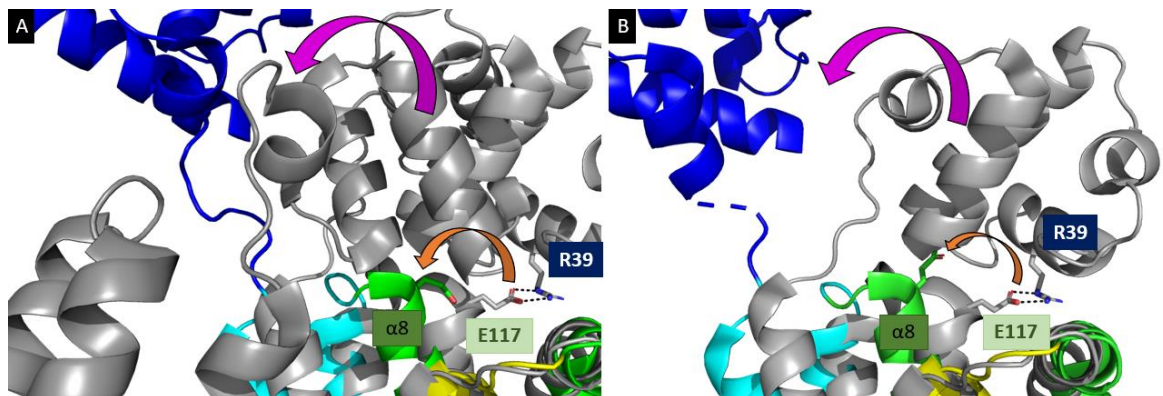


Figure 53 : Perte du pont salin R39 – E117

Superposition entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* (A) et *S. vestibularis* (B). Les protéines sont représentées en cartoon et les résidus importants sont labélisés et montrés en sticks. Les flèches oranges représentent le déplacement du résidu E117 lors de la fixation du peptide et la flèche violette représente le mouvement du domaine HTH.

D'autre part, les interactions hydrophobes entre la partie N-terminale du peptide XIP et les résidus aromatiques F/Y-171 et Y174 de l'hélice α_{10} induisent une petite rotation de l'hélice α_{10} et un changement de conformation de la boucle α_{10} - α_{11} (Figure 54 – A et B) ce qui induit une perte d'empilement entre les résidus H172 (de l'hélice α_{10}) et W135 (de la boucle α_{8} - α_{9}) (Figure 54 – C et D).

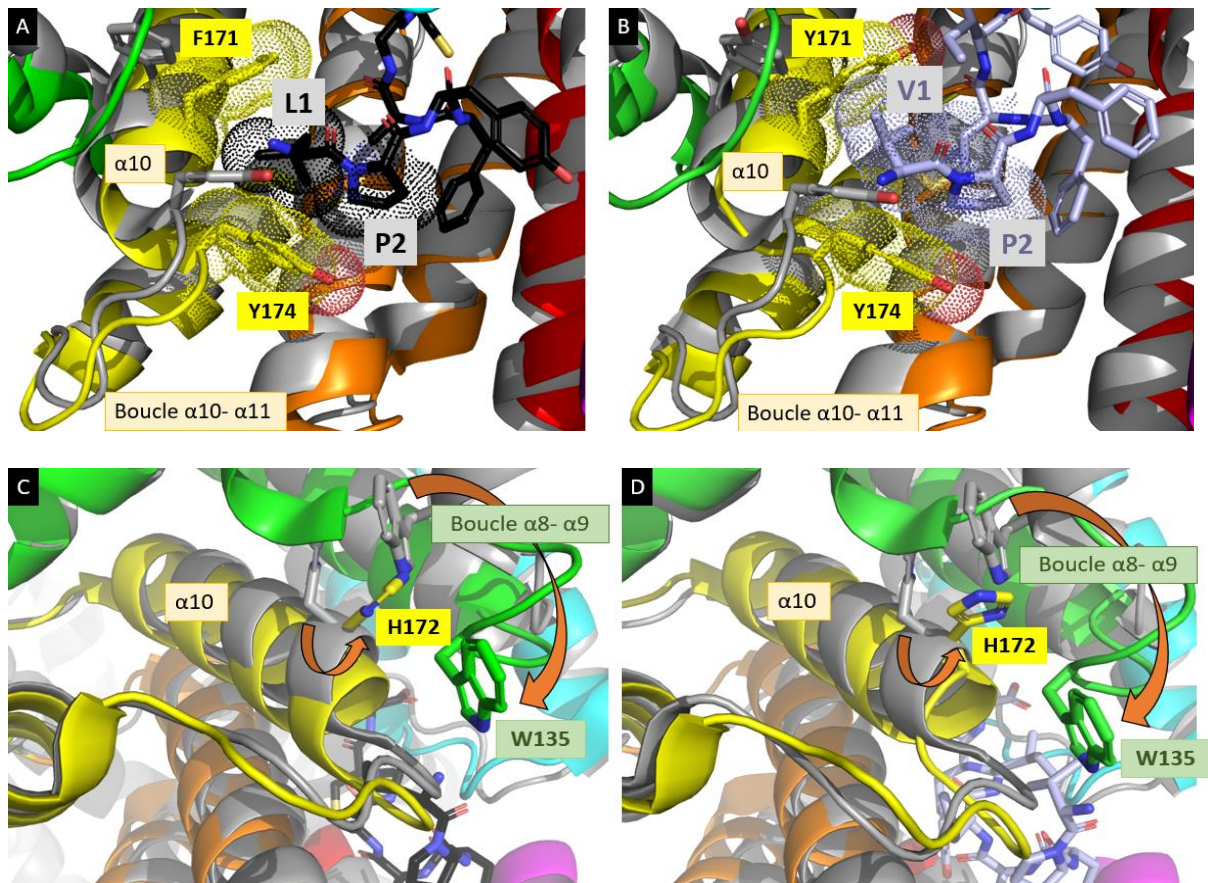


Figure 54 : Rôle de la partie N-terminale du peptide

Superposition entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* (A et C) et *S. vestibularis* (B et D). Les protéines sont représentées en cartoon et les résidus importants sont labélisés et montrés en sticks. Les flèches oranges représentent le déplacement des résidus lors de la fixation du peptide.

Ces deux changements provoquent la cassure de l'hélice α_9 au niveau du résidu D147 (TALAGAS *et al.* 2016) (Figure 55) entraînant la perte d'un deuxième pont salin entre les résidus E146 (du domaine TPR) et R35 (du domaine HTH) (Figure 55).

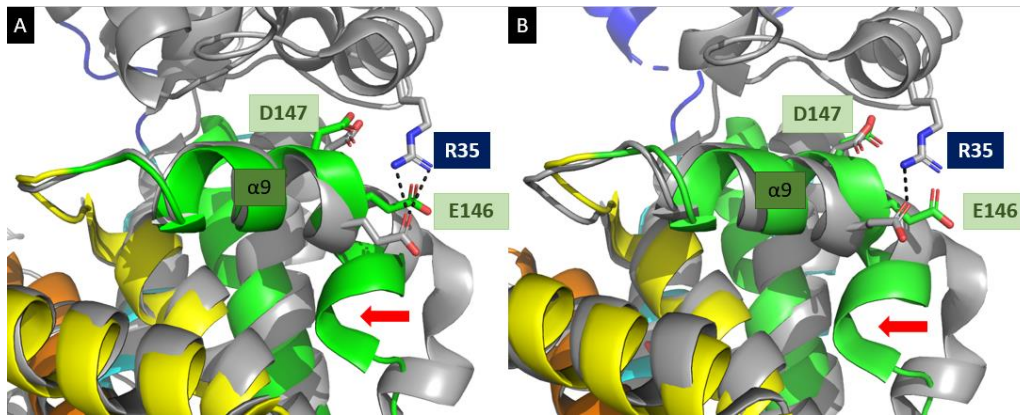


Figure 55 : Cassure de l'hélice α_9

Superposition entre les protéines apo (en gris) et les complexes avec le peptide XIP (en couleur) chez ComR *S. thermophilus* (A) et *S. vestibularis* (B). Pour chaque image, les protéines sont toujours représentés en cartoon et les résidus importants sont labélisés et montrés en sticks. Les flèches rouges représentent le déplacement de l'hélice α_9 lors de sa cassure au niveau du résidu D147

La perte des deux ponts salins entre les domaines HTH et TPR jouent très probablement un rôle essentiel dans le relargage des HTH.

1.4.4. Dimérisation

Le recrutement de l'hélice CAP entraîne un décalage de la boucle α_{15} -CAP portant le résidu conservé E282 qui va permettre de créer la dimérisation entre les deux protéines ComR. Chez *S. thermophilus*, la formation de deux ponts salins entre le résidu E282 et K246 (de l'hélice α_{14}) de la sous-unité voisine joue un rôle essentiel dans la dimérisation (TALAGAS *et al.* 2016) qui est nécessaire pour permettre aux domaines HTH libérés de se lier à l'ADN. Deux autres ponts salins sont formés entre K87 et D200 (Figure 56).

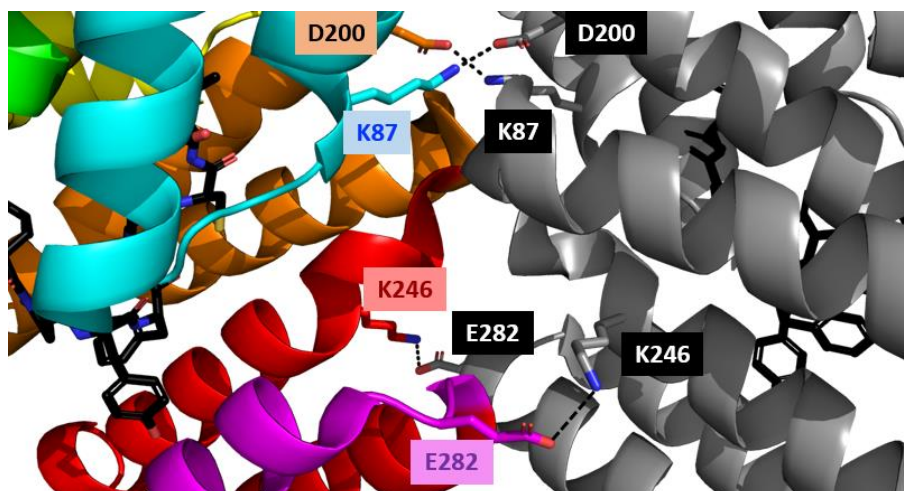


Figure 56 : Zoom sur la dimérisation de ComR_{Sth} en complexe avec XIP et l'ADN (Code PDB : 5JUB). Une des protéine est colorée suivant un gradient de couleur alors que la deuxième est représentée en gris. Les résidus impliqués dans les ponts salins sont labélisés et représentés en stick.

Chez *S. vestibularis*, bien que les résidus impliqués dans les ponts salins (E282, K246, K87, D200) soient conservés, seul les ponts salins entre K87 et D200 sont observés dans la structure du complexe binaire. Les ponts salins formés par les résidus E282 et K246 chez *S. thermophilus* ne le sont pas chez *S. vestibularis*, mais le résidu E282 fait une liaison hydrogène avec le résidu Y250 mieux placé (Figure 57).

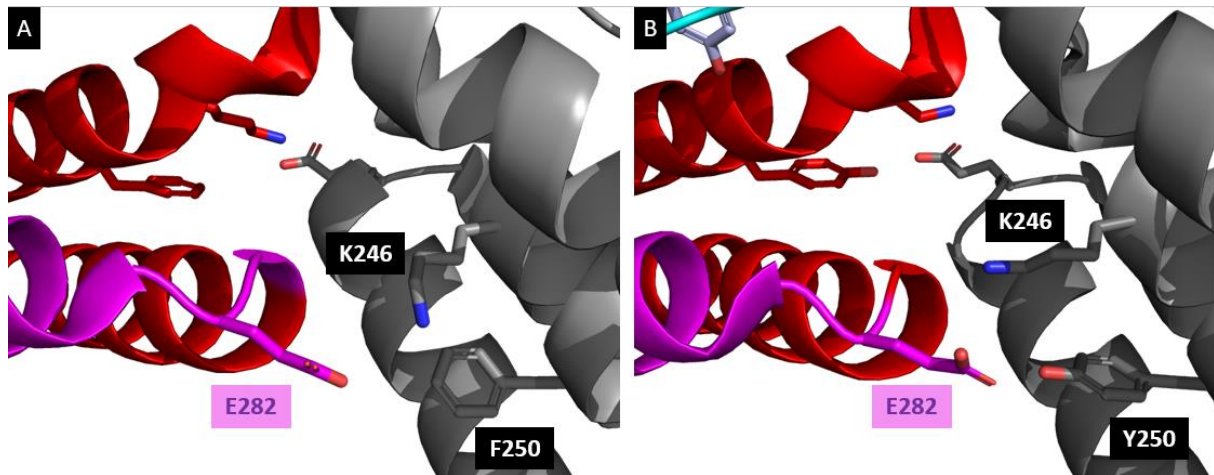


Figure 57 : Différence de dimérisation entre ComR_{Sth} et ComR_{Sve}

Les deux panels montrent la proximité des résidus 246, 250 et 282 chez *S. thermophilus* (A) et *S. vestibularis* (B). Alors que *S. thermophilus* forme un pont salin entre E282 et K246, *S. vestibularis* stabilise le dimère par liaison hydrogène entre E282 et Y250.

1.5. Vers la caractérisation de peptides inhibiteurs

L'étude précédente sur ComR_{Sth} (TALAGAS *et al.* 2016) avait permis d'identifier un peptide variant inactif, XIP_{Sth}-Y3A (LPAFAGCL), présentant une bonne affinité ($K_D = 32.6 \pm 5.4$ nM) avec la protéine (Figure 37).

J'ai analysé par DSC l'effet de sa fixation sur la stabilité de ComR_{Sth} (Figure 58) afin de comparer avec les résultats obtenus avec les peptides variants étudiés dans notre nouvelle étude.

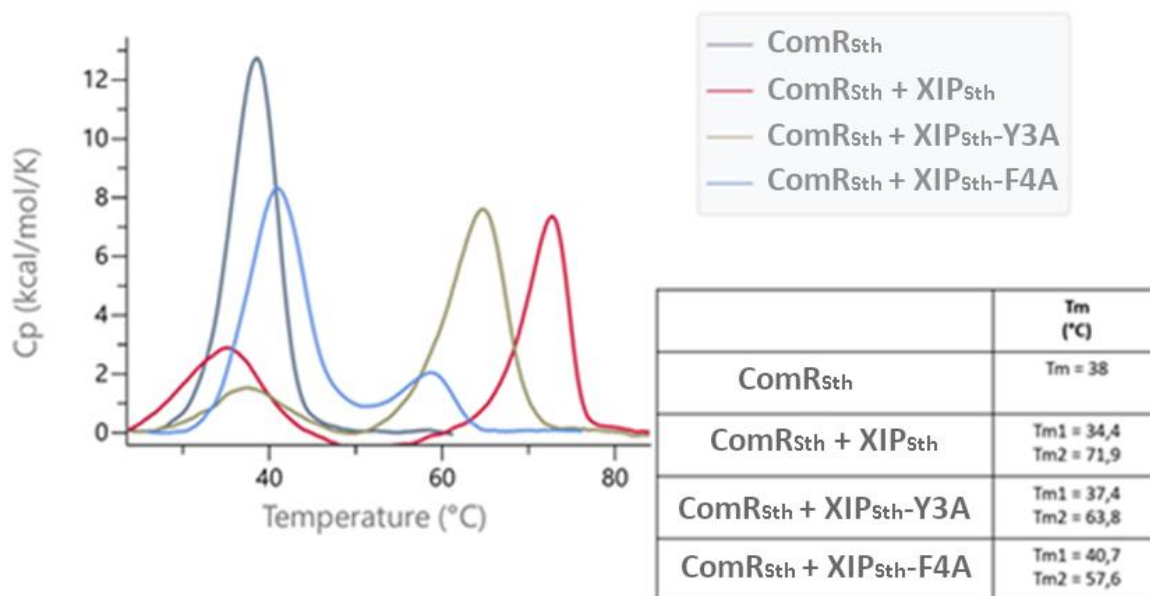


Figure 58 : DSC sur les complexes ComR_{sth}/XIP_{sth}-Y3A et ComR_{sth}/XIP_{sth}-F4A.

Les spectres de DSC ont été effectués sur une plage de température comprise entre 0 et 100°C avec une augmentation d'1°C par minute. La concentration de la protéine est toujours de 15µM et le peptide est toujours 10 fois en excès. Certains spectres possèdent plusieurs températures pouvant correspondre à différents états de la même protéine au sein d'un échantillon. Dans ce cas, les différents Tm sont mentionnés dans le tableau.

On voit que l'équilibre est majoritairement déplacé vers une forme intermédiaire (Tm = 63.8°C) moins stable que le dimère observé avec le peptide natif (LPYFAGCL) (Tm = 72°C). Pour contrôle, on voit qu'en présence du peptide inactif de faible affinité (K_D = 2.13µM), XIP_{sth}-F4A (LPYAAGCL), la protéine reste majoritairement sous une forme de Tm = 40.7°C correspondant à la forme apo monomérique. Seule une petite fraction de la protéine forme un complexe de Tm= 58°C dont on ne sait pas s'il correspond à une forme dimérique ou non.

Le shift de Tm induit par XIP_{sth}-Y3A (LPAFAGCL) est similaire à celui observé avec le peptide variant actif XIP_{Sve-v1} (VPFFMGYY) (Manuscrit - Figure S4-B). Ceci suggère que la fixation de XIP_{sth}-A3 induit sans doute la formation d'un dimère différent de la forme active.

Afin d'élucider le mode de fixation des trois peptides inactifs XIP_{sth}-Y3A (LPAFAGCL), XIP_{sth}-F4A (LPYAAGCL) et XIP_{sth}-L8A (LPYFAGCA), j'ai réalisé de nombreux essais de co-cristallisation avec la protéine ComR_{sth}. Avec le peptide le plus affiné, XIP_{sth}-A3, je n'ai obtenu que 2 cristaux dans une condition (0.1M HEPES pH7.5 / 1M Sodium Citrate) différente de la structure apo (Code PDB 5JUF). L'un d'eux a diffracté à 3.2Å de résolution, dans le groupe d'espace C222 avec comme paramètres de maille a = 63.57 ; b = 138.90 ; c = 76.73 ; α = 90 ; β = 90 ; γ = 90 quasiment identique aux paramètres de maille que l'on retrouve dans la

structure de la forme apo (C222 $a = 61.82$; $b = 141.68$; $c = 78.55$; $\alpha = 90$; $\beta = 90$; $\gamma = 90$). Avec les peptides de faible affinité, XIP_{Sth}-YA4 et XIP_{Sth}-L8A, j'ai obtenu de nombreux cristaux (Figure 59) dans des conditions souvent assez similaires entre elles mais différentes de celle de la forme apo (Code PDB : 5JUF).

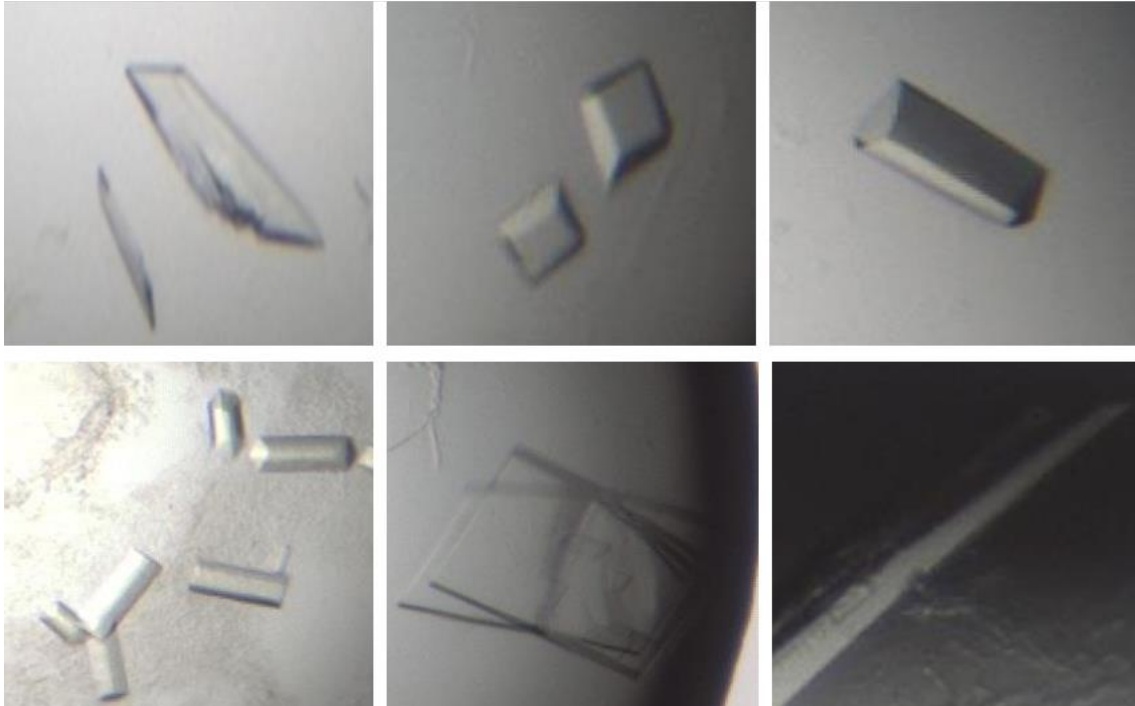


Figure 59 : Cristaux de ComRSth apo.

Les essais de cristallisation de ComRSth en présence des peptides inactifs XIP_{Sth}-Y3A, XIP_{Sth}-F4A ou XIP_{Sth}-L8A ont tous donnés des cristaux de la protéine seule malgré des formes et conditions différentes.

Une douzaine d'entre eux ont diffracté à moyenne-haute résolution ($3.5\text{\AA} > d > 2.3\text{\AA}$). La plupart diffractaient dans le même groupe d'espace et les mêmes paramètres de maille que la forme apo (Code PDB : 5JUF) (C222 ; $a = 61.8\text{\AA}$, $b = 141.7\text{\AA}$, $c = 78.5\text{\AA}$) et l'analyse des cartes de densités, après quelques cycles d'affinement, a confirmé l'absence de peptide fixé à la protéine qui était dans la conformation apo déjà décrite (TALAGAS *et al.* 2016).

1.6. Conformations alternatives de ComRSth apo

Un cristal obtenu en présence du peptide XIP_{Sth}-YA4 a cependant diffracté dans le groupe d'espace C2 avec 4 molécules par unité asymétrique (Tableau 3).

Data processing statistics	
Space group	C 2
Unit-cell parameters (Å)	a=271.79 b=73.72; c=78.31
Unit-cell angles (°)	$\alpha=90.0$; $\beta=104.97$; $\gamma=90.0$
Resolution range (Å)*	50.0-2.46 (2.55-2.46)
No. of unique reflections	54 104 (5195)
Completeness (%)	99.61 (96.60)
Redundancy	6.89 (6.42)
Mean I/ σ (I)	14.34 (2.54)
R _{meas} (%) ^a	10.9 (73.3)
Refinement statistics	
Resolution range	44.24-2.46 (2.55-2.46)
No. of molecules/a.u.	4
R _{work} (%) ^b	23.36 (34.76)
R _{free} (%) ^c	26.38 (36.35)
Ramachandran	
Favored (%)	97.14
Outliers (%)	0.17
R.M.S.D.	
Bond lengths (Å)	0.012
Bond angles (°)	1.27
Chirality	0.063
Planarity	0.007
Dihedral	13.612
Average B, all atoms (Å ²)	61.39

Tableau 3 : Statistiques des données cristallographiques de 6QER.

* Les nombres entre parenthèses représentent les valeurs dans la plus haute sphère de résolution.

(a) $R_{meas} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ où N est la multiplicité d'une réflexion particulière, $I_i(hkl)$ est l'intensité intégrée d'une réflexion particulière et $\langle I(hkl) \rangle$ est l'intensité moyennée de plusieurs réflexions équivalentes par symétrie. (b) $R_{work} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$, où $|F_{obs}|$ et $|F_{calc}|$ sont respectivement les amplitudes des facteurs de structure observés et calculés. (c) R_{free} est la même chose que R_{work} mais calculé sur une sous-partie du jeu de données (20%) qui n'est pas utilisée dans la procédure d'affinement de la structure.

La structure résolue par remplacement moléculaire à partir de la forme apo obtenue en C222 (Code PDB : 5JUF) (TALAGAS *et al.* 2016) a révélé après affinement, que 2 des 4 molécules présentaient des conformations distinctes du modèle initial au niveau du domaine TPR, et ce malgré l'absence de peptide lié.

La chaîne A présente en effet une nouvelle conformation de la boucle $\alpha 6$ - $\alpha 7$ alors que, dans la chaîne C, on remarque un décalage des deux hélices C-terminales CAP et $\alpha 15$ (Manuscrit - Figure S5). Les deux régions présentant des conformations alternatives sont celles qui participent au maintien de la conformation « fermée » active du domaine TPR. Cette nouvelle structure de ComR_{Sth} apo illustre donc la dynamique de la forme ouverte inactive du domaine

TPR en absence de peptide. On a vérifié par une analyse des contacts avec le serveur PISA (KRISINEL & HENRICK 2007) que les 4 chaines correspondaient bien à 4 monomères. J'ai déposé cette nouvelle structure de ComR_{StH} apo à la Protein Data Bank sous l'identifiant 6QER. Elle alimente la discussion sur le mécanisme moléculaire de ComR dans l'article ci-joint.

1.7. Perspectives

Parce que la plupart des espèces hautement pathogènes sont regroupées dans les protéines ComR de type II, j'ai également initié l'étude structurale de deux systèmes ComRS de ce type, *S. mutans* le pathogène responsable des caries dentaires (CLARKE 1924) et *S. pyogenes* une des bactéries les plus virulentes pouvant causer diverses infections mortelles qui entraînent la mort de plusieurs centaines de milliers de personnes par an (CARAPETIS *et al.* 2005). A ce jour, j'ai réussi à obtenir des cristaux avec ces deux protéines en complexe avec leur peptide XIP (Figure 60). Des jeux de données ont été collectés à moyenne résolution (environ 4.5 Å) mais, dans les deux cas, je n'ai pas réussi à résoudre le problème des phases par remplacement moléculaire. Une optimisation des conditions de cristallisation et un marquage à la sélénométhionine pour un phasage par diffusion anormale doivent donc être envisagés.

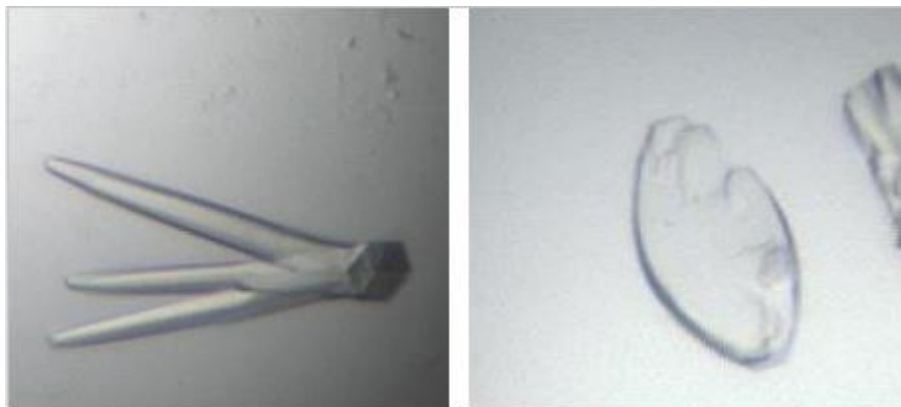


Figure 60 : Cristaux obtenus pour le complexe ComR/XIP chez *S. mutans* (à gauche) et *S. pyogenes* (à droite)

2. Etude structurale de deux paralogues de ComR chez *S. salivarius*

2.1. Introduction

Chez *S. salivarius* il a récemment été montré que ComR_{Ssa} contrôle non seulement la compétence mais également la biosynthèse de bactériocines qui sont des toxines peptidiques tuant d'autres bactéries (MIGNOLET *et al.* 2018). De manière intéressante, ce couplage entre compétence et prédation est régulé par deux paralogues de ComR, appelés ScuR (pour Salivarinicins Competence Uncoupling Regulator) et SarF (pour ScuR-associated Rgg factor) (MIGNOLET *et al.* 2019) (Figure 61).

Ce couplage entre compétence et prédation a particulièrement été étudié sur les systèmes QS indirects ComCDE et BlpCHR de *S. pneumoniae* (WHOLEY *et al.* 2016) mais existe également chez *S. mutans*, suggérant qu'il pourrait être présent chez tous les Streptocoques compétents (SHANKER *et al.* 2016).

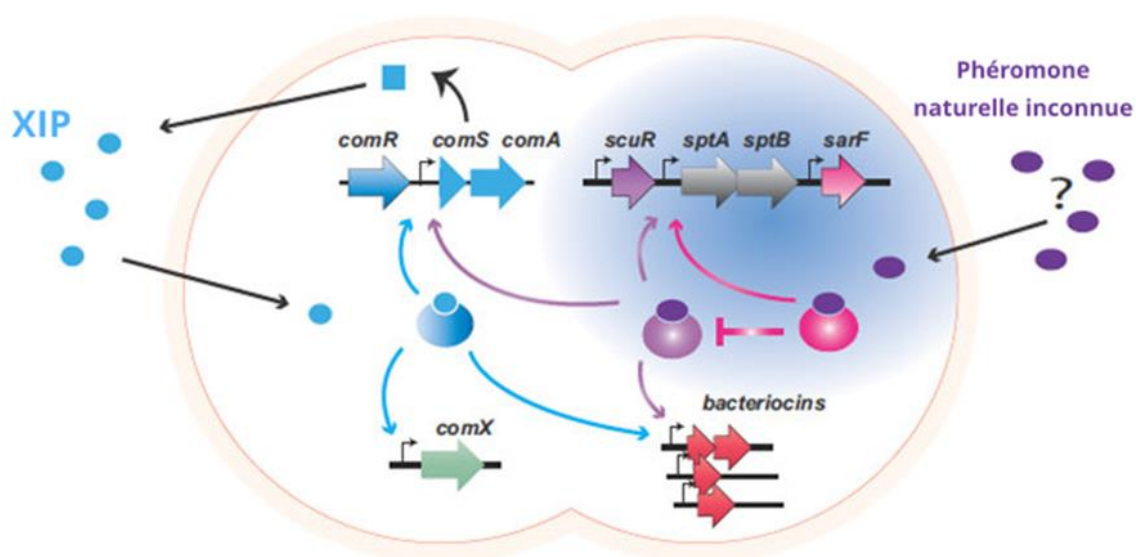


Figure 61 : Couplage entre compétence et prédation chez *S. salivarius* (MIGNOLET *et al.* 2019)

Les gènes *scuR* et *sarF* sont situés sur un locus différent des gènes *comR* et *comS*. Ils sont séparés par deux gènes qui pourraient coder pour deux sous-unités d'un transporteur de type ABC, appelées SptA et SptB (ScuR-promoted transporter A and B).

ScuR et SarF possèdent 82% d'identité de séquence entre elles et environ 41% d'identité de séquence avec ComR. Au niveau du domaine HTH, l'identité de séquence est encore plus forte puisqu'elle atteint 51% entre ComR et les deux paralogues et plus de 97% entre les deux paralogues (Figure 62).



Figure 62 : Alignement de séquence entre ComR_{Ssa} et ses paralogues ScuR et SarF.

Les positions des hélices sont indiquées par homologie avec la structure de ComR_{Sth}. Celles formant le domaine HTH sont représentées par un trait bleu au-dessus de l'alignement de séquences. Celles du domaine TPR sont indiquées par un gradient de couleur, du cyan au violet. Les résidus impliqués dans la séquestration du HTH sont encadrés en violet, ceux impliqués dans la dimérisation en orange, ceux impliqués dans la fixation du Cter du peptide en vert, les résidus qui maintiennent la conformation fermée du site actif sont encadrés en rouge et enfin les 5 résidus identifiés comme essentiels dans la spécificité de reconnaissance du peptide sont encadrés en bleu.

Les résidus impliqués dans la séquestration du domaine HTH sur le domaine TPR sont majoritairement conservés. On observe uniquement, chez ScuR comme chez SarF, deux substitutions conservatives (R35K et E118D), ce qui ne devrait pas empêcher la formation des ponts salins entre les deux domaines et suggère que le mécanisme de séquestration est conservé chez ScuR et SarF.

Les 4 résidus impliqués dans la formation des ponts salins stabilisant la dimérisation des domaines TPR (Figure 35) sont quant à eux beaucoup moins conservés. Dans les deux paralogues ScuR et SarF, la substitution E282G ne permet pas de faire un pont salin avec le résidu K246, qui est d'ailleurs substitué par une glutamine Q246 chez SarF. Il est donc probable que le mode de dimérisation de ComR ne soit pas strictement conservé chez ces deux paralogues.

Néanmoins, les résidus identifiés comme important pour la stabilité de la conformation fermée du site actif, R96 et D283, sont conservés dans ScuR et SarF, suggérant que ces deux paralogues pourraient donc potentiellement avoir un mécanisme d'activation semblable à ComR tout en ayant un mode de dimérisation différent.

La très forte conservation des domaines HTH suggère que les trois paralogues aient des séquences ADN cibles très proches. Pourtant, on observe une sélectivité de ces 3 protéines dans la reconnaissance de l'ADN. ComR_{Ssa} reconnaît les séquences promotrices des promoteurs PcomS et PcomX, mais aussi PsptA, le promoteur d'un putatif ABC transporteur, et PslvX, celui d'un gène contrôlant la production de bactériocines (Tableau 4), ScuR ne reconnaît pas PcomX, déconnectant ainsi les processus de prédation et de compétence. SarF quant à lui reconnaît uniquement PsptA.

Promoteur	Séquence	ComR S. salivarius	ScuR	SarF
PcomS	AATGGTGGTGACATAAATGTCACTAC	+	+	-
PcomX	TTTTATAGTGACATATATGTCTGCTA	+	-	-
PsptA	TAACGAGTCAAAGTGACATAGATGTCCTTTTGATTCGTTA	+	+	+
PslvX	CTCCATAGTGACATTTATGTCACTAT	+	+	-

Tableau 4 : Tableau récapitulatif des promoteurs reconnus par ComR et ses paralogues.

Un « + » indique que la protéine se fixe sur le promoteur.

La comparaison de séquences des trois paralogues révèle également que les 5 résidus qui sont identifiés comme essentiels dans la spécificité de reconnaissance du peptide chez ComR sont quasiment tous conservés entre SarF et ScuR, à l'exception de la substitution A248G, mais sont tous différents de ceux observés à ces positions chez ComR_{Ssa} (et ComR_{Sth}). Ceci peut expliquer pourquoi l'activité de ScuR et de SarF semble indépendante du peptide ComS (MIGNOLET *et al.* 2019).

Contrairement à ce que l'on peut observer pour ComR et les autres RNPPs, aucune séquence codant pour un potentiel peptide n'a pu être identifiée en amont ou en aval des gènes ScuR et SarF. Le criblage d'une banque d'ADN codant pour des peptides de 12 résidus a néanmoins permis d'identifier 22 peptides capables d'activer les deux régulateurs ScuR et SarF (MIGNOLET *et al.* 2019). Les 7 résidus C-terminaux, aléatoirement variés présentent une certaine ressemblance qui a permis de proposer une séquence consensus commune (Figure 63).

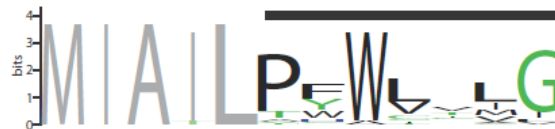


Figure 63 : Séquence consensus des 22 peptides activateurs de ScuR et SarF.

Les 5 premiers résidus marqués en gris sont les résidus non variables. Parmi les résidus variables, la taille d'un résidu est proportionnelle à son taux d'apparition à une position donnée.

Cette séquence consensus rappelle les peptides XIP de type I. L'octapeptide LPFWLILG, appelé BI7, très proche de la séquence consensus a été utilisé pour la caractérisation structurale et fonctionnelle de ScuR et SarF.

Il a été montré par polarisation de fluorescence (Figure 64-A) que le peptide synthétique BI7 (LPFWLILG) se fixe avec une forte affinité sur ScuR, avec un $EC_{50} = 16.05$ nM, du même ordre de grandeur que celle du complexe $ComR_{Ssa}/ComS_{Ssa}$ ($EC_{50} = 110.39$ nM). De manière surprenante, BI7 est également reconnu par $ComR_{Ssa}$ ($EC_{50} = 324.39$ nM) mais n'interagit que très faiblement avec SarF (MIGNOLET *et al.* 2019). Enfin, aucun des deux paralogues ne fixe le peptide $ComS_{Ssa}$ (LPYFTGCL) (Figure 64-B).

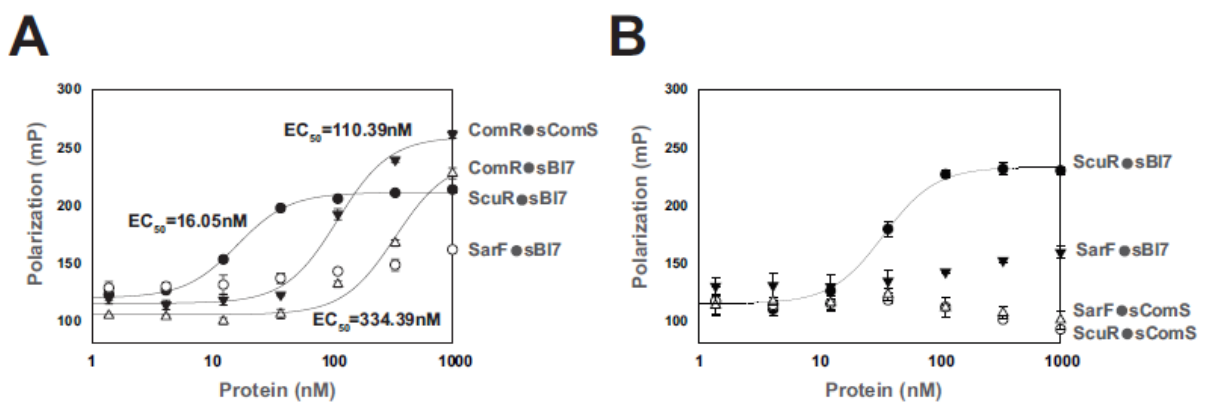


Figure 64 : Mesure d'interaction par polarisation de fluorescence.

A- Comparaison de la fixation du peptide synthétique BI7 sur les protéines ComR, ScuR et SarF avec l'interaction biologique entre $ComR_{Ssa}$ et $ComS_{Ssa}$. B- Comparaison de l'affinité de ScuR et SarF pour le peptide synthétique BI7 et le peptide naturel $ComS_{Ssa}$.

Le résidu 248 situé sur l'hélice $\alpha 14$ ayant été montré comme jouant un rôle particulièrement important dans la sélectivité du peptide chez ComR_{Sth}, la substitution A248G pourrait expliquer la différence d'affinité de ScuR et SarF pour le peptide BI7.

Afin d'élucider le fonctionnement de ScuR et SarF, et mieux comprendre la régulation du couplage entre compétence et prédation chez *S. salivarius*, j'ai donc initié, au cours de la deuxième partie de ma thèse, l'analyse structurale de ces deux paralogues de ComR.

2.2. Comparaison structurale des protéines ScuR et SarF

2.2.1. Purification des deux paralogues

A partir des plasmides de type pBAD fournis par nos collaborateurs, j'ai exprimé les 2 protéines recombinantes portant une étiquette Strep-tag au C-terminus dans une souche Top10 *d'E. coli*. J'ai ensuite adapté le protocole de purification déjà utilisé pour ComR_{Sth} et ComR_{Sve}.

L'étape de chromatographie d'exclusion de taille (SEC) sur colonne Superdex S200 16/60 a révélé une différence du volume d'élution (V_e) des deux protéines de MM identiques (37 kDa), suggérant deux états oligomériques distincts (Figure 65). Selon la calibration de la colonne, la protéine ScuR, élue à $V_e = 85.4$ mL, serait monomérique alors que SarF, élue à $V_e = 77.4$ mL, serait dimérique. On voit également sur le profil d'élution de ScuR (Figure 65-A) un pic minoritaire qui pourrait correspondre à une fraction dimérique de la protéine alors que sur celui de SarF (Figure 65-B) on voit un épaulement qui pourrait s'apparenter à un état d'oligomérisation encore supérieur de la protéine. On voit également dans le volume mort de la colonne que SarF semble s'agréger davantage que ScuR.

L'analyse des fractions d'élution par SDS-PAGE montre que les deux échantillons sont purs (Figure 66). A partir de 3.2 litres de culture, et en regroupant les fractions du pic majoritaire de chacun des deux chromatographies, j'ai obtenu, pour ScuR, un échantillon purifié de 500 μ L à 14 mg/mL, et pour SarF, un échantillon purifié de 500 μ L à 11 mg/mL.

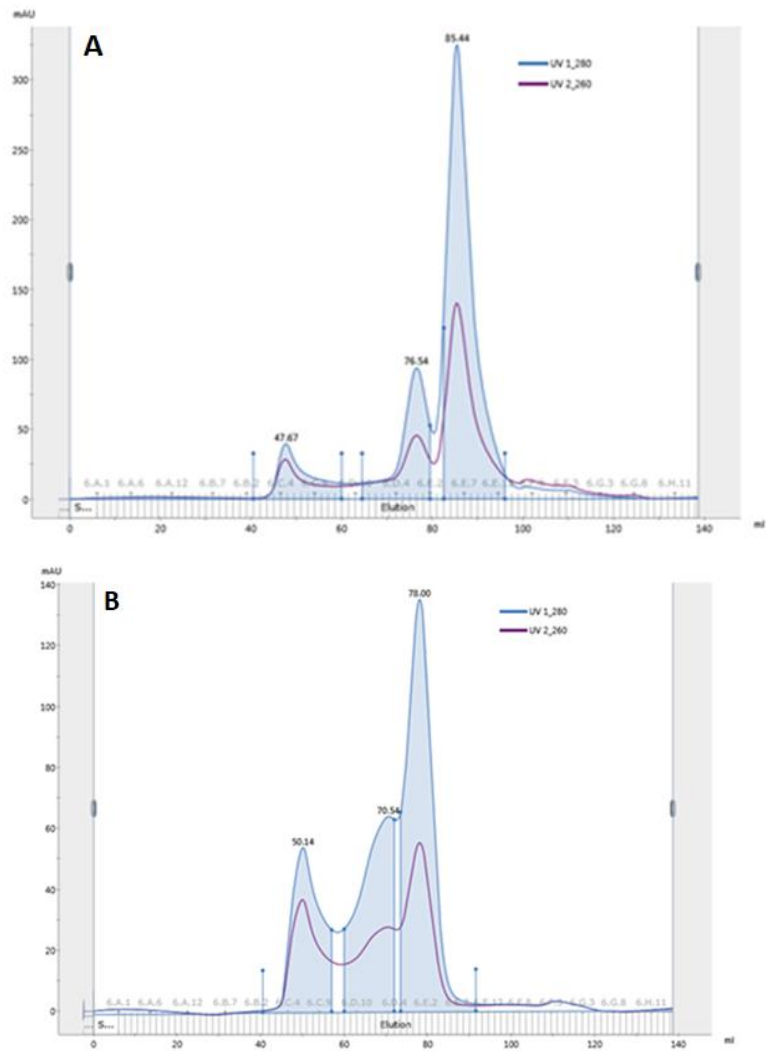


Figure 65 : Chromatographes en sortie de colonne Strep-Tactin.
Le chromatographe A représente la protéine ScuR et le chromatographe B la protéine SarF.

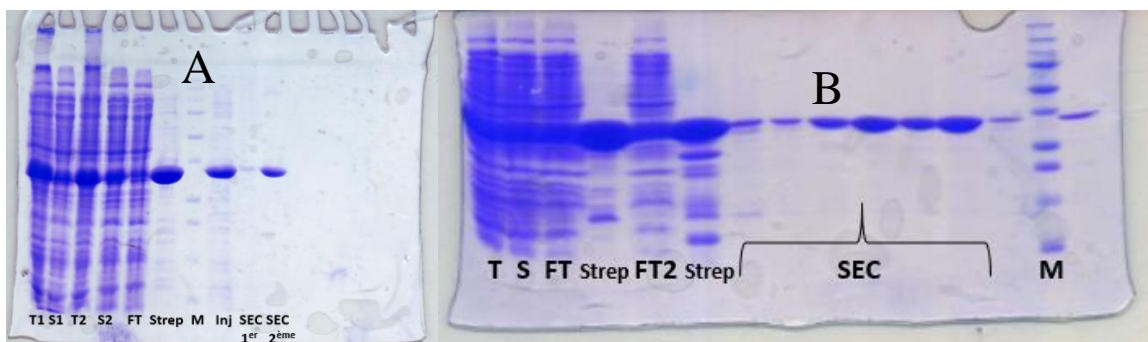


Figure 66 : Gels SDS-PAGE de contrôle de la pureté pour les protéines ScuR (A) et SarF (B).
Sur chaque gel sont représentés : la fraction totale (T) et soluble (S), le flow through (FT), une fraction du pic en sortie de colonne Strep-Tactin (Strep), en sortie de gel filtration (SEC) ainsi qu'un marqueur de taille (M)

2.2.2. Analyse DSC

J'ai montré que, sur ComR, la transition de la forme apo monomérique vers un complexe dimérique induite par la fixation du peptide XIP se traduit par une forte augmentation de la température de transition (T_m). J'ai donc testé si le même phénomène est observé avec la fixation du peptide synthétique BI7 sur ScuR et SarF (Figure 67).

En absence de peptide, la protéine ScuR, monomérique selon l'analyse par SEC, possède un T_m de 47.5°C, alors que Sarf présente un T_m de 57°C, pouvant refléter une stabilisation intrinsèque par dimérisation.

En présence de BI7, malgré des affinités très différentes observées par fluorescence (Figure 64), les deux protéines présentent un profil identique, avec un T_m de 62.5°C. Cette valeur est plus faible que celle observée avec les complexes ComR/XIP de $T_m = 75^\circ\text{C}$, mais le peptide utilisé n'est pas la phéromone naturelle de ces deux paralogues. Ce résultat suggère néanmoins que la fixation de BI7 entraîne la dimérisation de ScuR, de manière semblable à ce que l'on observe avec ComR. Pour SarF, déjà dimérique en absence de peptide, le shift de T_m induit par BI7, plus faible mais non négligeable (de 57 à 62°C), montre que, malgré la faible affinité observée en fluorescence, le peptide s'est bien fixé sur la protéine, entraînant un changement de conformation vers une forme plus stable (Figure 67).

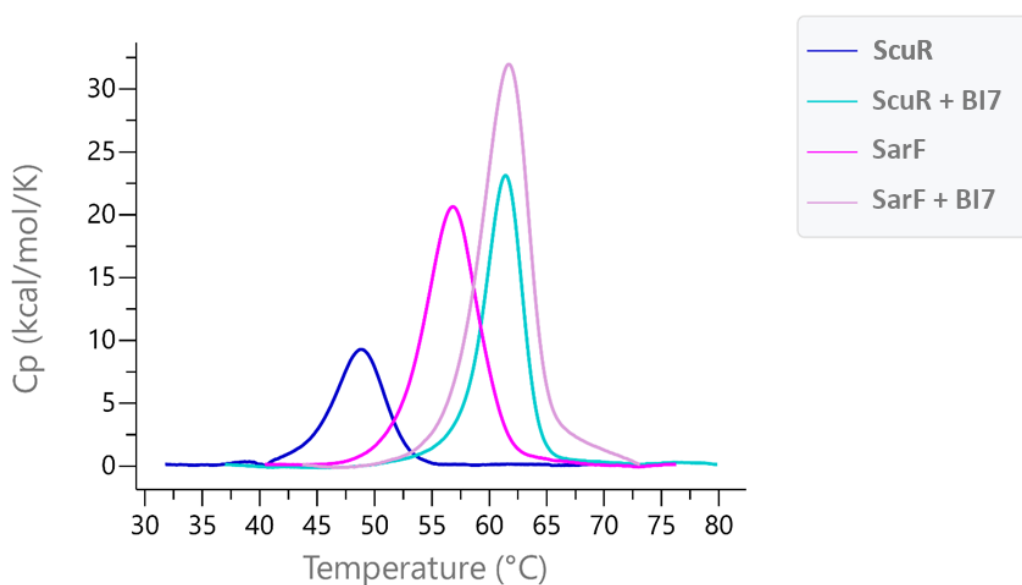


Figure 67 : DSC sur ScuR et SarF en absence et en présence de BI7.

Les spectres de DSC ont été effectués sur une plage de température comprise entre 0 et 100°C avec une augmentation d'1°C par minute. La concentration de la protéine est toujours de 15 μM et le peptide est toujours 10 fois en excès.

2.2.3. Analyse par SEC-MALS

Afin de confirmer l'effet du peptide sur l'état d'oligomérisation de ScuR et SarF, des expériences de Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) ont été réalisées en absence et en présence de BI7.

La forme apo de SarF présente un pic majoritaire correspondant à une masse molaire mesurée d'environ 69.5 kDa, confirmant que la protéine est majoritairement sous forme dimérique en solution (MM=37 kDa). Comme sur la colonne SEC (Figure 68), on retrouve un pic minoritaire pouvant correspondre à la présence d'une fraction de protéine sous forme trimérique (masse d'environ 104 kDa) ainsi qu'un pic correspondant à une petite fraction monomérique (MM = environ 36 kDa).

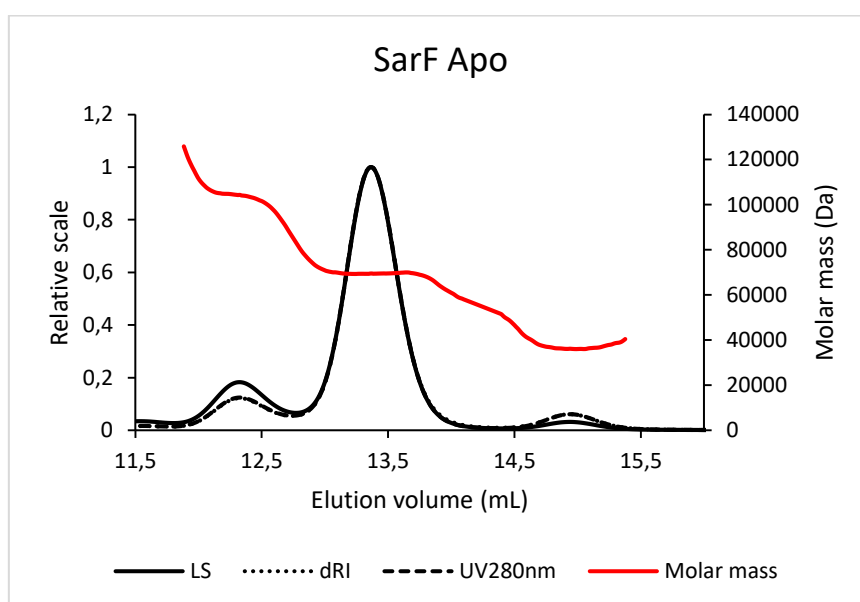


Figure 68 : Analyse par SEC-MALS de la protéine SarF apo

L'analyse du complexe SarF-BI7 donne un spectre très similaire à celui observé avec la protéine seule, avec une masse molaire mesurée d'environ 71.7 kDa pour la forme majoritaire (Figure 69). Néanmoins, la légère augmentation de la masse molaire correspondant à l'ajout de deux fois la masse du peptide (MM=958.21Da) ainsi que l'augmentation du coefficient d'extinction massique mesuré (1.026 mL/mg/cm contre 0.953 mL/mg/cm pour SarF apo) indiquent que le peptide BI7 s'est bien fixé sur SarF. Ce résultat démontre donc que l'ajout de BI7 ne change pas le degré d'oligomérisation de la protéine SarF. La hausse de T_m observée par DSC en présence de peptide (Figure 67) correspond donc à un changement de structure du dimère et non à un changement du degré d'oligomérisation.

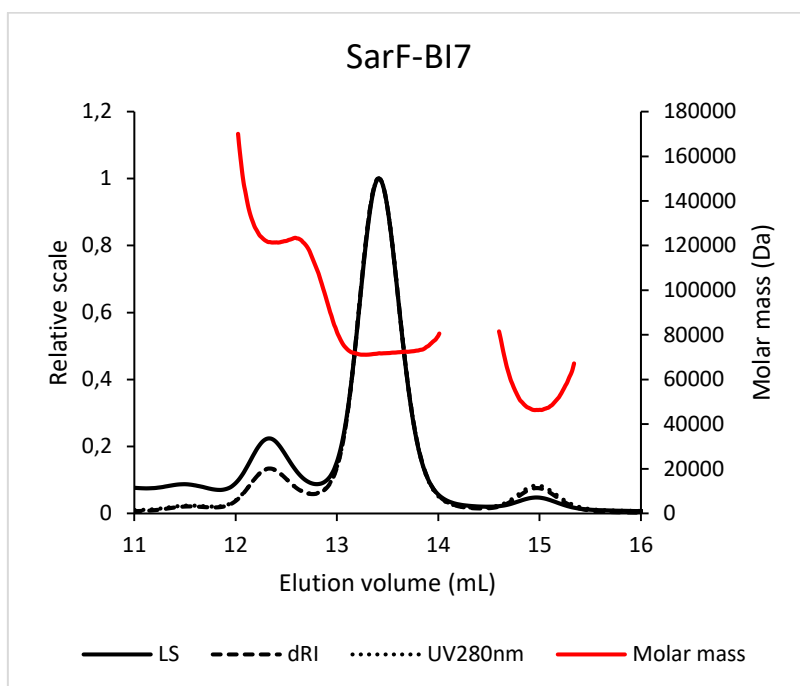


Figure 69 : Analyse par SEC-MALS de la protéine SarF-BI7

Comme déjà observé en SEC, ScuR apo présente un pic majoritaire correspondant à une protéine monomérique de masse molaire de 34.6 kDa (Figure 70). On constate également un pic minoritaire correspond à une espèce protéique de masse molaire mesurée de 71.2 kDa suggérant qu'une petite partie de l'échantillon est présente sous forme dimérique

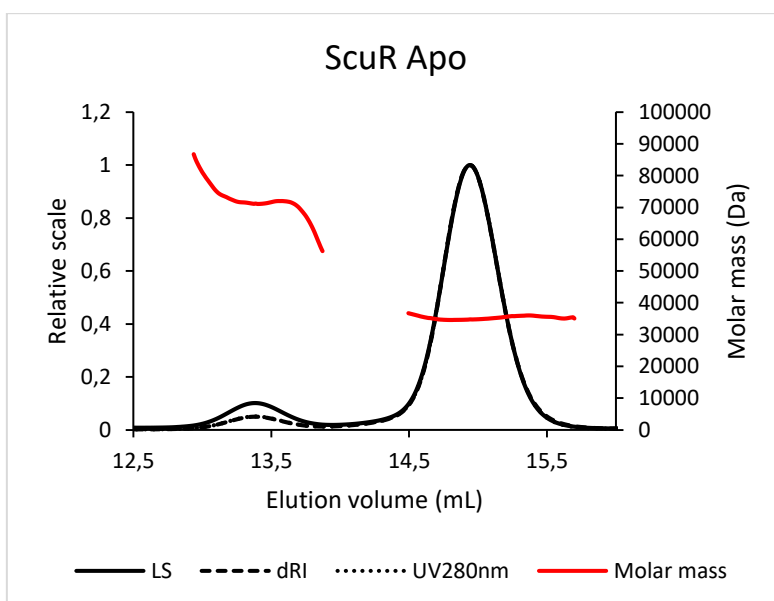


Figure 70 : Analyse par SEC-MALS de la protéine ScuR apo.

En présence de peptide BI7, le spectre de la protéine ScuR est plus homogène, avec un seul pic correspondant à une protéine de masse molaire de 66.4 kDa compatible avec la formation d'un dimère (Figure 71).

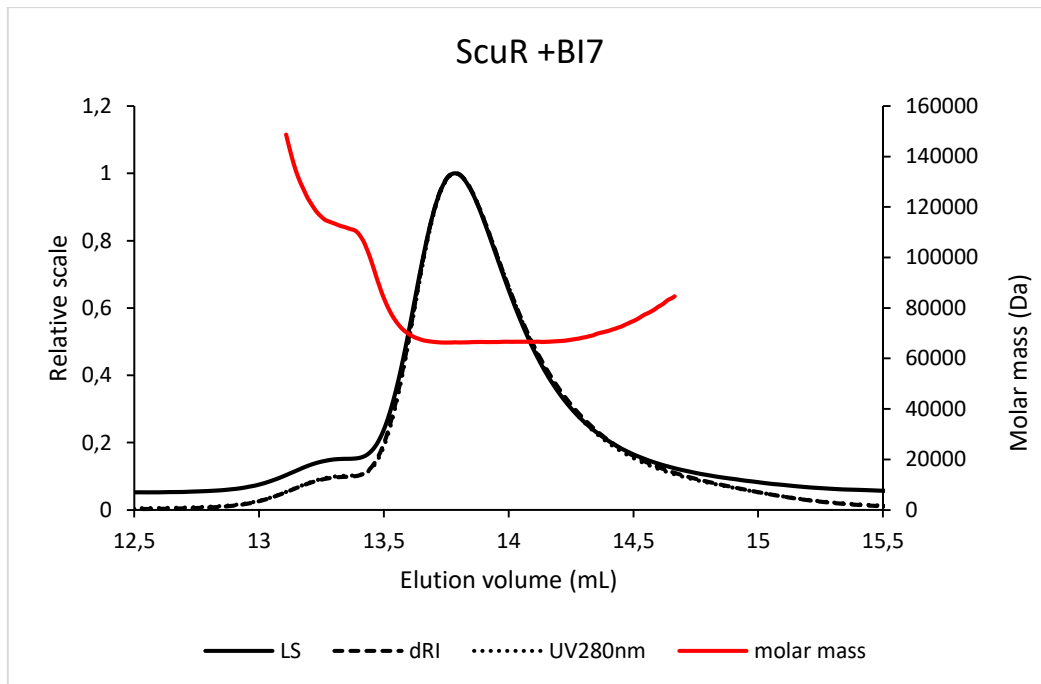


Figure 71 : Analyse par SEC-MALS de la protéine ScuR-BI7

En conclusion, ces résultats montrent que les deux protéines, bien que présentant 84% d'identité de séquence, ne se comportent pas de la même façon en solution. Alors que ScuR passe d'une forme apo monomérique à un complexe dimérique en présence de peptide, se comportant ainsi comme ComR +/- XIP en SEC-MALS (TALAGAS *et al.* 2016), SarF en revanche est dimérique en absence comme en présence de peptide.

2.2.4. Analyse par SAXS

Afin d'aller plus loin dans la caractérisation structurale de ScuR et SarF, j'ai effectué des mesures par SAXS sur ces deux protéines en absence et présence du peptide BI7. Après le passage dans la Superdex 200 Increase 10x300 (GE) et avant l'entrée dans la cellule SAXS, la DO_{280nm} de chaque échantillon est mesurée à l'aide d'un spectromètre UV. Une superposition des différents profils d'éluion est représentée dans la (Figure 72). En accord avec les expériences de SEC-MALS, ScuR apo est éluée après 29.5 minutes, tandis que le complexe ScuR-BI7 est élué à 26.8 minutes, corroborant la transition monomère-dimère quand le peptide

se fixe. La protéine SarF, avec ou sans peptide, est éluée principalement après 26.4 minutes, confirmant sa forme dimérique.

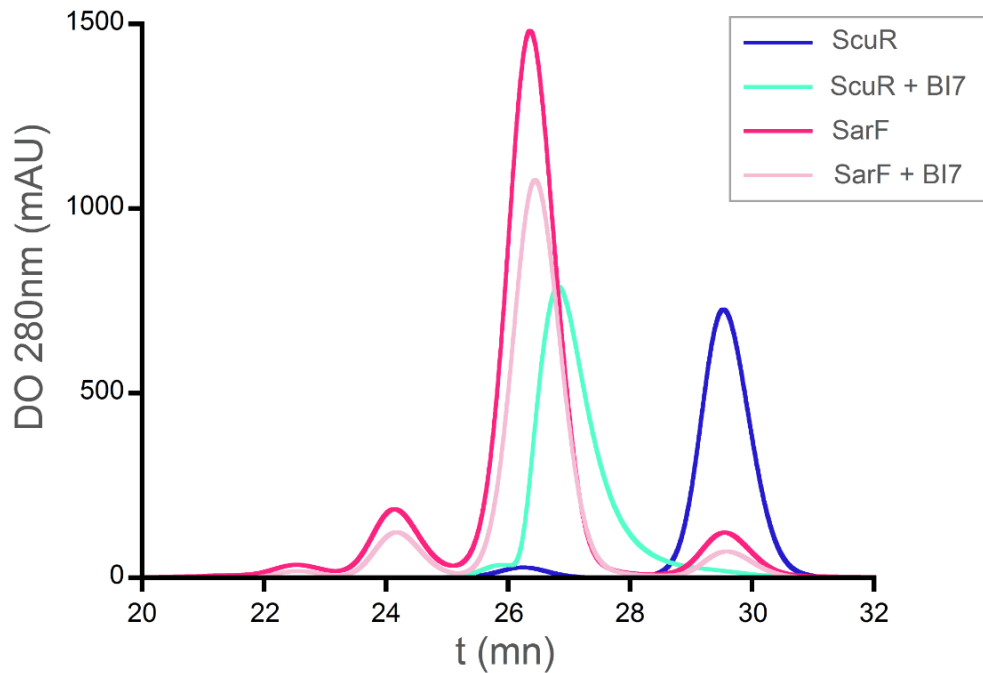


Figure 72 : Chromatogrammes des échantillons passés en SAXS en sortie de SEC.

La concentration initiale diffère selon l'échantillon. Alors que SarF apo a été injecté à 9.3mg/mL et SarF-BI7 à 8.8mg/mL, les concentrations de ScuR apo (4.2mg/mL) et ScuR-BI7 (5.8mg/mL) plus faible peuvent expliquer pourquoi l'absorbance est moindre pour cette protéine

2.2.4.1. Exploitation des données de diffusion

A partir des courbes de diffusion des rayons X obtenues toutes les 3 secondes tout au long de l'éluion (ce qui permet d'obtenir ~320 courbes par échantillon), il est possible pour chaque protéine d'obtenir, grâce au module SAXS du logiciel US-SOMO (BROOKES *et al.* 2016) l'évolution, au cours de l'éluion, du rayon de giration (R_g), de la valeur de l'intensité diffusée à l'origine $I(0)$ ainsi que celle de la masse moléculaire de l'objet diffusant. Les (Figures 73 - A à D) montrent les résultats ainsi obtenus pour les quatre échantillons étudiés, SarF apo, SarF-BI7, ScuR apo et ScuR-BI7. Cette analyse confirme les résultats obtenus par SEC-MALS : SarF apo et SarF-BI7 sont majoritairement en dimères, ScuR apo est presque exclusivement en monomère et ScuR-BI7 en dimère

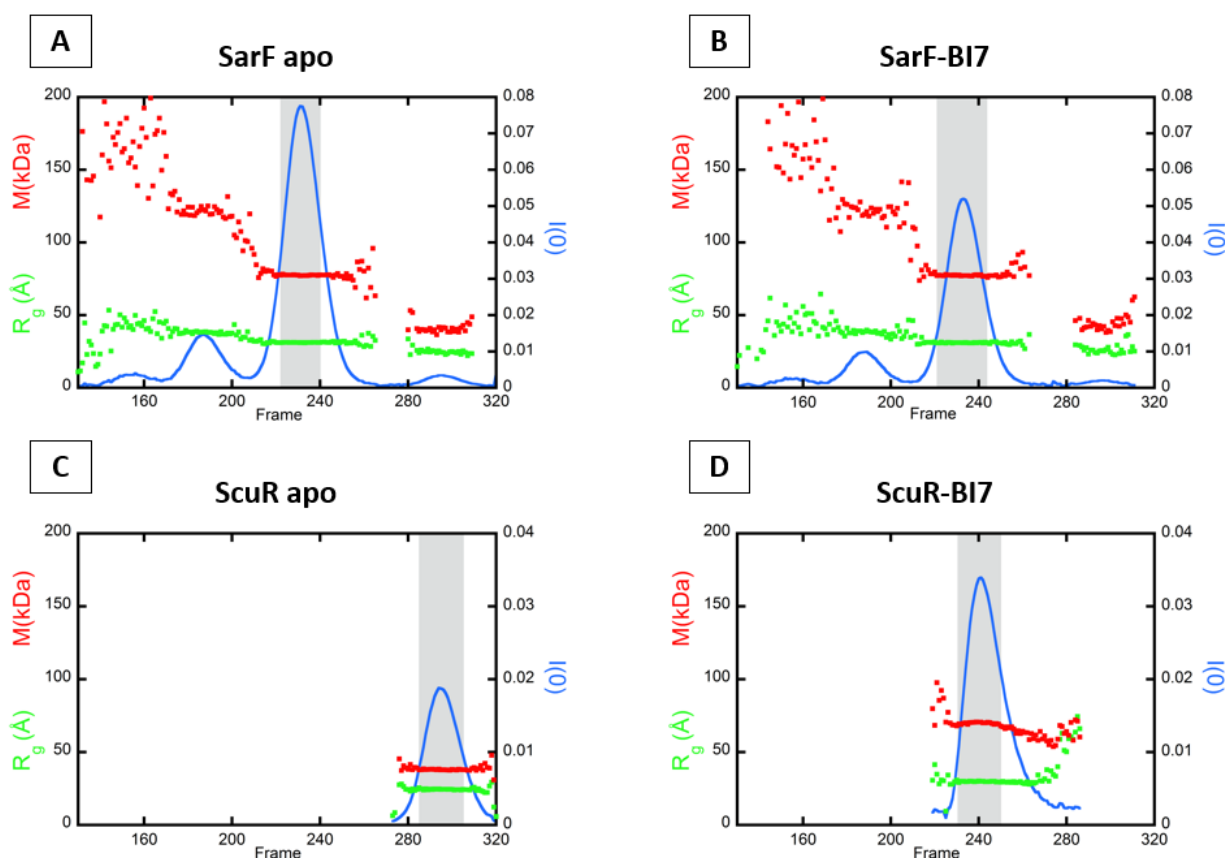


Figure 73 : Analyse par SAXS de ScuR et SarF avec et sans peptide BI7.

Evolution le long du pic d'éluion des différents paramètres extraits des courbes de SAXS obtenues toutes les 3 secondes. En abscisse, le numéro de la courbe de SAXS. En ordonnée : en vert, le rayon de giration (Å) ; en rouge, la masse molaire de l'objet diffusant (kDa) ; en bleu, l'intensité $I(0)$ diffusée en $q=0$.

Comme détaillé dans le paragraphe Matériel et Méthodes, la seconde étape consiste à moyenner les courbes identiques au voisinage du maximum d'éluion, dans la région grisée indiquée sur les différents panels de la (Figure 73).

L'analyse des quatre courbes moyennées ainsi obtenues, représentatives des quatre protéines étudiées, conduit aux paramètres R_g , $I(0)$, au diamètre maximum de la molécule D_{max} et la masse molaire de l'objet (Tableau 5).

	SarF apo	ScuR apo	SarF-BI7	ScuR-BI7
Data collection parameters				
Used frames	222-240	285-305	221-244	230-250
Guinier analysis				
R _g Guinier [Å]	31.2 ± 0.2	24.4 ± 0.2	31.1 ± 0.2	29.8 ± 0.2
qR _g -range	0.20-1.30	0.18-1.29	0.27-1.29	0.23-1.29
P(r) analysis				
R _g [Å]	31.2 ± 0.2	24.7 ± 0.2	31.0 ± 0.2	29.9 ± 0.2
D _{Max} [Å]	100 ± 5	85 ± 5	105 ± 5	105 ± 5
q-range [Å ⁻¹]	0.0064-0.35	0.0073-0.35	0.0087-0.35	0.0082-0.35
Molecular mass determination				
MM _{sequence} [kDa]	37.3	37.5	38.2	38.5
MM _{correlation volume} [kDa]	77.6 ± 7	38.0 ± 4	77.3 ± 7	69.8 ± 7
DADIMODO analysis				
q-range		0.0064-0.35		0.0068-0.35
Model number		10		10
χ ²		1.01-1.14		1.12-1.32

Tableau 5 : Tableau récapitulatif du résultat des analyses des courbes SAXS.

Ces quatre courbes peuvent être comparées visuellement si on divise l'intensité $I(q)$ par la concentration en protéine correspondante (Figure 74). On observe ainsi la transition monomère-dimère de ScuR quand la protéine fixe le peptide, puisque les quantités $I(0)/c$ (qui sont proportionnelles à la masse de l'objet diffusant) sont dans un rapport 2 (0.027 pour la forme apo et 0.054 pour le complexe ScuR-BI7). On observe aussi que les dimères SarF (en rose) et SarF-BI7 (en rose clair) donnent des courbes quasiment identiques. Par contre, les dimères ScuR-BI7 et SarF-BI7 conduisent à des courbes de SAXS $I(q)$ différentes.

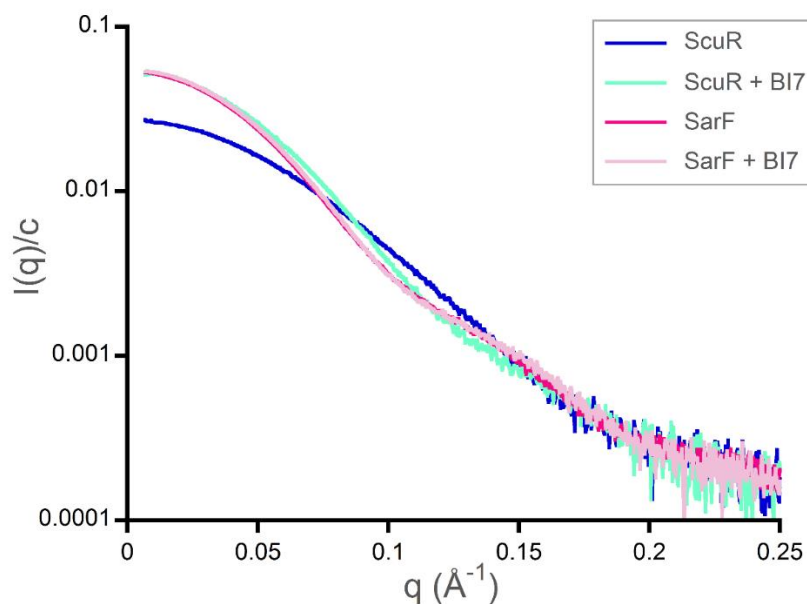


Figure 74 : Représentation logarithmique des courbes $I(q)/c$
 Les valeurs $I(q)$ sont normalisées par la concentration c correspondante.
 Dans ces conditions, $I(0)$ est directement proportionnelle à la masse de l'objet diffusant

Cela est bien visible si l'on compare pour ces deux complexes les fonctions de distribution de distances (Figure 75) ou si l'on représente les courbes de diffusion en représentation de Kratky (Figure 76), qui donne $(qR_g)^2 I(q)/I(0)$ en fonction de qR_g et qui permet de mieux mettre en évidence les différences.

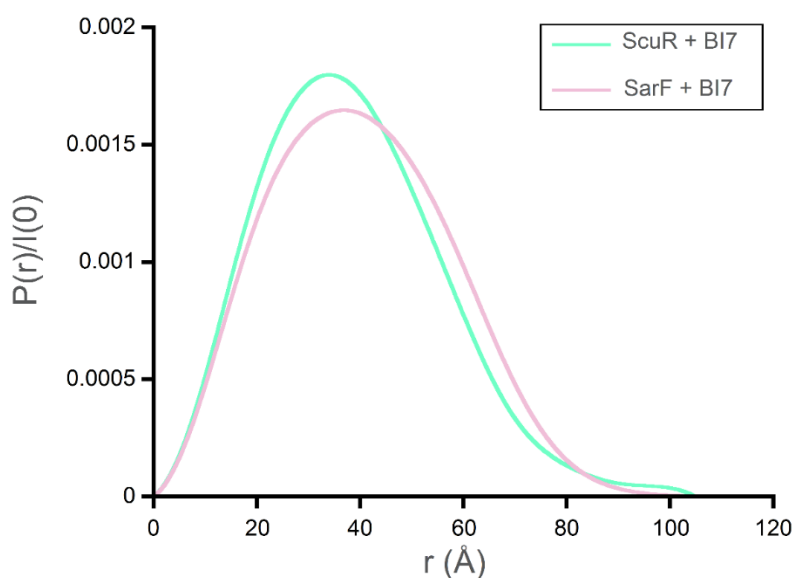


Figure 75 : Fonctions de distribution des distances r des complexes ScuR-BI7 et SarF-BI7
 Les courbes sont représentées sous la forme $P(r)/I(0)$, afin de s'affranchir de la concentration de l'échantillon.

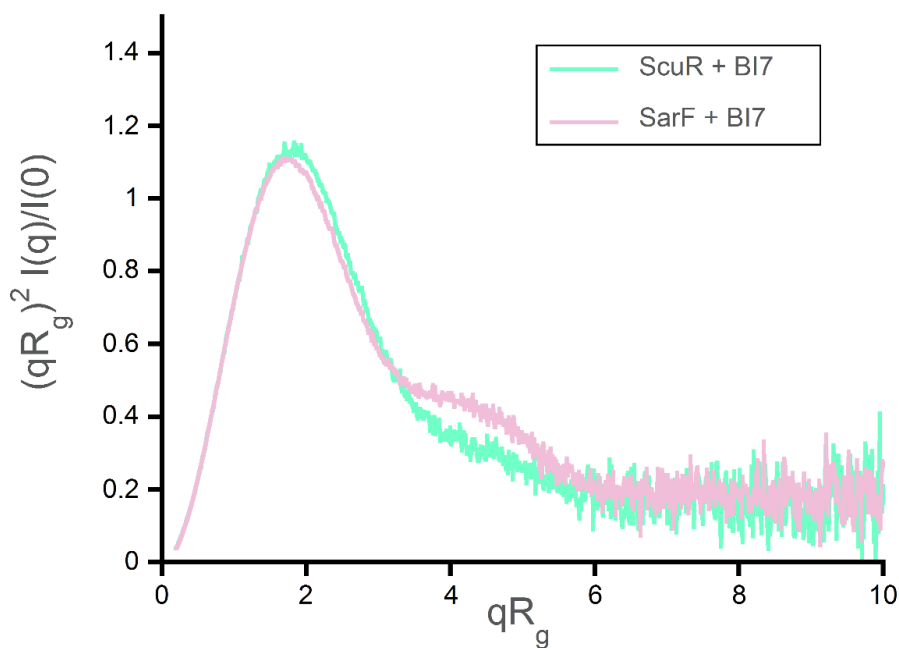


Figure 76 : Comparaison des complexes ScuR-BI7 et SarF-BI7, en représentation de Kratky. Cette représentation, sans dimension, $(qR_g)^2 I(q)/I(0)$ en fonction de qR_g , permet de mieux faire apparaître la différence entre les deux complexes.

La comparaison des deux distributions de distances $P(r)$ sur la courbe $P(r)/I(0) = f(r)$ (Figure 75) démontre qualitativement, du fait du décalage de la courbe vers de plus fortes valeurs de r , que le dimère SarF-BI7 est moins compact que le dimère ScuR-BI7

De plus, sur la représentation de Kratky (Figure 76), seule la courbe de SarF-BI7 présente un maximum secondaire vers $q=0.14-0.16 \text{ \AA}^{-1}$ caractéristique de la corrélation entre les deux monomères du dimère. Ceci indique que le mode de dimérisation n'est pas le même pour SarF-BI7 et pour ScuR-BI7.

L'étape suivante consiste à tenter de faire une modélisation tout-atome des protéines seules et des complexes, qui permette de rendre compte des courbes de diffusion.

2.2.4.2. Modélisation par homologie de ScuR

La protéine ScuR se comportant en solution de manière identique à ComR, un modèle par homologie de la forme apo de ScuR a été élaboré en utilisant le serveur SWISS-MODEL. Nous sommes partis de la structure cristalline de ComR_{Sth} (PDB ID 5JUF), mais un modèle similaire aurait été obtenu à partir de la structure de ComR_{Sve} (PDB ID 6HU8) étant donné le rmsd très faible entre les deux structures (0.8Å sur 298 atomes C α alignés).

Les extrémités N et C-terminales de ce modèle, qui comporte les résidus 5 à 308, ont été ajoutées à l'aide du logiciel MODELLER. Leur position a ensuite été affinée grâce au logiciel DADIMODO qui positionne les parties flexibles de façon à ajuster la courbe calculée à partir du modèle sur la courbe expérimentale. La (Figure 77 - A) montre 10 modèles ainsi obtenus, qui ne diffèrent que par l'extension C-terminale de l'étiquette Strep-tag, et qui donnent un excellent ajustement à la courbe expérimentale ($1.01 < \chi^2 < 1.14$) (Figure 77 - B).

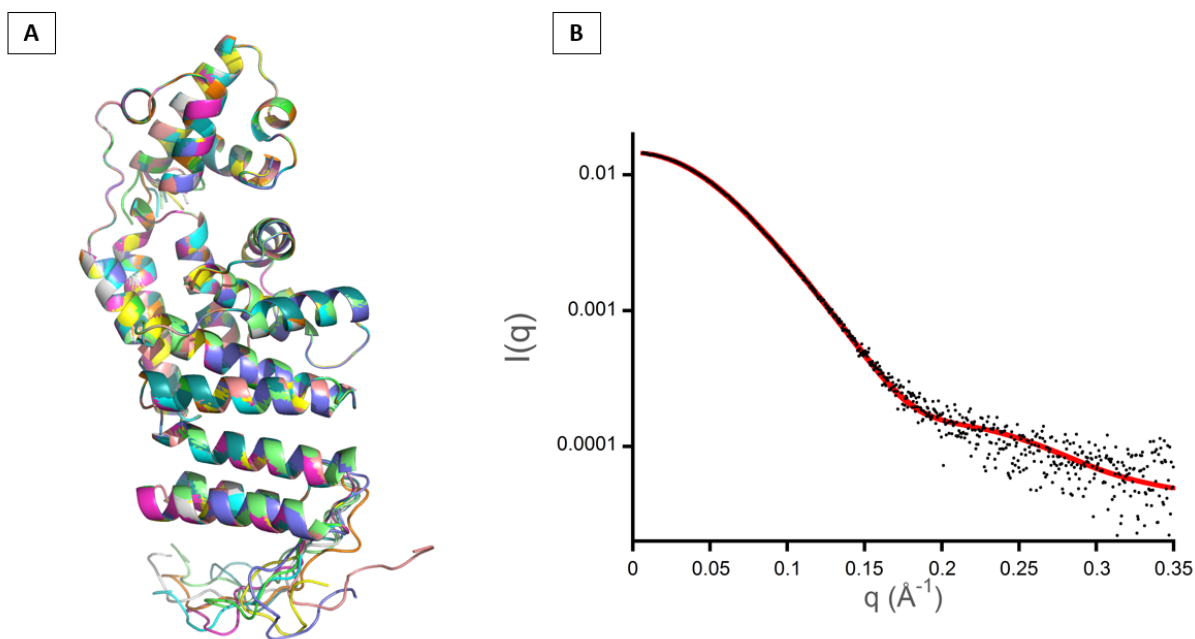


Figure 77 : Modélisation de ScuR apo

A- 10 modèles de ScuR apo, de structure identique à celle de ComR_{Sth} (PDB ID 5JUF) compatibles avec les résultats de SAXS. B- Comparaison de la courbe calculée (trait rouge) sur l'un des modèles, avec les données expérimentales (points noirs). Les autres modèles donnent des résultats identiques.

Ce résultat, couplé à l'analyse de séquence qui montrait majoritairement une conservation des résidus impliqués dans la séquestration du domaine HTH, montre que l'on peut donc faire

raisonnablement l'hypothèse que ScuR dans sa forme apo a une structure analogue à celle de ComR, avec des ponts salins qui maintiennent une interaction entre les domaines HTH et TPR.

En ce qui concerne le complexe binaire ScuR-BI7, la modélisation a été conduite de la même façon que pour ScuR apo, en commençant par la construction avec SWISS-MODEL d'un modèle de dimère de ScuR-BI7, à partir de la structure cristalline de ComR_{Sth}/XIP_{Sth}/ADN (PDB ID 5JUB). Ce modèle comporte deux sous-unités constituées des résidus 8-308, chacun avec leur peptide associé, auxquelles on ajoute avec MODELLER les extrémités manquantes. La courbe de SAXS calculée sur ce modèle de départ, pour lequel la position des domaines HTH est probablement imposée par l'ADN, donc probablement différente de celles du complexe sans ADN, est très éloignée de la courbe expérimentale. Néanmoins en symétrisant les deux domaines HTH, on arrive à diminuer l'écart entre courbe calculée et courbe expérimentale. Dans une seconde étape, la position des extrémités N- et C-terminales est affinée à l'aide du programme Dadimodo qui ajuste la courbe calculée à la courbe expérimentale, en laissant de plus les domaines HTH libres de se mouvoir par rapport aux domaines TPR, les linkers HTH-TPR étant considérés comme flexibles.

Dans les 10 modèles obtenus qui permettent un ajustement très correct avec la courbe expérimentale ($1.12 < \chi^2 < 1.32$), la position des deux domaines HTH varie peu : ils restent proches l'un de l'autre et proches des TPR, en position croisée (Figure 78).

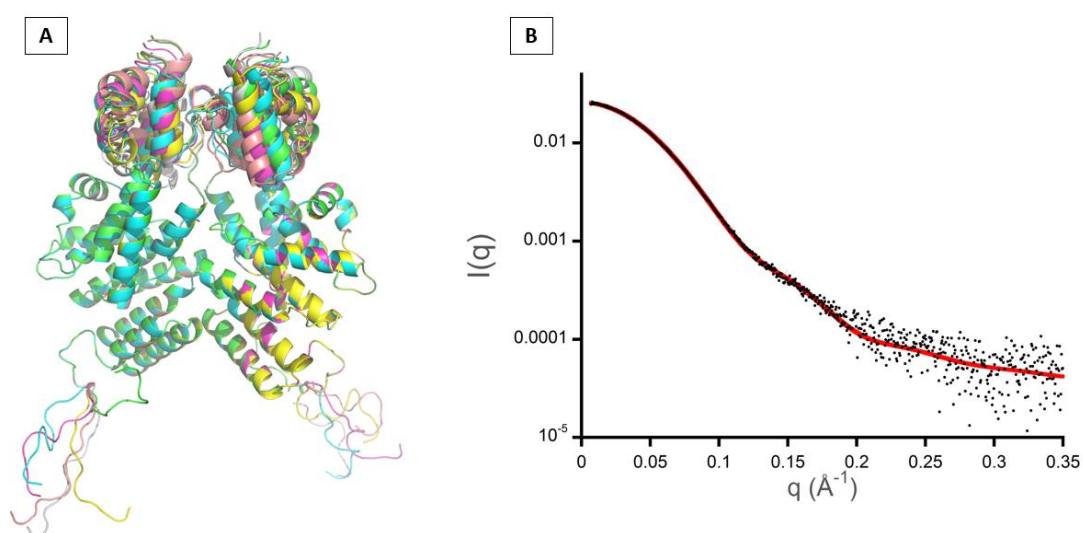


Figure 78 : Modélisation de ScuR-BI7

A- 5 modèles de ScuR/B17 (pour ne pas surcharger la figure), compatibles avec les résultats de SAXS. **B-** Comparaison de la courbe calculée (trait rouge) sur l'un des modèles, avec les données expérimentales (points noirs). Les autres modèles donnent des résultats identiques.

Nous avons vu ci-dessus que les courbes de SarF et SarF-BI7 sont très semblables, toutes deux caractéristiques d'un dimère. L'analyse par SAXS ne permet donc pas d'expliquer l'augmentation de T_m observée par DSC lors de la fixation de BI7 sur SarF. Ces courbes étant différentes de la courbe obtenue pour le dimère de ScuR-BI7, il n'est donc pas surprenant de constater que des modèles du dimère de SarF, réalisés à partir des structures de ComR comme pour le complexe ScuR-BI7, ne permettent pas de décrire la courbe expérimentale (Figure 79).

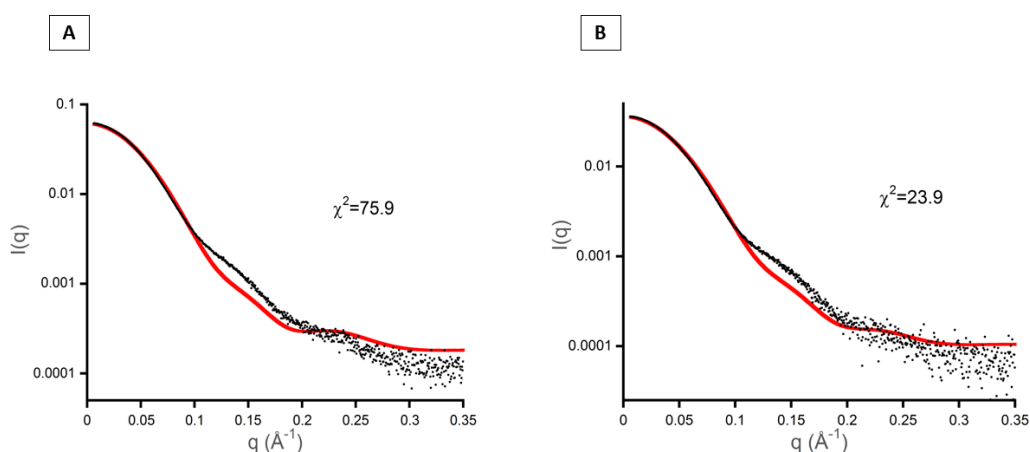


Figure 79 : Courbes expérimentales et calculées pour les modèles SarF apo et SarF-BI7
La Comparaison à la courbe expérimentale (points noirs) de la courbe calculée (courbe rouge) sur un modèle de dimère de SarF sans (a) ou avec (b) peptide, montre une absence de fit.

On pourrait chercher à décrire la courbe expérimentale en permettant aux domaines HTH de décrire des mouvements de grande amplitude, ce qui permettrait de mieux rendre compte de l'allure de la courbe dans la région $0.1 < q < 0.2 \text{\AA}^{-1}$, mais ceci a peu de sens en l'absence d'informations sur l'interface des deux domaines TPR. Puisque SarF est en dimère même sans peptide, alors que ScuR et ComR_{Sth} sont monomériques en l'absence de peptide et ne dimérisent qu'en présence du peptide, il y a lieu de penser que l'interface des deux domaines TPR de SarF pourrait être différente de celle de ScuR et ComR_{Sth}.

Nous disposons donc de trop peu d'information sur l'interface entre les domaines TPR des deux monomères de SarF, pour pouvoir extraire un modèle tout-atome à partir des data de SAXS, et il faut impérativement avoir recours à la cristallographie.

2.2.5. Perspectives

J'ai réalisé des essais de cristallisation avec la protéine SarF seule, en complexe binaire avec BI7 et en complexe ternaire avec BI7 et un fragment d'ADN de 22bp issue du promoteur pSptA (séquence 5'-AAAGTGACATAGATGTCCTTTT-3'). Le complexe binaire SarF-BI7 n'a jamais cristallisé et les différents cristaux obtenus avec la forme apo ne diffractaient pas. Par chance, le complexe ternaire SarF-BI7-pSptA a donné des cristaux diffractant à basse résolution. Après optimisation des conditions de cristallisation, j'ai obtenu un cristal plus gros que la moyenne (Figure 80) qui a diffracté à moyenne résolution.



Figure 80 : Cristal du complexe SarF-BI7-pSptA.

La condition de cristallisation est la suivante : 26% PEG4000, 0.2M Ammonium Sulfate, 0.1M Sodium Acetate pH 5.7.

Le cristal appartient au groupe d'espace C2 (Tableau 6)

LATTICE-CHARACTER	BRAVAIS-LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROM & DEGREES)						REINDEXING TRANSFORMATION											
			a	b	c	alpha	beta	gamma	1	0	0	0	0	1	0	0	0	0	1	0
* 44	aP	0.0	72.0	72.2	154.5	98.9	98.7	101.8	1	0	0	0	0	1	0	0	0	0	1	0
* 14	mC	5.7	90.9	111.9	154.5	90.1	104.0	89.9	1	1	0	0	-1	1	0	0	0	0	1	0
31	aP	102.9	72.0	72.2	159.2	77.8	77.7	101.8	1	0	0	0	0	1	0	0	1	1	1	0
10	mC	106.7	90.9	111.9	159.2	90.1	109.7	90.1	1	1	0	0	1	-1	0	0	-1	-1	-1	0
41	mC	188.4	306.4	72.2	72.0	101.8	101.6	85.4	0	-1	-2	0	0	-1	0	0	-1	0	0	0
37	mC	190.3	306.4	72.2	72.0	101.8	101.6	85.4	0	-1	-2	0	0	-1	0	0	-1	0	0	0
17	mC	214.2	111.9	90.9	160.2	87.1	110.3	90.1	1	-1	0	0	1	1	0	0	0	1	1	0
16	oF	218.0	90.9	111.9	300.4	90.1	93.1	90.1	1	1	0	0	1	-1	0	0	-1	-1	-2	0

Tableau 6 : Indexation du jeu de donnée du complexe SarF-BI7-pSptA.

Le fichier CORRECT.LP permet de visualiser, entre autres, le réseau retenu pour l'indexation des données. Dans notre cas, le groupe d'espace choisit est entouré en rouge.

Les données, traitées par XDS, ont été coupées à 3.3Å de résolution (Tableau 7)

RESOLUTION LIMIT	SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano	
	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE										
9.68	6105	934	947	98.6%	4.6%	4.7%	6105	35.73	5.0%	99.8*	-30	0.683	771
6.92	11454	1572	1574	99.9%	6.3%	5.9%	11454	26.23	6.8%	99.8*	-26	0.728	1439
5.67	13496	2015	2019	99.8%	14.6%	14.2%	13496	12.81	15.9%	98.8*	-17	0.722	1871
4.92	16842	2390	2393	99.9%	20.1%	19.7%	16842	10.36	21.7%	98.6*	-11	0.744	2253
4.41	18542	2697	2703	99.8%	23.7%	23.4%	18541	9.02	25.7%	98.2*	-11	0.755	2539
4.02	18790	2946	2948	99.9%	38.5%	37.9%	18790	5.59	42.1%	95.7*	-12	0.769	2704
3.73	22247	3226	3227	100.0%	67.5%	71.4%	22247	3.35	73.1%	90.1*	-13	0.694	3034
3.49	24481	3428	3429	100.0%	113.0%	128.2%	24481	1.99	121.9%	75.7*	-10	0.638	3283
3.29	23913	3448	3638	94.8%	170.6%	208.3%	23858	1.23	184.2%	55.9*	-8	0.596	3178
total	155870	22656	22878	99.0%	15.9%	16.7%	155814	8.29	17.2%	99.8*	-13	0.697	21072

Tableau 7 : Statistiques du jeu de données du complexe SarF-BI7-pSptA.

Le jeu est complet à une résolution de 3.3Å

J'ai ensuite utilisé le logiciel Matthews_coef de CCP4 pour analyser le contenu de l'unité asymétrique. En définissant comme poids moléculaire le complexe dimérique SarF-BI7-pSptA, on détermine sans ambiguïté que le cristal peut contenir deux complexes par unité asymétrique (Tableau 8).

Nmol/asym	Matthews Coeff	%solvent	P(3.29)	P(tot)
1	4.85	74.66	0.02	0.01
2	2.43	49.32	0.98	0.99
3	1.62	23.98	0.00	0.00

Tableau 8 : Détermination du nombre de complexes SarF-BI7-pSptA par unité asymétrique

J'ai alors essayé de résoudre le problème des phases par remplacement moléculaire, en utilisant comme modèle les différentes structures de ComR, sous forme monomérique ou dimérique, en considérant les domaines HTH et TPR comme deux objets indépendants. Malheureusement, aucune solution n'a pu être obtenue à ce jour, ce qui corrobore les résultats de SAXS, suggérant que la protéine SarF adopte une structure différente de ses paralogues, et en particulier de ScuR malgré les 82% d'identité de séquence.

J'ai donc produit et purifié la protéine SarF marquée à la sélénométhionine afin de tenter un phasage par diffusion anormale. La protéine marquée a cristallisé dans les mêmes conditions que la protéine native en présence de BI7 et d'ADN mais les cristaux ne diffractaient pas. J'ai alors réalisé de nouveaux criblages de conditions de cristallisation mais aucune nouvelle piste n'a pu être identifiée à ce jour.

III. Discussion

A une époque où les traitements antibiotiques classiques sont de moins en moins efficaces, on cherche des stratégies alternatives pour lutter contre les infections microbiennes. Le principal problème des antibiotiques actuels est la mort cellulaire qu'ils entraînent. A cause de ça, ils vont induire des mutants d'échappement capables de résister aux nouveaux traitements. Une solution potentielle à ce problème pourrait venir du ciblage du QS comme approche thérapeutique. On appelle ça le quorum quenching (QQ) (RAFFA *et al.* 2005; RUTHERFORD & BASSLER 2012; MION *et al.* 2019).

La principale force du QQ est de limiter la pression de sélection et d'éviter l'apparition de résistances ce qui est particulièrement vrai pour les bactéries pathogènes étant donné que les mécanismes de virulence ne sont généralement pas essentiels à la survie de la bactérie (RASKO & SPERANDIO 2010). Cependant, ce point de vue n'est pas partagé par tous. En effet, la variation des systèmes QS au sein des souches d'une même espèce risque de favoriser le développement de souches résistantes au QQ employé (DEFOIRDT *et al.* 2010). D'autre part, comme on a pu le voir au cours de ce manuscrit, il n'est également pas rare que certains processus bactériens majeurs comme la virulence ou la compétence soient régulés par plusieurs systèmes QS au sein d'une même bactérie. Ainsi, chez *S. aureus*, la virulence est contrôlée à la fois par les systèmes Agr et Sae (ZHU & KAUFMANN 2013). De ce fait, l'inhibition du mécanisme de quorum sensing pourrait atténuer le pouvoir pathogène de *S. aureus*, mais pas le faire disparaître totalement. Le QQ est donc peut-être à considérer plus comme un complément aux antibiotiques que comme une stratégie antibactérienne en soi.

De nombreuses études basées sur l'approche QQ pour le développement de nouveaux antimicrobiens sont déjà bien avancées chez les bactéries à gram négatif. Par exemple, dès 2006 chez *Pseudomonas aeruginosa*, on faisait déjà des expérimentations animales, à succès, sur des cibles thérapeutiques pouvant inhiber le mécanisme du QS (LE BERRE *et al.* 2006; SCHUSTER *et al.* 2013).

Différentes stratégies sont envisagées pour inhiber le QS à AHL des bactéries à gram négatif (Figure 81). De nombreuses enzymes comme les lactonases, les acylases ou les oxydoréductases (FETZNER 2015) sont capables de dégrader les AHLs, ce qui en font des enzymes intéressantes pour le QQ. A défaut de dégrader les molécules signal, il est également

possible de les empêcher d'interagir avec leur partenaire en les séquestrant à l'aide de «molécules cages » comme la cyclodextrine (KATO *et al.* 2006). Enfin, des molécules inhibitrices capable d'entrer en compétition avec la phéromone pour se fixer sur le site du récepteur ont également été découvertes, certaines naturelles (TANG & ZHANG 2014) et d'autres synthétiques (SCUTERA *et al.* 2014).

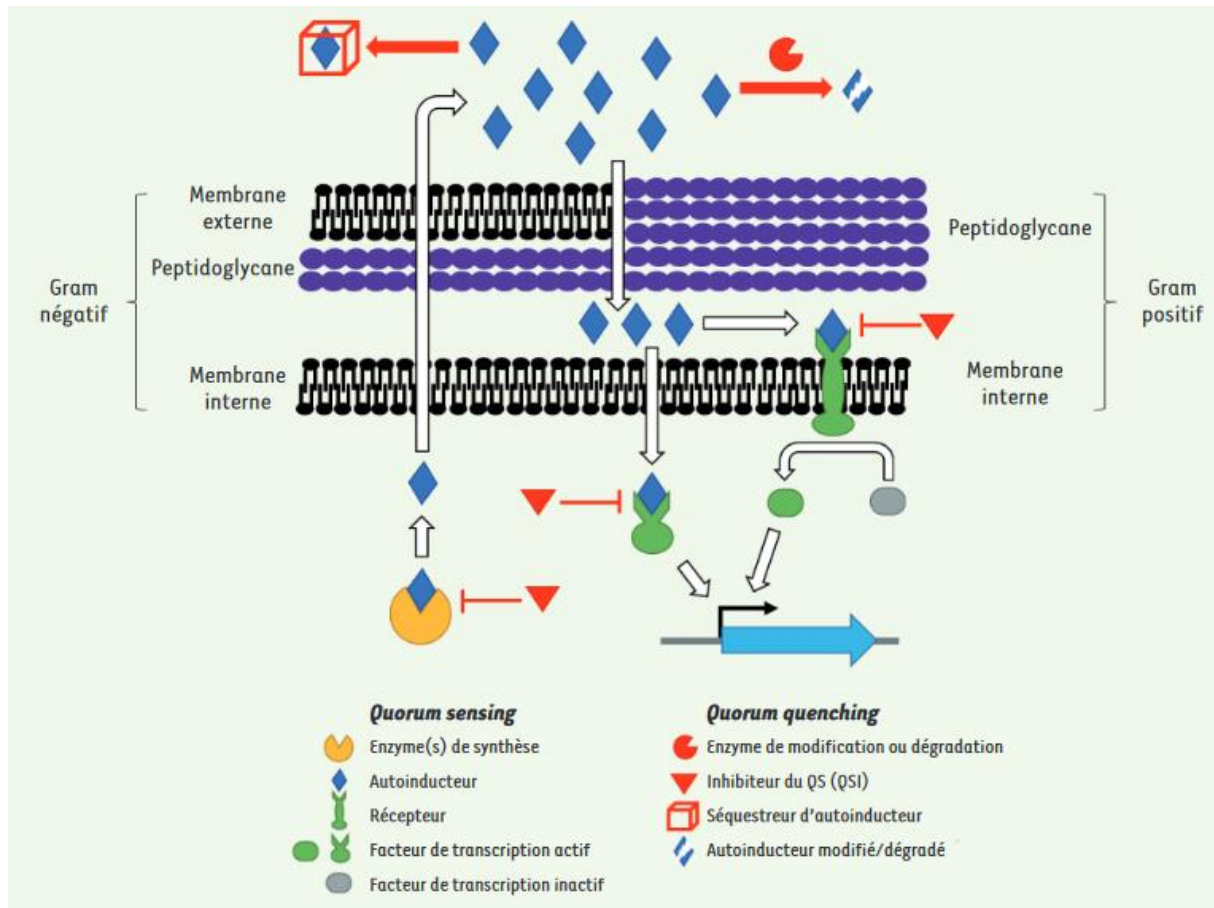


Figure 81 : Les différentes stratégies de QQ chez les bactéries à Gram négatif (MION *et al.* 2019)

Chez les bactéries à gram positif dont le QS repose sur des phéromones peptidiques (AIP), les études de QQ concernent essentiellement les systèmes QS indirects, mieux caractérisés que les systèmes directs à RNPP. Ainsi, chez le pathogène *Staphylococcus aureus*, des travaux visant de QQ sont également en cours (KIM *et al.* 2017; TAN *et al.* 2018). On peut notamment mentionner la mise au point de molécules cages qui agissent comme anticorps synthétiques capables de séquestrer des AIPs, à l'instar de la cyclodextrine (PARK *et al.* 2007). A contrario, l'inhibition des systèmes QS directs des bactéries à gram positif n'a pas encore été exploitée (VADAKKAN *et al.* 2018). Cibler le système ComRS qui régule la compétence chez les

streptocoques permettrait d'inhiber le transfert de gènes entre différentes souches et donc d'empêcher la transmission de résistances aux antibiotiques.

Durant ce travail de thèse, nous avons pu décrire de manière très précise le mécanisme d'activation du récepteur ComR par sa phéromone XIP. Nous avons également démontré que la substitution d'un seul résidu de ce peptide pouvait totalement modifier la sélectivité de reconnaissance de ce peptide par les récepteurs de différentes espèces. Ces résultats concernent le mécanisme d'activation des récepteurs ComR de type I. *S. salivarius*, bien que faiblement virulent, a été associé à diverses infections, comme la méningite et la bactériémie. Néanmoins, les streptocoques les plus pathogènes sont ceux qui possèdent un système ComRS de type II. Ainsi, il existe des souches de *S. pneumoniae* résistantes à la pénicilline qui posent un énorme problème de santé publique (MAMISHI *et al.* 2014) alors que *S. pyogenes* est une des bactéries les plus virulentes pouvant causer diverses infections mortelles qui entraînent la mort de plusieurs centaines de milliers de personnes par an (CARAPETIS *et al.* 2005).

Il convient donc, afin de pouvoir cibler les espèces pathogènes, de vérifier que le mécanisme moléculaire est bien commun aux trois types de système ComRS. Un premier élément de réponse est apporté par la structure cristalline de la forme apo de ComR_{Ssu} de *S. suis* (Code PDB : 5FD4) qui appartient au type III (SHANKER *et al.* 2016). On y reconnaît la structure monomérique caractérisée par la séquestration du domaine HTH par le domaine TPR. La superposition avec ComR_{sth} apo ([Figure 82](#)) montre, que, avec un rmsd de 2.75Å sur 271 résidus, la structure est globalement conservée. Néanmoins, on observe dans la poche de fixation de ComR_{Ssu}, à la place du peptide, des résidus provenant de l'étiquette 6xHis N-terminale d'une molécule voisine. De ce fait, la structure de ComR_{Ssu} ne correspond pas réellement à la forme apo ouverte du récepteur.

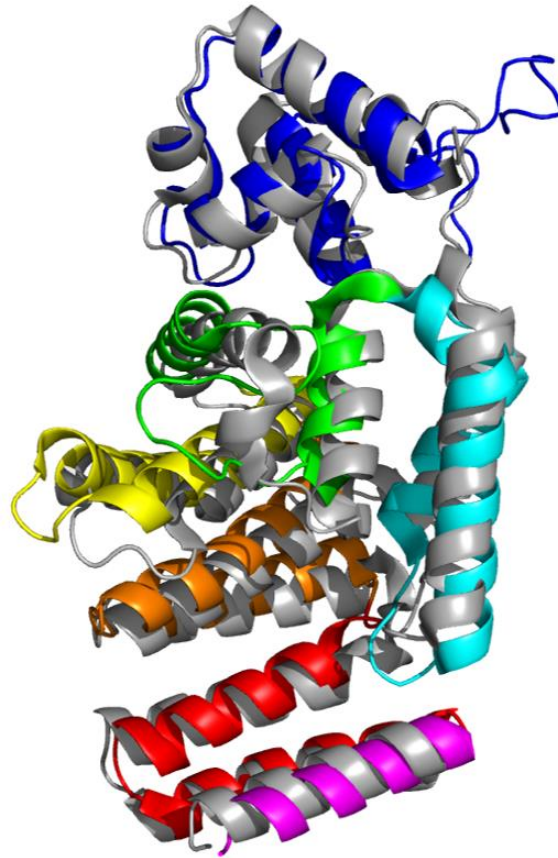


Figure 82 : Superposition de la forme apo de ComR *S. thermophilus* (Code PDB : 5JUF) en gris et ComR *S. suis* (Code PDB : 5FD4) en couleur

Comme mentionné plus haut, j'ai également commencé à travailler sur les protéines ComR de *S. mutans* et *S. pyogenes*. Lors de l'étape de purification par chromatographie d'exclusion par la taille, j'ai pu vérifier qu'elles étaient monomériques en absence de peptide, ce qui suggère un mécanisme commun à tous les systèmes ComRS.

La comparaison de séquences de ComR de types I, II et III ([Figure 83](#)), qui montre une forte conservation des résidus directement impliqués dans ce mécanisme, supporte également cette hypothèse.

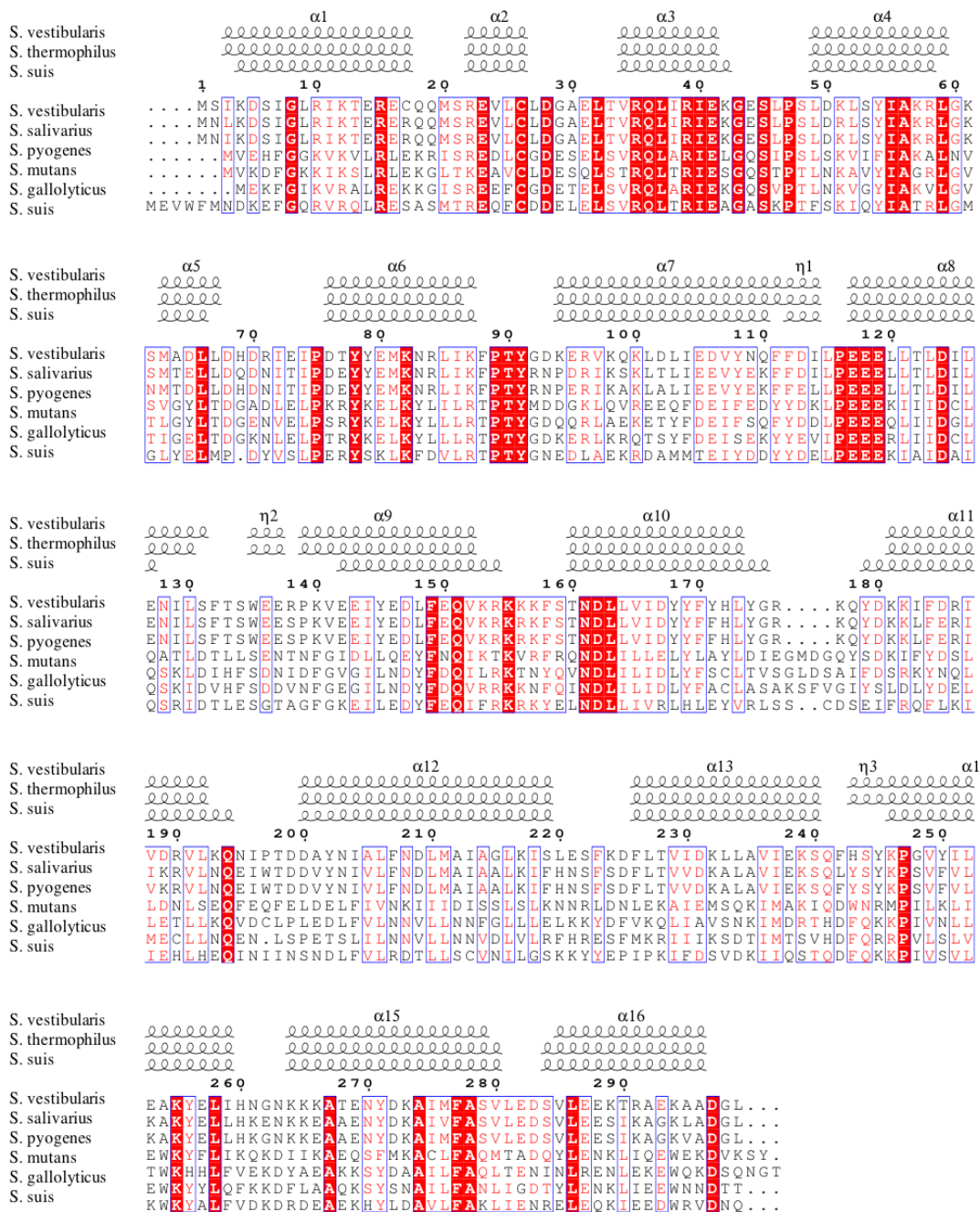


Figure 83 : Alignement de séquence entre protéines ComR de type I (*S. vestibularis*, *S. thermophilus* et *S. salivarius*), de type II (*S. pyogenes*, *S. mutans* et *S. gallolyticus*) et de type III (*S. suis*)

Si l'on regarde les résidus impliqués dans la séquestration du domaine HTH chez *S. thermophilus*, on retrouve une méthode de séquestration similaire chez *S. suis* avec une conservation de presque tous les résidus (Figure 84).

En effet, seule une mutation conservatrice entre une Arginine (R51) et une Lysine (K55) est observée. Chez les ComR de type II, le résidu E146 de ComR_{Sth} est remplacé par une asparagine ou une glutamine, ce qui suggère qu'un pont salin serait remplacé par une liaison hydrogène, résultant en une séquestration plus fragile du domaine HTH.

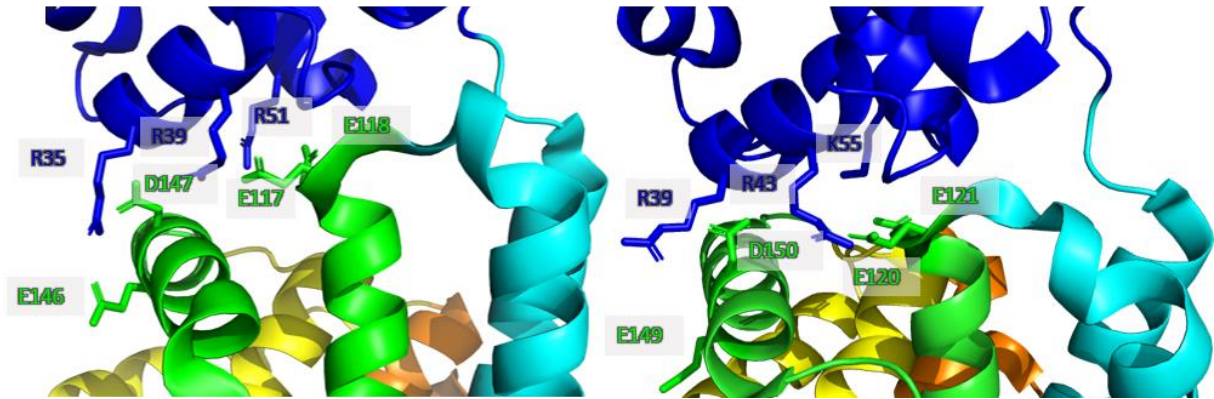


Figure 84 : Zoom sur la séquestration des domaines HTH chez ComR *S. thermophilus* apo (Code PDB : 5JUF) à gauche et ComR *S. suis* apo (Code PDB : 5FD4) à droite.

Au niveau de l'interface de dimérisation, les résidus formant les ponts salins chez ComR_{Sth} sont globalement conservés chez ComR_{Ssu} (Figure 83). Néanmoins, une insertion de 2 résidus au niveau de la boucle α 10- α 11 ne permet pas de savoir si le pont K87-D200 de ComR_{Sth} correspondrait à une interaction entre R90 et N204, N206 ou D207 chez ComR_{Ssu}. Chez les ComR de type II, l'insertion est de 4 et non pas 2 résidus, ce qui rend encore plus aléatoire l'identification du résidu pouvant faire un pont salin avec l'équivalent du résidu K87 de ComR_{Sth} ainsi que ceux pouvant former l'équivalent de l'interaction K246-E282 observée chez ComR_{Sth} et ComR_{Sve}. Néanmoins, les domaines HTH fixant l'ADN sous forme de dimère, il est très probable que la fixation du peptide induise la dimérisation des ComR de types II et III d'une façon très semblable à ce qui a été observé dans le type I.

Si l'on regarde au niveau de la fermeture de la poche, la paire hydrophobe Y91, L286 stabilisant la forme active fermée chez les ComR de type I est strictement conservée alors que le pont salin entre R96 et D283 pourrait-être remplacé par une liaison hydrogène impliquant une asparagine à la place de D283 ou ne pas avoir d'équivalent chez ComR_{Ssu} chez lequel on observe une leucine à la position 96.

Il est également intéressant de regarder si le mode de fixation du peptide et les détails de son mode d'activation pourraient être conservés entre les différents systèmes ComRS. Les résidus

T90 et K100 de ComR_{Sth}, qui jouent un rôle essentiel dans l'ancrage du peptide par son carboxylate C-terminal, sont conservés (Figure 83).

Dans les systèmes de type I, le bon positionnement de la chaîne latérale hydrophobe du résidu 8 est essentiel pour l'ancrage du COO⁻. Chez ComR_{Ssu} on observe dans la poche qui accueille cette partie du peptide la présence du résidu R170 (équivalent au résidu D167 de ComR_{Sth}), qui pourrait former un pont salin avec le glutamate E8 de XIP_{Ssu} WGTWVEE (Figure 85). Les peptides de type II ont également un résidu hydrophobe en position 8, ce qui explique certainement l'absence totale de cross-talk entre ComR_{Ssu} et les systèmes de type II (SHANKER *et al.* 2016).

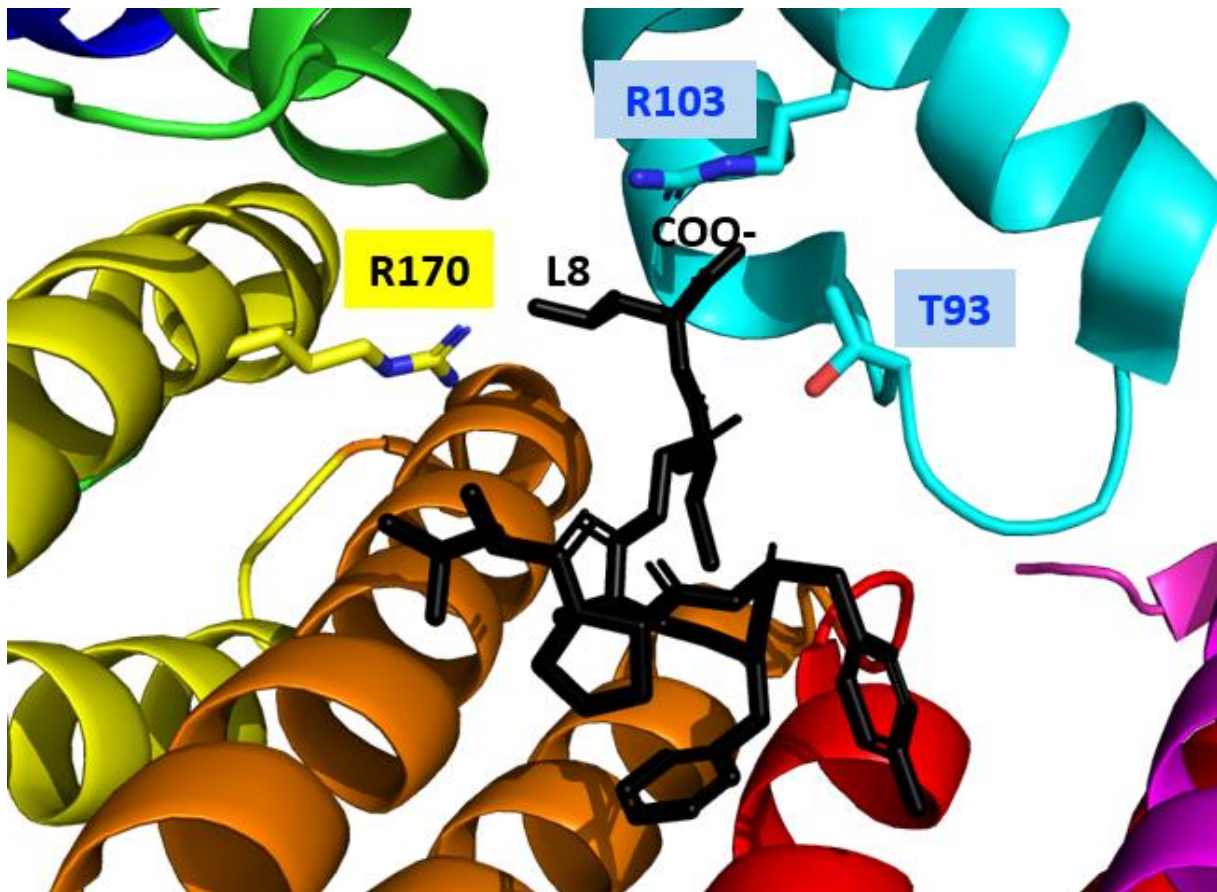


Figure 85 : Zoom de la poche de fixation de *S. suis* dans laquelle on a introduit le peptide XIP « LPYFAGCL » de *S. thermophilus*.

Les séquences des peptides XIP de type II et III étant très différentes des peptides de type I, en particulier à cause de l'absence de Proline en position 2, il est difficile de modéliser le

positionnement du reste du peptide au sein de la poche de fixation de leurs récepteurs respectifs. De plus, des essais de docking *in silico* des systèmes étudiés au cours de ma thèse se sont avérés très difficiles à réaliser du fait de la grande flexibilité de ces ligands. Il serait donc intéressant de confirmer la conservation du mécanisme et élucider le mode de fixation des peptides XIP d'un système de type II, par exemple ComR de *S. pyogenes* ou *S. mutans* avec lesquels j'ai obtenu des cristaux à optimiser.

En conclusion, cette étude structure-fonction ouvre la voie vers la conception rationnelle de peptides ou pseudo-peptides inhibiteurs du récepteur ComR régulant la compétence.

Un autre récepteur QS de type RNPP, le régulateur de virulence PlcR de *B. cereus* fait déjà aujourd'hui l'objet d'une recherche d'inhibiteurs basée sur l'utilisation de peptides synthétiques modifiés (YEHUDA *et al.* 2018; YEHUDA *et al.* 2019). Des dérivés de son peptide activateur PapR qui inhibent la production de facteurs de virulence sans affecter la croissance bactérienne ont ainsi été découverts.

IV. Matériel et méthodes

1. Production des protéines

1.1. Plasmides utilisés

Tous les plasmides utilisés dans les purifications de protéines sont des plasmides de type pBAD fournis par l'équipe de Pascal Hols (LIBST, Louvain la Neuve, Belgique) (FONTAINE *et al.* 2013). Ils possèdent une cassette de résistance à l'Ampicilline. L'expression du gène d'intérêt est sous le contrôle d'un promoteur du même nom inductible à l'arabinose. Ils codent pour une protéine fusionnée à l'extrémité C-terminale avec une étiquette de type Strep-tag II de 8 AA (WSHPQFEK).

Dans un premier temps, on transforme avec 1 μ L de notre plasmide d'intérêt 100 μ L de souches *E. coli* XL1-Blue ce qui va permettre de surexprimer notre plasmide. On réalise ensuite une extraction d'ADN plasmidique grâce à une mini préparation d'ADN à partir du kit GenJET Plasmid Miniprep Kit (ThermoFisher SCIENTIFIC) en adaptant le protocole fourni. Le plasmide ainsi extrait est conservé à -20°C. Afin de connaître la concentration et de s'assurer de la pureté du plasmide (ratio 260/280 > 1.8), un spectre est réalisé au nanodrop.

1.2. Surexpression des protéines

L'expression des protéines ComR_{StH} et ComR_{Sve} a été réalisée dans une souche d'expression d'*E. coli* BL21 Gold alors que ScuR et SarF ont été exprimées dans une souche Top 10 d'*E. coli*. On transforme par choc thermique 100 μ L d'une culture de bactéries compétentes préparées en milieu 2*YT à laquelle on a ajouté 1 μ L de plasmide. Le mélange est incubé 45 secondes à 42°C puis 2 minutes dans la glace. Les bactéries transformées sont alors mises en culture à 37°C pendant une 1H dans 900 μ L d'un milieu nutritif 2*YT.

Cet aliquot de bactéries transformées est ensuite dilué dans 10mL de milieu 2*YT complémenté par 20 μ L d'ampicilline (50mg/mL), l'antibiotique correspondant à la résistance portée par le plasmide pBAD pour éliminer les bactéries non transformées, puis est mise en culture à 37°C pendant toute une nuit sous agitation constante (200rpm). Le lendemain, lorsque la pré-culture

est arrivée à saturation, elle est diluée dans un erlenmeyer de 2 litres contenant 800mL de milieu 2*YT complété par 1.6mL d'ampicilline (50mg/mL) et mise en culture à 37°C sous agitation constante. La croissance bactérienne est suivie par lecture au spectrophotomètre de la densité optique à 600nm (DO₆₀₀) sur des aliquots de 1mL prélevés stérilement. L'expression est induite lorsque la DO₆₀₀ est comprise entre 0.4 et 0.5. Avant induction de l'expression du gène d'intérêt par ajout de 8mL d'Arabinose (1,3M) la culture bactérienne est refroidie 10 minutes dans la glace puis les bactéries sont remises en culture à 30°C sous agitation pendant 4 heures. La culture est ensuite centrifugée 20min à 4400g et à 4°C. Avant d'être stocké à -20°C, chaque culot bactérien obtenu à partir d'une culture de 800mL est repris, selon sa masse, dans 30 à 40mL de tampon A (100mM Tris HCl, pH8, 150mM NaCl, 1mM EDTA) supplémenté avec 10% (v/v) glycérol.

1.3. Marquage des protéines au Sélénium

Dans une première pré-culture, 20mL de milieu 2*YT sont inoculés par 100µL de bactéries *E. coli* Top 10 transformées par le plasmide pBAD d'intérêt puis incubés à 37°C sous agitation constante (200rpm) pendant 6 à 8 heures. Ensuite, 45mL d'un milieu synthétique pauvre (Tableau 9) sont inoculés avec 5mL de cette première pré-culture puis incubés une nuit entière à 37°C sous agitation (200rpm).

Le lendemain matin, 40mL de cette deuxième pré-culture sont utilisés pour ensemercer 760mL du même milieu pauvre. Lorsque la DO₆₀₀ atteint une valeur d'environ 0.6, on ajoute 25mL d'une solution de sélénométhionine (Se-MET) (ACROS Organics) à 2.5g/L ainsi que 10mL d'un mélange d'acides aminés (Sigma-Aldrich). La solution de sélénométhionine et le cocktail d'acides aminés sont préalablement solubilisés dans du NaOH 0.25M.

On met ensuite la culture bactérienne à incuber pendant 1H à 30°C sous agitation constante (200rpm). Enfin, on réalise un choc thermique 10min dans la glace avant d'induire avec 8mL d'Arabinose (1.3M) et de laisser la culture pousser pendant 4H à 30°C. Les culots bactériens sont alors récupérés par centrifugation (4400g, 20min, 4°C) et re suspendus comme décrit précédemment, dans le tampon A (100mM Tris HCl, pH8, 150mM NaCl, 1mM EDTA et 10% (v/v) Glycérol).

Vol (µL)	Ingrédients	Concentration stock
400	Glycérol	50% v/v
5000	Minéraux :	g/L
	KH ₂ PO ₄	136
	KOH	40
	NH ₄ Cl	11
500	MgSO ₄ , 7H ₂ O	80 g/L
250	CaCl ₂ , 2H ₂ O	10 g/L
50	Thiamine	2 g/L
12.5	FeSO ₄ , 7H ₂ O	10 g/L
50	ZnSO ₄ , 7H ₂ O	400 mg/L
	MnSO ₄ , 1H ₂ O	400 mg/L
	CuSO ₄ , 5H ₂ O	40 mg/L
	Na ₂ MoO ₄	200 mg/L
	H ₃ Bo ₃	500 mg/L
	KI	100 mg/L
100	Antibiotique (Ampiciline)	50 mg/mL
38637	H ₂ O	

Tableau 9 : Composition du milieu synthétique pauvre

2. Purification par chromatographie d'affinité

Après décongélation au bain marie, chaque culot bactérien est incubé 30min dans la glace en présence d'une tablette d'inhibiteurs de protéases (Thermo Scientific) et de 1µL de benzonase (Sigma-Aldrich) afin de dégrader l'ADN et l'ARN.

Les culots sont ensuite lysés par sonication (Branson Sonifier). Pour les bactéries *E. coli* BL21 Gold utilisées pour l'expression des protéines ComR_{Sth} et ComR_{Sve}, je faisais 4 cycles (40 sec ON / 30 sec OFF), à environ 35% d'énergie et un pulse par seconde avec des pulses de 0.9 à 0.7 secondes. Pour les bactéries *E. coli* Top 10 utilisées pour l'expression des protéines ScuR et SarF, j'ai augmenté à 8 cycles (40 sec ON / 30 sec OFF), 35% d'énergie et des pulses de 0.7

secondes. Les extraits cellulaires vont ensuite être centrifugés pendant 30min à 20 000g, 4°C, afin de récupérer, dans le surnageant, la fraction soluble.

La fraction soluble de nos extraits cellulaires est ensuite purifiée selon un protocole en deux étapes de chromatographie liquide réalisées à 4°C, à l'aide d'un système FPLC de type Akta pure (GE healthcare) placé en chambre froide. Les profils d'élution sont suivis par lecture de l'absorbance à 280 et 260nm. Entre deux purifications, toutes les colonnes sont équilibrées en 20% d'isopropanol et conservées à 4°C. La colonne est ensuite lavée par 3 volumes H₂O avant de l'équilibrer dans le tampon de purification.

L'étiquette Strep-tag II C-terminale portée par les protéines étudiées ici possède une forte affinité avec un dérivé de la streptavidine qu'on nomme Strep-Tactin. La fraction soluble de nos extraits cellulaires a donc été déposée sur une colonne StrepTrap HP de 5mL (GE Healthcare) équilibrée en tampon A (100mM Tris HCl, pH8, 150mM NaCl, 1mM EDTA). Seule la protéine d'intérêt portant le Strep-tag est retenue sur la colonne. Elle est éluée par compétition avec 15mL de tampon A complété par 2.5mM de D-desthiobiotine. L'élution se fait brutalement, en une seule étape, et non par gradient, afin d'éviter que l'échantillon purifié ne soit trop dilué dans de nombreuses fractions, ce qui diminuerait le rendement de la purification.

Afin d'éliminer d'éventuels contaminants résiduels, mais surtout de séparer la fraction correctement repliée de la protéine d'intérêt d'une éventuelle fraction agrégée et de changer de tampon, on réalise ensuite une chromatographie d'exclusion par la taille (SEC). Pour ComR qui est monomérique (MM 37kDa) en absence de peptide, nous utilisons une colonne Superdex-S75 HR16/60 (GE Healthcare) qui sépare les protéines entre 3 et 70kDa. Parce que nous ne savions pas si ScuR et SarF allaient être monomérique ou dimérique en solution, j'ai opté pour une Superdex-S200 HR16/60 (GE Healthcare) qui sépare les protéines entre 10 et 100 kDa. Avant injection sur la colonne préalablement équilibrée par 1.2 volumes de tampon SEC de plus faible molarité (20mM Tris HCl, pH8, 150mM NaCl, 1mM EDTA), l'échantillon est, si besoin, concentré jusqu'à un volume de 5mL (correspondant au volume de la boucle d'injection utilisée) par ultrafiltration sur membrane de cut-off 10 kDa (Vivaspin-10) à 4000rpm, 4°C,

A chaque nouvelle étape de purification, des aliquots de 10µL des différentes fractions d'élution sont prélevés afin d'être analysés sur gel SDS-PAGE et par mesure du spectre d'absorbance sur un spectrophotomètre de type nanodrop 200 (Thermo Scientific). Cette mesure permet

également de vérifier que le ratio DO_{260}/DO_{280} est < 0.6 , valeur attendue pour un échantillon protéique sans contamination par des acides nucléiques. Enfin, la concentration des échantillons protéiques purifiés est calculée en appliquant la loi de Beer-Lambert à partir du coefficient d'extinction molaire théorique à 280nm calculée à partir de la composition en acides aminés de chacune de nos protéines d'intérêt (Tableau 10). Après regroupement des différentes fractions composant le pic d'intérêt, les échantillons étaient concentrés par ultrafiltration sur membrane de cut-off 10 kDa (Vivaspin-10) à 4000rpm, 4°C jusqu'à atteindre une concentration de quelques mg/ml (environ 0.15mM) en tampon SEC sauf pour quelques essais de cristallographie ou la concentration des protéines pouvait dépasser les 20mg/mL. L'échantillon était ensuite alliqué, refroidi rapidement dans l'azote liquide avant d'être stocké à -20°C.

Protéine	Coefficient extinction molaire
ComR <i>S. thermophilus</i>	33 350
ComR <i>S. vestibularis</i>	29 340
ScuR	38 280
SarF	36 790

Tableau 10 : Coefficients d'extinctions molaire
Valeurs calculées avec l'outil Prot-Param de ExPASy

Lors des premières purifications de ComR_{Sve}, je me suis aperçu que le rapport $Abs(260)/Abs(280)$ de l'échantillon en sortie de Strep-Tactin était supérieur à 0.6, suggérant une contamination par des acides nucléiques, et ce malgré l'ajout de benzonase dans le tampon de lyse. J'ai donc réalisé une seconde étape de chromatographie d'affinité en utilisant, cette fois-ci, une colonne greffée à l'héparine, un polymère de disaccharides mimant la structure de l'ADN. Néanmoins, lors du dépôt de l'échantillon et du lavage de la colonne, aucune absorbance à 260nm pouvant refléter l'élimination d'ADN n'était observé. De plus, les fractions éluées par ajout de NaCl et contenant la protéine présentaient toujours un rapport $Abs(260)/Abs(280) > 0.6$ (Figure 86). Ce résultat suggérait donc qu'il n'y avait pas de contamination avec des acides nucléiques ou nucléotides, mais que ce rapport élevé pouvait plutôt être dû à un problème d'agrégation de la protéine. L'analyse de l'échantillon élué de la colonne d'héparine par SEC confirma cette hypothèse, avec une élution de la protéine dans le volume mort de la colonne. J'ai réussi à régler ce problème en modifiant les réglages du sonicateur utilisé pour la lyse des bactéries.

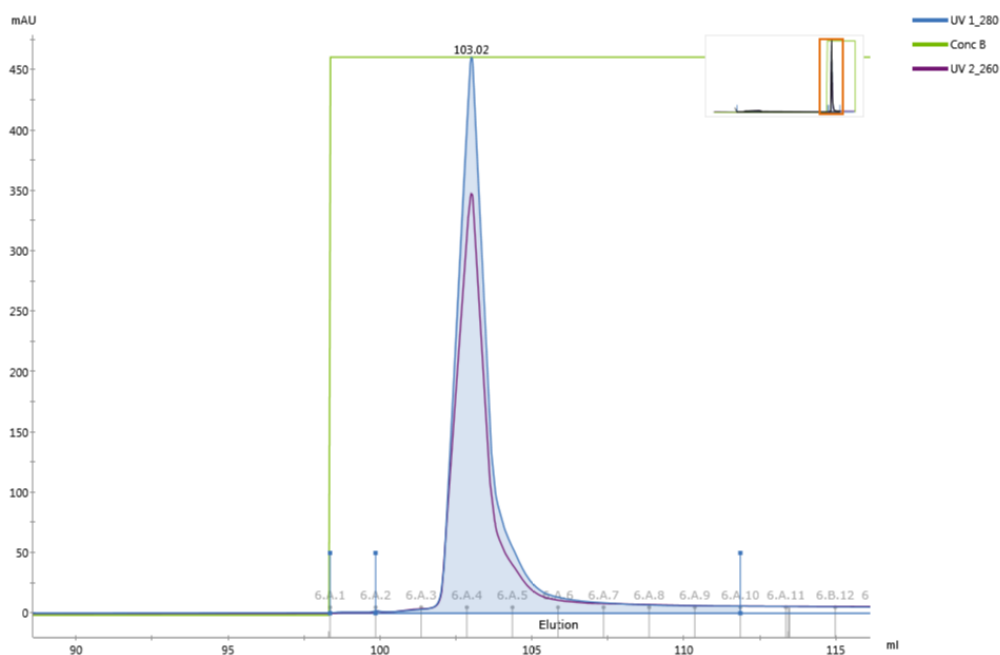


Figure 86 : Profil d'élution de ComR_{Sve} sur colonne d'héparine

3. Préparation des ligands

Les peptides utilisés sont des peptides synthétiques lyophilisés (GenScript) d'une pureté > 95%. Etant peu solubles en eau, ils sont re-suspendus dans 100% de diméthylsulfoxyde (DMSO), à une concentration finale de 50 mM, afin de pouvoir réaliser les essais de cristallisation à des concentrations élevées de protéines tout en ayant, dans les mélanges finaux, le peptide en excès.

Les oligonucléotides purifiés par HPLC et lyophilisés par le producteur (Eurofins) sont remis en suspension dans le tampon SEC (20mM Tris HCl pH8, 150mM NaCl, 1mM EDTA). Afin d'obtenir un fragment d'ADN double brin, les deux brins complémentaires sont mélangés de manière équimolaire à une concentration finale de 2mM et hybridés en les plongeant dans de l'eau portée à ébullition puis en les laissant refroidir lentement à température ambiante sur une nuit.

4. Mesures interactions peptide-protéine

4.1. Calorimétrie différentielle à balayage (DSC)

L'effet de la fixation des peptides sur la stabilité des récepteurs étudiés ici a été analysé par DSC. Le principe de cette technique consiste à faire subir à l'échantillon un gradient de température qui va progressivement déplacer l'équilibre de la forme native vers la forme dénaturée de la protéine.

Lorsqu'une protéine se dénature, elle absorbe de la chaleur. La différence de température entre les deux cellules (ΔT) est mesurée et la puissance utilisée pour ramener ΔT à 0 permet de tracer la courbe de la calorimétrie C_p (kcal/mole/°C) en fonction de la température. La température de transition entre les deux états natif et dénaturé (T_m) est déterminée au sommet du pic dont l'aire correspond à l'enthalpie (ΔH) qui reflète les changements de conformation de la protéine (Figure 87). Plus la protéine est stable et plus son T_m va être élevé.

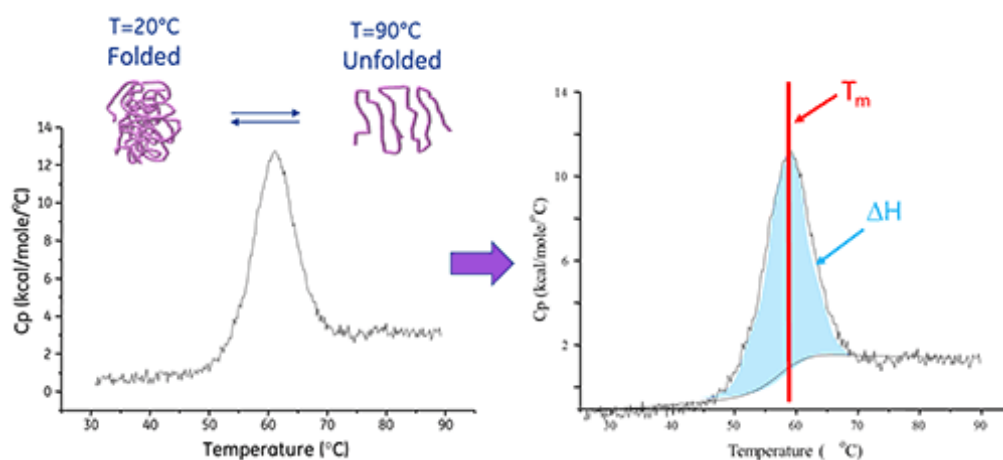


Figure 87 : Principe de la DSC

Les peptides étant suspendus dans 100% DMSO, des mesures préliminaires ont été réalisées avec chaque protéine afin de tester quel pourcentage de solvant elles supportent sans se dénaturer. Le T_m de toutes les protéines reste inchangé jusqu'à environ 10% DMSO mais, pour plus de sécurité, toutes les mesures ont été réalisées avec seulement 2% DMSO. La concentration en protéine était fixée à 15 μ M. Pour l'analyse de complexes, les peptides étaient ajoutés à une concentration de 150 μ M et, le cas échéant, l'ADN à 45 μ M.

Les mesures ont été réalisées sur un système automatisé MicroCal PEAQ-DSC (Malvern Panalytical). Les échantillons (350 μ L) sont stockés sur plaques de type Deepwell 96 puits thermostatées à 15°C. Chaque échantillon et le tampon correspondant sont prélevés pour analyse dans des cuves progressivement chauffées de 20 à 100°C à une vitesse de 1°C par minute. Les données ont été analysées par le logiciel MicroCal PEAQ-DSC.

4.2. Calorimétrie par titration isotherme (ITC)

J'ai utilisé l'ITC pour tenter de mesurer la constante de dissociation K_D caractérisant les complexes protéine/ligand étudiés. Le principe de cette technique est basé sur le fait que, lorsqu'une interaction entre biomolécules a lieu, de la chaleur est absorbée ou libérée. L'appareil est constitué de deux cellules, l'une contenant le tampon de référence et l'autre la protéine. Le ligand est progressivement injecté dans la cellule d'échantillon à l'aide d'une seringue et à chaque injection on mesure l'énergie nécessaire (en μ cal/sec) pour compenser la petite différence de température avec la cellule de référence. La variation de température décroît progressivement jusqu'à saturation de la protéine par le ligand. La superficie de chaque pic est intégrée en fonction du rapport de concentration ligand/protéine, ce qui permet de tracer un isotherme qui va permettre de calculer les différents paramètres d'interaction (Figure 88). La pente correspond à la constante d'association à l'équilibre, K_A , inverse de la constante de dissociation K_D , qui renseigne sur l'affinité. Le rapport molaire correspondant au centre de l'isotherme renseigne sur la stœchiométrie de l'interaction, c'est-à-dire sur le nombre de sites n de fixation du ligand sur la protéine. Enfin, l'isotherme permet aussi de mesurer l'enthalpie (ΔH) de la réaction, et donc de calculer les autres paramètres thermodynamiques de l'interaction, ΔG et ΔS .

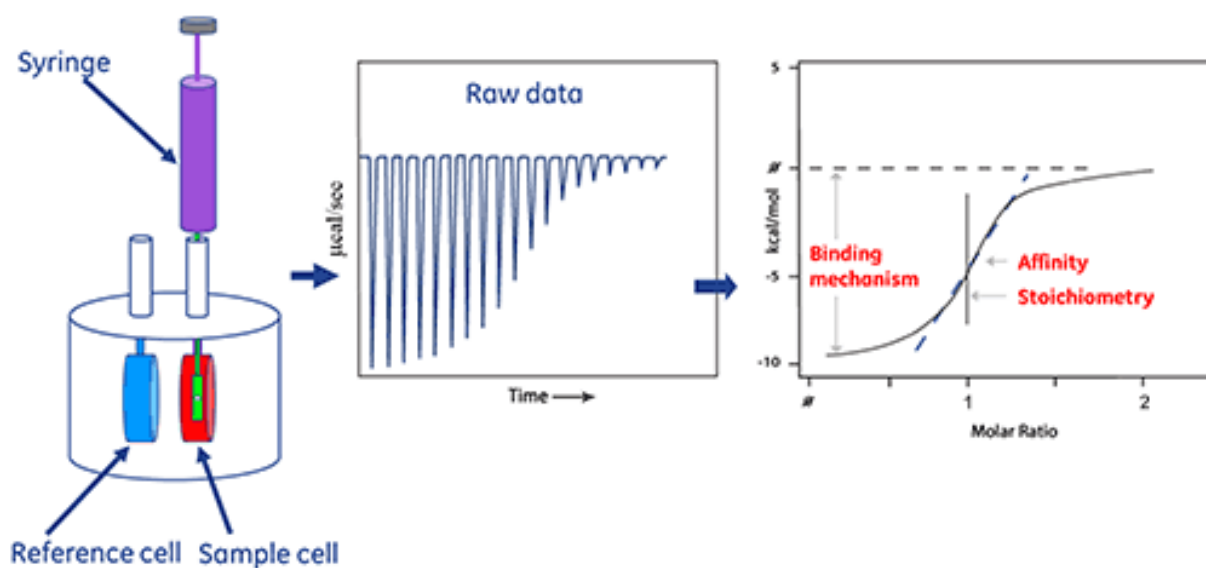


Figure 88 : Principe de l'analyse par ITC
Exemple d'une réaction exothermique avec dégagement de chaleur

Ces mesures sont très sensibles aux différences de tampon entre les deux cellules. Ainsi, des différences minimales de concentration en DMSO dans lequel le peptide est ressolubilisé provoquent d'importants pics artificiels qui masquent le signal dû à l'interaction protéine/peptide. Comme pour l'analyse en DSC, les mesures d'ITC ont été réalisées avec une concentration de 2% de DMSO.

Les mesures ont été réalisées à une température constante de 20°C sur un appareil MicroCal iTC200 (Malvern Panalytical), constitué de deux cellules de 200µL. La concentration en protéine dans la cellule d'échantillon était de 50µM et la seringue contenait le peptide à une concentration environ 10 fois supérieure à celle de la protéine, ajustée en fonction de l'affinité du complexe afin de voir toute la courbe de saturation. Le peptide était progressivement injecté à raison de 2µL toutes les 180 secondes. Les résultats ont été analysés avec le logiciel Origin.

4.3. Size Exclusion Chromatography Multi-Angle Light Scattering (SEC-MALS)

Le degré d'oligomérisation des protéines ScuR et SarF en présence et absence de peptide a été mesuré par SEC-MALS. Cette technique permet, à partir d'une chromatographie d'exclusion par la taille (SEC), de déterminer la masse molaire d'un objet en solution en s'affranchissant de la calibration de la colonne. Les molécules séparées en fonction de leur forme et de leur poids

moléculaire sur la colonne sont détectées par diffusion de lumière à plusieurs angles (90° et 6°) (MALS), ce qui permet de déterminer le rayon de giration de la molécule et d'en déduire sa masse. La machine comporte également un spectromètre capable de mesurer l'absorbance UV à 280nm et un réfractomètre permettant de mesurer l'indice de réfraction (RI) de l'échantillon. Comme pour les autres analyses d'interaction avec les peptides resolubilisés en 100% DMSO, les expériences ont été réalisées en tampon SEC + 2% DMSO final. Nous avons utilisé une colonne Superdex 200 10/300 GL increase (GE Healthcare) sur laquelle des échantillons de 100µL ont été déposés à une concentration de 2mg/ml. Les données ont été analysées par le logiciel OmniSEC.

5. Analyse structurale en solution par diffusion des rayons X aux petits angles (SAXS)

5.1. Principe

En attendant de pouvoir résoudre les structures de ScuR et SarF à haute résolution par cristallographie, j'ai utilisé le SAXS pour essayer de savoir si la structure de ComR était globalement conservée chez ces deux paralogues. L'intérêt de cette technique, qui renseigne uniquement sur la structure à basse résolution des macromolécules, réside essentiellement dans la possibilité de les étudier dans un environnement proche des conditions physiologiques.

Comme la cristallographie, le SAXS est basé sur l'interaction entre les rayons X et les électrons de la matière. Les ondes diffusées par un échantillon irradié par un rayonnement X monochromatique sont enregistrées sur un détecteur bidimensionnel, donnant lieu à une image de diffusion 2D ([Figure 89](#)) ([PUTNAM et al. 2007](#)). Les orientations des molécules en solution étant quelconques, cette image est isotrope, et la quantité que l'on va analyser lors d'une expérience de SAXS résulte de l'intégration radiale de l'image : on obtient ainsi une intensité diffusée $I(q)$ en fonction du module du vecteur de diffusion q , avec $q = (4\pi\sin\theta/\lambda)$ où 2θ représente l'angle de diffusion et λ la longueur d'onde du rayonnement X incident.

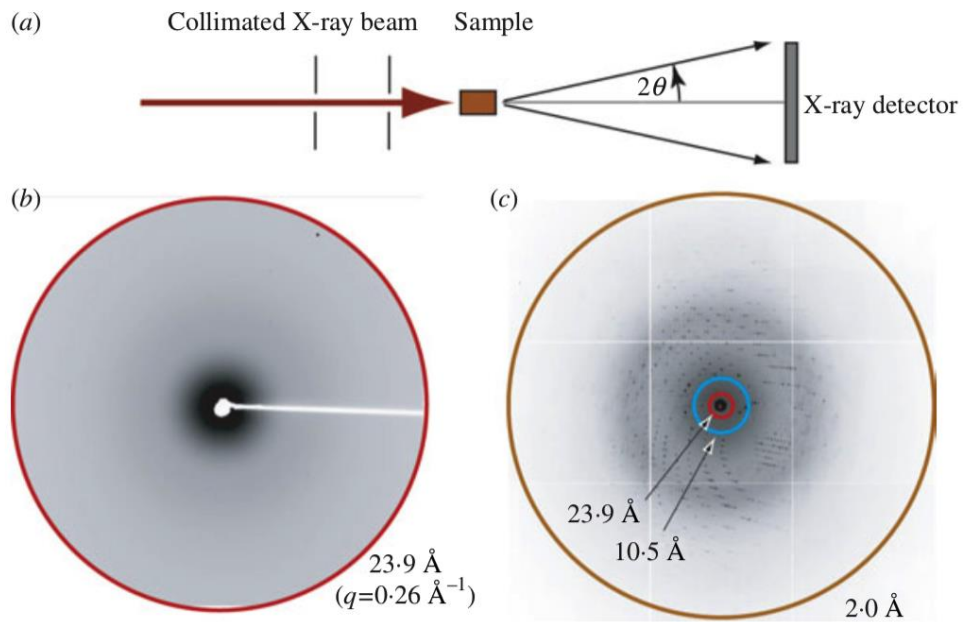


Figure 89 : Principe du SAXS

A- Représentation schématique d'une expérience de diffusion ou diffraction des rayons X. **B-** Image 2D obtenue dans une expérience de SAXS sur une solution de protéine. Dans cette expérience, le vecteur de diffusion q a une valeur maximale $q_{\max}=0.263 \text{ \AA}^{-1}$. La distance correspondante dans l'espace réel (communément désignée comme étant la résolution) vaut $2\pi/q_{\max}=23.9 \text{ \AA}$. **C-** Image de diffraction enregistrée sur un cristal. Le cercle rouge au centre de l'image correspond à la région explorée par SAXS dans l'image (b).

Le SAXS est une méthode dite de contraste car le signal de diffusion de la protéine $I_{\text{protéine}}(q)$ est obtenu en faisant la différence entre le signal de diffusion de la solution de protéine et celui du tampon, et est ainsi relié à la différence de densité électronique $\Delta\rho(r) = \rho(r) - \rho_s$ entre la protéine et le tampon :

$$I(q) = N \left\langle \left| \int_V \Delta\rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d\vec{r} \right|^2 \right\rangle$$

où N est le nombre de protéines, V le volume d'une protéine, et $\langle \rangle$ la moyenne sur toutes les orientations. A noter que cette expression n'est valable que si la solution est « idéale » c'est-à-dire si toutes les particules sont identiques et sans interaction.

De cette équation peuvent être extraites des informations sur l'état de polydispersité/d'agrégation de l'échantillon, sur l'état d'oligomérisation (grâce à la détermination de la masse) et sur la forme moyenne des molécules. Pour cela, les représentations ci-dessous sont utilisées :

- La représentation logarithmique de $I(q)$ en fonction de q^2 , qui permet notamment d'extraire la valeur du rayon de giration R_g grâce à l'approximation de Guinier, valable dans la région des petites valeurs de q :

$$\ln(I(q)) = \ln(I(0)) - \frac{q^2 R_g^2}{3}$$

où le rayon de giration R_g correspond à l'expression :

$$R_g^2 = \frac{\int_v \Delta\rho(\vec{r}) r^2 d\vec{r}}{\int_v \Delta\rho(\vec{r}) d\vec{r}}$$

Cette relation est valable en général pour des valeurs de qR_g inférieures à 1.3, mais pour des valeurs plus faibles (typiquement < 0.8) pour une molécule non structurée ou très allongée. Lorsque la courbe $\ln(I)$ en fonction de q^2 n'est pas linéaire, cela montre la présence d'agrégation dans la solution. $I(0)$ est proportionnelle à la masse de la particule. La détermination de $I(0)$ donne donc accès à l'état d'oligomérisation.

- La fonction de distribution des distances $P(r)$, qui est la transformée de Fourier de l'intensité diffusée $I(q)$ et qui est l'histogramme de toutes les distances r à l'intérieur de la protéine.

$$P(r) = TF(I(q)) = \frac{r^2}{2\pi^2} \int I(q) \frac{\sin qr}{qr} q^2 dq$$

La fonction de distribution des distances a une forme de gaussienne si la protéine est globulaire. Par contre elle est asymétrique si la protéine est allongée ou non structurée. La valeur maximale à laquelle $P(r)$ s'annule correspond à l'extension maximale D_{\max} de la molécule.

Le R_g et le $I(0)$ peuvent également être obtenus à partir de cette fonction par les relations suivantes, qui ont l'avantage de prendre en compte toute la courbe de diffusion et pas seulement la région de Guinier, plus sensible à la présence d'agrégats. :

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr} \quad I(0) = 4\pi \int_0^{D_{max}} P(r) dr$$

-La représentation de Kratky, $(qR_g)^2 I(q)/I(0)$ en fonction de qR_g , permet une estimation rapide de la forme de la molécule. Pour une protéine globulaire, la fonction correspond à une courbe en cloche et présente un maximum de 1,104 pour un qR_g de $\sqrt{3}$ quelle que soit la masse de la protéine étudiée. Dans le cas extrême des protéines non structurées la courbe croît indéfiniment.

5.2. Protocole expérimental

Le paramétrage de l'expérience est résumé dans le (Tableau 11). Le détecteur était placé à 2m de notre échantillon et l'énergie du faisceau réglé à 12.0 keV correspondant à une longueur d'onde de 1.033 Å.

	SarF apo	ScuR apo	SarF-BI7	ScuR-BI7
Data collection parameters				
Instrument	SWING (SOLEIL)			
Detector	EIGER 4M (Dextris)			
Wavelength [Å]	1.033204			
q-range [Å ⁻¹]	0.005 - 0.6			
Absolute scaling	Comparison with scattering from pure H ₂ O			
Exposure time [s]	2.99			
SEC-SAXS column	Superdex 200 Increase 10x300 (GE)			
Loading concentration [mM]	0.25	0.11	0.24	0.15
Injection volume [µL]	100	100	100	100
Flow rate [mL.min ⁻¹]	0.5			
Temperature [°C]	15			

Tableau 11 : Paramètres utilisés pour les mesures de SAXS.

Peu avant l'expérience, la protéine est décongelée, si besoin mélangée au peptide (que l'on met 10 fois en excès), centrifugée puis placée dans un rack à 10°C. Un robot automatisé se charge d'injecter la protéine dans une colonne de chromatographie d'exclusion de taille Superdex 200 10/300 (SEC) préalablement équilibrée dans le tampon SEC (20mM Tris HCl, pH8, 150mM NaCl, 1mM EDTA). La protéine est éluée à 0.5mL/min. Une cellule UV est placée entre la colonne de chromatographie et la cellule de SAXS, qui est un capillaire de 1.5mm de diamètre placé sous vide. La colonne ainsi que la cellule SAXS sont thermostatés à 15°C tout au long de l'expérience.

Les courbes de diffusion des rayons X sont collectées tout au long de l'éluion, à raison d'un frame toutes les 3 secondes. Les courbes mesurées avant le passage de la protéine correspondent à la diffusion du tampon. Elles sont moyennées et soustraites des courbes de protéines, afin d'obtenir le signal de diffusion de la protéine uniquement. On collecte typiquement 160 images par tampon et 700 images par échantillon. L'intensité de diffusion est obtenue en échelle absolue en normalisant par la diffusion de l'eau. Ce traitement initial du signal est fait à l'aide du logiciel FoxTrot, développé sur la ligne SWING.

Nous avons ensuite utilisé le module SAXS du programme US-SOMO (UltraScan Solution Modeler) pour extraire de chaque courbe de diffusion $I(q)$ les paramètres caractéristiques de l'échantillon. On peut ainsi suivre l'état d'oligomérisation ainsi que la taille de l'objet diffusant tout au long de l'éluion et déterminer, à partir de la loi de Guinier, le rayon de giration et l'intensité à l'origine $I(0)$ ainsi que la masse moléculaire, par la méthode du volume de corrélation. La seconde étape consiste à comparer les différentes courbes afin de moyennner les courbes identiques correspondant à chaque pic d'éluion. On obtient alors une courbe moyenne représentant notre molécule d'intérêt. C'est cette courbe $I(q)$ que l'on va analyser. L'étape suivante de l'analyse consiste à déterminer la fonction de distribution des distances $P(r)$, définie ci-dessus, en utilisant par exemple le programme GNOM (SVERGUN 1992), et donc à obtenir la dimension maximale de l'objet diffusant D_{max} .

5.3. Modélisation

Les étapes de l'analyse présentées ci-dessus sont modèles-indépendantes. Pour aller plus loin, on peut chercher à modéliser l'objet diffusant, en ayant recours à deux types de modélisation :

- i) la modélisation *ab initio* (à l'aide des programmes DAMMIN/DAMMIF (FRANKE &

SVERGUN 2009) ou GASBOR (SVERGUN *et al.* 2001) qui donne une enveloppe de la protéine, ou ii) la modélisation atomique, qui consiste à construire des modèles atomiques par homologie à partir de structures cristallographiques existantes, et à valider ces modèles atomiques à l'aide des données de SAXS.

Dans notre cas, la modélisation *ab initio* est peu informative, elle conduit à des enveloppes globalement globulaires, qui ne permettent pas de donner des informations sur le comportement de la protéine.

Nous avons donc eu recours à la modélisation tout-atome. Pour ce faire, différents modèles de ScuR et SarF ont été construits par homologie avec les structures de ComR grâce au logiciel SWISS-MODEL (WATERHOUSE *et al.* 2018). Les régions manquantes (par exemple le Strep-Tag C-terminal) ont été ajoutées manuellement avec le logiciel COOT (EMSLEY & COWTAN 2004). La conformation de ces régions, ainsi que le linker flexible entre le domaine HTH et le domaine TPR, a été ensuite affinée à l'aide du programme DADIMODO (RUDENKO *et al.* 2019), qui ajuste la courbe calculée sur le modèle tout-atome à la courbe expérimentale, en jouant sur les régions flexibles. Typiquement, une dizaine de modèles DADIMODO sont construits pour chaque protéine.

6. Cristallographie

6.1. Cristallogenèse

6.1.1. Principe de la cristallisation

La cristallogenèse consiste à trouver les conditions permettant de faire passer une protéine d'une forme soluble vers un état solide ordonné, le cristal. On utilise pour cela la méthode de diffusion en phase vapeur (Figure 90).

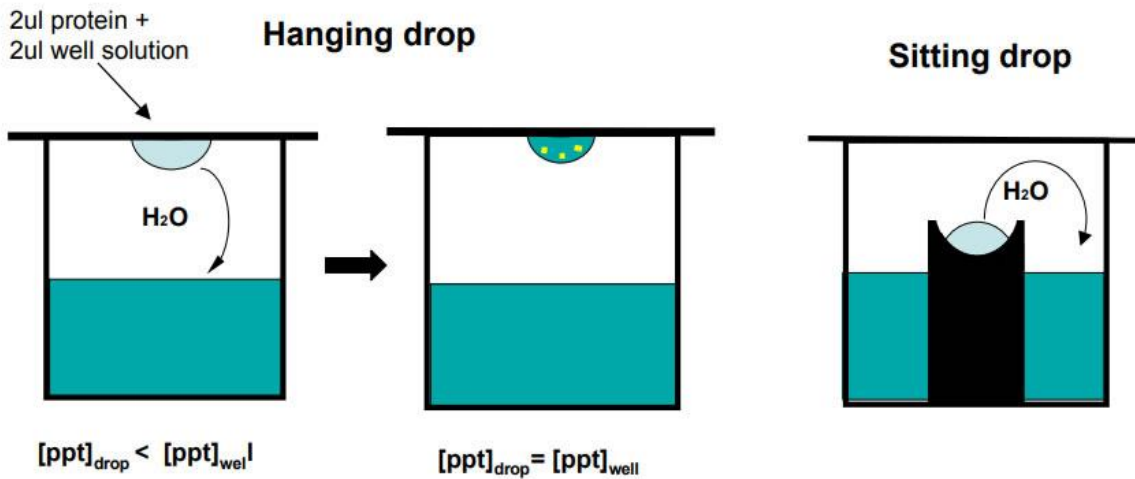


Figure 90 : Schéma d'une expérience de cristallisation en phase vapeur

Une goutte de protéine est mise à équilibrer au-dessus d'un réservoir contenant une solution d'agent précipitant. La goutte de protéine est diluée avec une goutte de la solution de cristallisation du réservoir. La concentration en agent précipitant est donc plus faible dans la goutte que dans le réservoir, ce qui génère une évaporation d'eau de la goutte vers le réservoir afin d'équilibrer les concentrations en agent précipitant. La diminution du volume de la goutte entraîne une augmentation des concentrations en agent précipitant et en protéine, et éventuellement à l'apparition de cristaux.

On peut décrire le comportement de la goutte de cristallisation par un diagramme de phase (Figure 91) qui montre comment cette technique de cristallisation permet à la protéine de dépasser son seuil de solubilité pour passer d'un état sous-saturée soluble vers une solution sursaturée permettant la formation de germes cristallins qui vont pouvoir grossir et former un cristal.

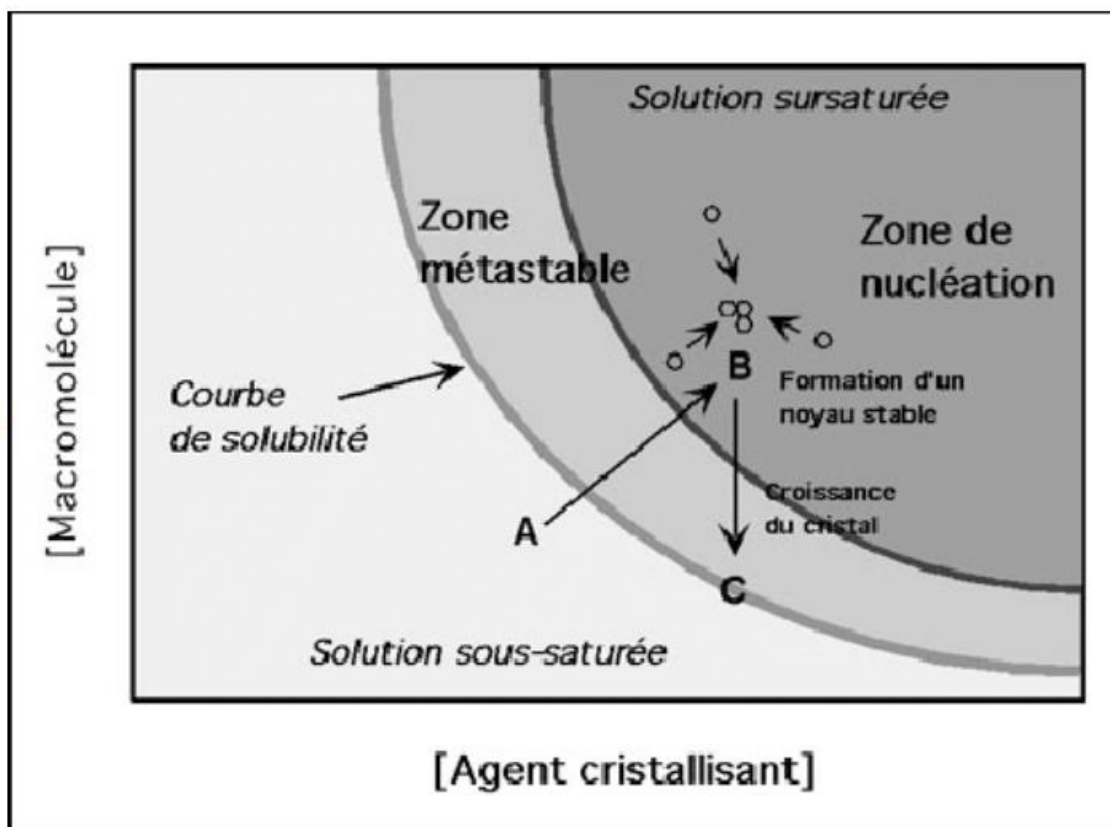


Figure 91 : Diagramme de phase type d'une expérience de cristallisation.

Les points A, B et C décrivent comment la protéine traverse la courbe de solubilité dans une expérience par diffusion en phase vapeur

6.1.2. Criblage des conditions de cristallisation

Les essais de cristallisation ont été réalisés sur la plateforme de l'I2BC (<https://www.i2bc.paris-saclay.fr/spip.php?article281>). On utilise des plaques de 96 puits à fond plat facilitant l'imagerie des gouttes (tpp LABTECH) pouvant équilibrer 3 gouttes par puits (Figure 92), ce qui permet de tester jusqu'à 288 conditions par boîte en faisant varier, par exemple, la concentration en protéine, la nature ou la concentration des ligands. La plateforme dispose d'une dizaine de kits commerciaux (QIAGEN) couvrant un large panel de conditions de cristallisation (nature et concentration en agent précipitant, pH, additifs...).



Figure 92 : Plaque ttp LABTECH triple sitting drop.

A droite, un zoom sur un puits et les 3 emplacements identiques permettant de déposer 3 gouttes.

J'ai utilisé le robot pipeteur TECAN (Figure 93) pour transférer les solutions des kits stockés en plaques de type Deepwell vers les puits des boîtes de cristallisation.



Figure 93 : Photo du robot pipeteur TECAN

J'ai ensuite utilisé les robots de cristallisation Cartesian ou Mosquito (Figure 94) pour effectuer le dépôt des gouttes de cristallisation.



Figure 94 : Robots de cristallisation.

Le robot Cartésien (photo de gauche) utilise 8 aiguilles et 8 corps de pompe pour faire des gouttes de 100 à 300nL de protéine + 100 à 300nL de solution du réservoir. Pour faire une plaque entière (96 conditions), il faut prévoir au minimum 25 μ L de protéine.

Le robot de cristallisation Mosquito (photo de droite) utilise des pointes jetables pour faire des gouttes de 50 à 200nL. Cela permet d'éviter toute contamination et le Mosquito présente également l'avantage de prélever au minimum 17 μ L d'échantillon pour 96 conditions. Néanmoins, son coût de fonctionnement est plus élevé

Les plaques contenant nos échantillons sont ensuite placées dans un robot de visualisation (Formulatrix) à 4°C ou 18°C. Le robot prend régulièrement une photo de chaque goutte, en suivant une suite de Fibonacci (0, 1, 2, 3, 5, 8, 13, 21 et 34 Jours) pendant un mois. Par la suite, il est possible de demander manuellement des photos au robot à la date que l'on souhaite. Les photos sont alors analysées grâce au logiciel CRIMS, qui permet d'annoter manuellement chaque goutte (Figure 95). Des observations fréquentes permettent de surveiller l'évolution des gouttes et, en cas d'apparition de cristaux, les congeler rapidement dans l'azote liquide, pour éviter qu'ils ne se dégradent au cours du temps.

The screenshot displays the CRIMS software interface. At the top, the IBBMC logo is visible. Below it, the user name 'H. Thuiller Jordan' is shown. The main area features a 96-well plate grid on the left, a large microscopic image of a crystal in the center, and a scoring schema on the bottom left. The grid shows a color-coded drop in well A4. The scoring schema lists categories from 0 (Clear drop) to 6 (Single crystal). A 'Set Score' panel with numeric buttons is located below the image. At the bottom, a sample information table is displayed.

Well	Salt	Buffer	pH	Main Precipitant	Component 2	Component 3	Component 4	final pH	Cryoprotectant	Reagent number	Supplier
A04		0.1 M sodium acetate	4.6	25 % (v/v) PEG 550 MHE						4	Qiagen/Vector

Figure 95 : Capture d'écran de l'interface du logiciel CRIMS

Il est possible d'annoter chaque goutte avec un panel de couleur afin de retrouver facilement les conditions les plus intéressantes.

6.1.3. Optimisation des cristaux

Lorsqu'aucune condition des kits ne permet d'obtenir un cristal suffisamment gros (environ 50 μ m de côté) et bien formé pour être utilisé en diffraction des rayons X, on affine les conditions les plus prometteuses, donnant des cristaux de mauvaise qualité ou trop petits, en faisant varier les différents paramètres de la condition de cristallisation (concentrations en protéine et en agent précipitant, pH, température, ajout d'additifs...). Ces optimisations peuvent être réalisées en nano-gouttes (avec les robots de cristallisation Cartesian ou Mosquito) en utilisant le robot pipeteur Hamilton (Figure 96) pour préparer des plaques de 96 puits à partir de solutions stock « maison » (agent cristallisant, tampon...).



Figure 96 : Robot de pipetage Hamilton.

J'ai également fait des optimisations manuelles sur des plaques Greiner de 24 puits en utilisant la technique de la goutte suspendue (Figure 97).

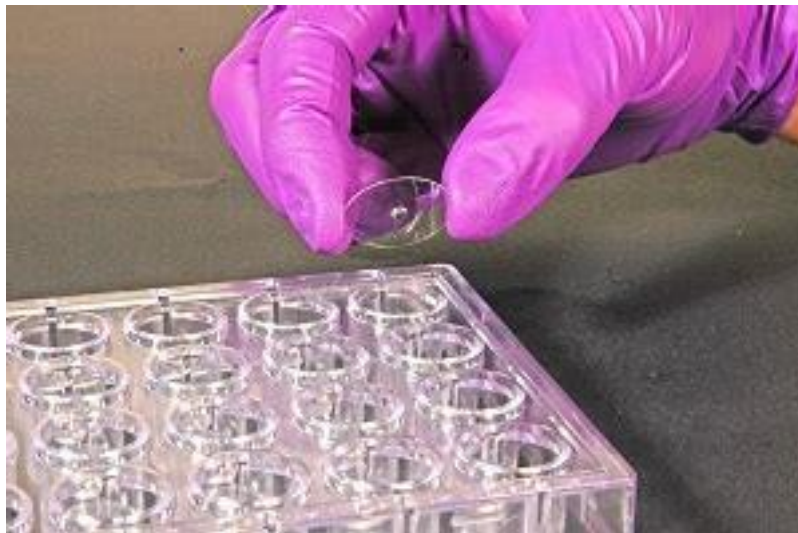


Figure 97 : Photo d'une plaque 24puits permettant de faire des optimisations manuelles.

La goutte, qui est un mélange entre la protéine et l'agent précipitant est placée sur une lamelle puis cette dernière est retournée pour être placée au-dessus d'un puits contenant de l'agent précipitant seul dans un système clos. Des échanges de vapeurs permettent de faire passer la protéine dans la zone métastable.

Les optimisations manuelles utilisent des volumes plus gros (500 μ L pour le puits et 1 μ L ou plus pour la goutte), ce qui permet parfois d'obtenir des cristaux plus gros. Les plaques de 24 puits ne sont pas prises en charge par le robot de visualisation et les gouttes doivent donc être observées régulièrement au microscope.

Lorsqu'après optimisation, seuls des micro-cristaux (ou des aiguilles très fines) ont pu être obtenus, j'ai utilisé la technique de micro-ensemencement pour essayer de les faire grossir. Des gouttes fraîchement équilibrées contre une solution d'agent précipitant légèrement en-dessous du seuil de cristallisation sont inoculées avec différentes dilutions (1/10, 1/100, 1/1000 typiquement) de germes cristallins broyés puis re suspendus dans la solution de cristallisation.

Les cristaux finalement obtenus sont transférés dans une solution cryoprotectante, en général la solution de cristallisation est complétée avec 25%-30% de glycérol, pour être congelés dans l'azote liquide où ils seront conservés jusqu'à ce qu'ils puissent être testés en diffraction aux rayons X.

6.2. Analyse par diffraction aux rayons X

6.2.1. Théorie de la diffraction

Un cristal est un solide qui résulte de l'empilement périodique d'une molécule dans les trois directions de l'espace (3D), ce qui forme un réseau. La maille élémentaire constitue l'unité de répétition de base d'un cristal. Elle correspond au plus petit élément permettant la construction du cristal par les translations du réseau. La maille du cristal est caractérisée par trois vecteurs $\{a, b \text{ et } c\}$ et trois angles (α, β, γ) qui définissent les paramètres de maille (Figure 98).

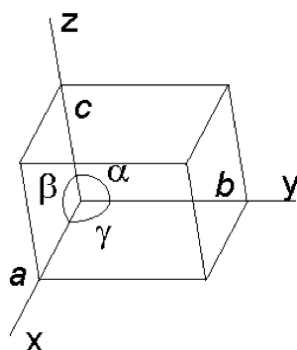


Figure 98 : Schéma d'une maille élémentaire.

Suivant les valeurs de ces 6 paramètres, il existe 7 systèmes cristallins (que l'on appelle également réseaux de Bravais) définissant la maille d'un cristal : réseau monoclinique, triclinique, orthorhombique, tétragonal, trigonal, hexagonal ou cubique. La maille peut être multiple et contenir plusieurs motifs équivalents par symétrie. On appelle alors unité asymétrique (UA) la portion de maille contenant ce motif. Il peut être composé d'une ou plusieurs molécules. On reconstitue la maille en appliquant au motif de base les symétries du cristal. Le système cristallin et les éléments de symétrie qui relient les UAs entre elles définissent ce qu'on appelle le groupe d'espace.

L'assemblage des molécules en un réseau cristallin régulier définit des plans réticulaires. Lorsqu'un cristal est placé dans un faisceau incident de longueur d'onde du même ordre de grandeur que la distance inter-réticulaire, c'est-à-dire l'Angström, cela donne lieu à un phénomène de diffraction. Chaque onde diffractée est générée par des interférences constructives entre les ondes diffusées par l'ensemble des plans réticulaires formés par le réseau cristallin (Figure 99).

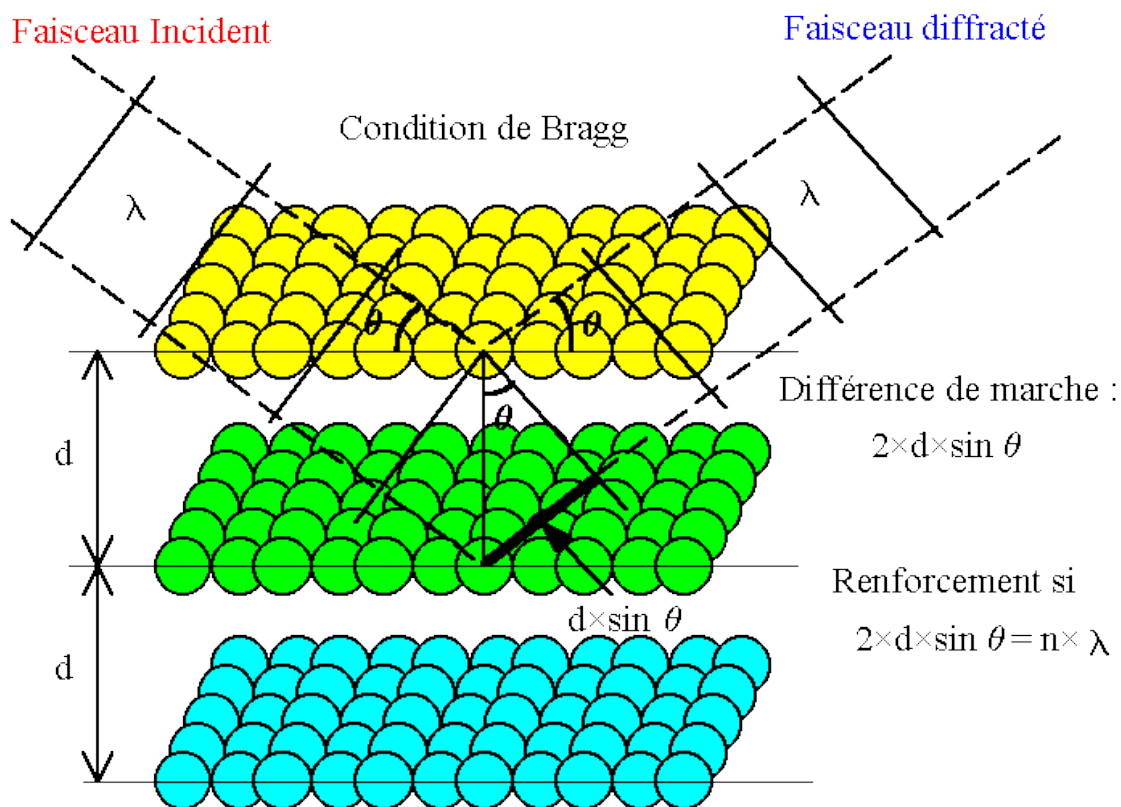


Figure 99 : Schéma expliquant le phénomène de diffraction au sein d'un réseau cristallin.

On n'a donc pas diffusion mais diffraction de la lumière dans des directions particulières de l'espace, définies par la loi de Bragg :

$$2 d \sin \theta = n \lambda$$

Où θ est l'angle d'incidence du faisceau incident, d la distance entre deux plans réticulaires et λ la longueur d'onde du faisceau de rayon X utilisé

Les ondes diffractées forment ce qu'on appelle le réseau réciproque que l'on peut visualiser en enregistrant leur impact sur un détecteur. On fait tourner le cristal sur lui-même dans le faisceau incident de façon à enregistrer l'ensemble des ondes diffractées. A chaque orientation du cristal correspond une image appelé cliché de diffraction (Figure 100).

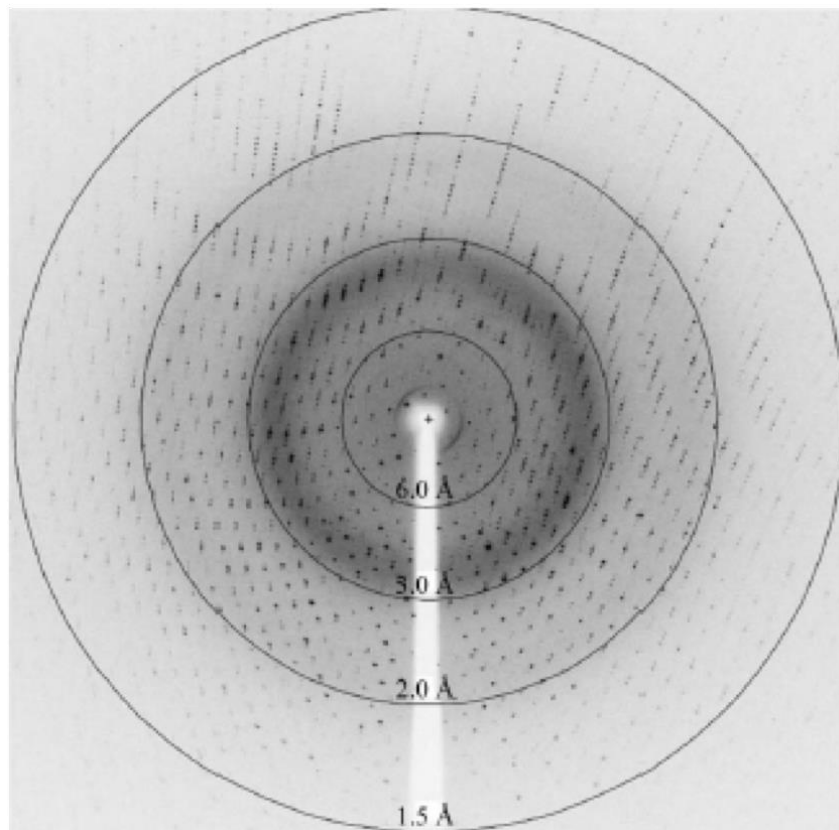


Figure 100 : Cliché de diffraction

En analysant la position et l'intensité des tâches de diffraction, on peut caractériser le réseau réciproque et ses éléments de symétrie, ce qui permet de déterminer le groupe d'espace auquel appartient le cristal.

Chaque onde diffractée est définie par un nombre complexe F , appelé facteur de structure:

$$F_{hkl} = |F_{hkl}| \exp(i\varphi)$$

Les indices hkl correspondent aux coordonnées de la tache de diffraction dans le réseau réciproque : $\mathbf{s} = h \mathbf{a}^* + k \mathbf{b}^* + l \mathbf{c}^*$

L'intensité de la tâche est égale au carré du module du facteur de structure, $|F_{hkl}|^2$

L'information sur la phase φ de l'onde, qui ne peut pas être déterminée par l'analyse des clichés de diffraction, est perdue.

Il existe heureusement différentes méthodes qui permettent d'estimer les phases $\varphi(hkl)$ qui, bien qu'approximatives, permettront de calculer, par transformée de Fourier inverse, la densité électronique ρ en tout point de coordonnées (xyz) du cristal :

$$\rho(x y z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h k l)| \exp[-2\pi i(hx + ky + lz) + i\varphi(h k l)]$$

On pourra alors, par analyse de la carte de densité électronique, construire la structure atomique de la protéine cristallisée (Figure 101):

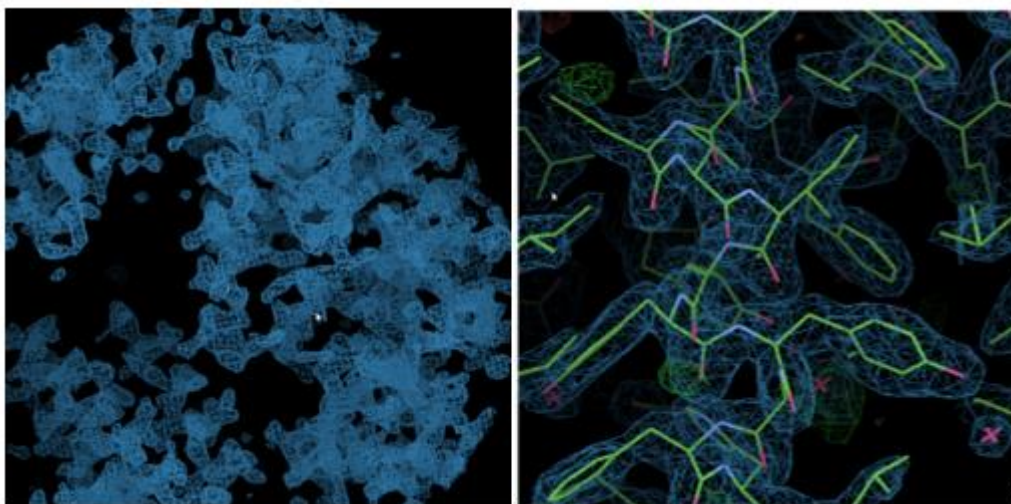


Figure 101 : Reconstruction d'une structure protéique dans sa carte de densité associée

A partir des positions atomiques ainsi identifiées ($\mathbf{r} = x \mathbf{a} + y \mathbf{b} + z \mathbf{c}$), il est possible de calculer, par transformée de Fourier, de nouveaux facteurs de structures $F(hkl)$:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

Chaque facteur de structure $F(hkl)$ correspond en effet à la composante de la transformée de Fourier de la densité électronique du cristal dans une direction donnée.

On obtient ainsi un nouveau jeu de phases et, par une procédure itérative, on améliore progressivement la qualité de la carte de densité et on affine les coordonnées atomiques de la protéine.

Les différentes techniques pouvant être utilisées pour retrouver l'information sur les phases sont :

- le remplacement isomorphe, basé sur la différence observée entre les réseaux réciproques obtenus à partir de deux cristaux isomorphes, un cristal dit natif et un cristal dans lequel un atome lourd a été introduit, par trempage ou co-cristallisation. Cette technique permet d'identifier les positions des atomes lourds dans la maille et d'obtenir une première estimation des phases des ondes générées par le cristal de la protéine.
- le remplacement moléculaire, basé sur le positionnement d'un modèle structural préexistant dans la maille du cristal étudié.
- la diffusion anormale, propriété de certains atomes qui perturbe la loi de Friedel, permet de retrouver leur position et donc de calculer des phases.

Au cours de cette étude, j'ai utilisé la méthode du remplacement moléculaire. Par superposition des cartes de Patterson générées par transformée de Fourier à partir, d'une part, des modules des facteurs de structure observés et, d'autre part, des coordonnées atomiques du modèle utilisé, on obtient une matrice de rotation et de translation qui permet de positionner le modèle dans l'unité asymétrique du cristal étudié. A partir de ces positions atomiques, des facteurs de structures peuvent être calculés, desquels on extrait l'information sur les phases, que l'on peut alors associer aux modules observés pour calculer une carte de densité électronique.

En corrigeant progressivement les positions atomiques, on peut recalculer des phases moins entachées d'erreurs et la carte de densité est de mieux en mieux définie. On s'aide pour cette procédure d'affinement des contraintes géométriques caractéristiques des protéines.

6.2.2. Démarche expérimentale

J'ai utilisé la suite de programme XDS (KABSCH 2010) pour traiter tous mes jeux de données. Dans un premier temps, le programme recherche sur les clichés de diffraction les taches les plus intenses et évalue le bruit de fond. À partir de la position des réflexions intenses identifiées, le programme propose des systèmes cristallins et des paramètres de mailles compatibles avec ces positions. Les solutions proposées sont classées en fonction d'une pénalité qui rend compte des écarts entre positions prédites et observées. On choisit le système cristallin de plus haute symétrie, tout en conservant la plus faible pénalité. Une fois la maille et le groupe d'espace identifiés, on peut attribuer à chaque tache de diffraction les indices hkl. L'étape d'intégration permet ensuite de mesurer l'intensité $I(h, k, l)$ de chaque tache.

Il est ensuite nécessaire de mettre à la même échelle ces intensités afin de compenser par exemple la variation éventuelle du faisceau incident ou la dégradation du cristal au cours de la collecte. On peut alors calculer la moyenne des réflexions équivalentes par symétrie et réduire le jeu de données complet à un jeu de réflexions uniques. Le facteur « R_{merge} » permet d'évaluer la déviation moyenne des réflexions équivalentes et d'alerter sur un éventuel problème de d'identification du groupe d'espace.

A cette étape, on va sélectionner la résolution maximale à laquelle on doit couper notre jeu de données. Pour cela, on se base sur plusieurs critères. Les données doivent être complètes à au moins 95%. Le rapport signal sur bruit (I/σ) doit être supérieur à 1 et le coefficient de corrélation (CC) doit être supérieur à 0.5.

Enfin, la dernière étape du traitement des données consiste à la conversion des intensités moyennées en modules observés $|F_{\text{hkl}}|_{\text{obs}}$.

Ayant déterminé la maille du réseau cristallin et sa multiplicité, on peut déterminer le nombre de molécules à placer dans chaque unité asymétrique. Cette étape a été réalisée à l'aide du programme MATTHEWS_COEF (KANTARDJIEFF & RUPP 2003) de la suite de programme CCP4 (WINN *et al.* 2011) qui permet de calculer le coefficient de Matthews (C_M) :

$$c_M = \frac{V}{M \cdot n \cdot Z_a}$$

avec :

V : Volume de la maille élémentaire calculé à partir des paramètres de maille

M : Masse moléculaire de la protéine

n : Nombre d'unités asymétriques par maille

Z_a : Nombre de molécule dans l'unité asymétrique

En se basant sur les statistiques observées dans les cristaux de protéine et en faisant varier Z_a , on détermine la valeur Z_a de plus forte probabilité, qui donne une estimation du nombre de molécules au sein de l'UA et donc de la maille.

Pour le phasage par remplacement moléculaire, j'ai utilisé les programmes Phaser (McCOY 2006) et Molrep (VAGIN & TEPLYAKOV 2010). Ces deux logiciels fonctionnent par comparaison des cartes de Patterson générées par transformée de Fourier des modules des facteurs de structure, et qui représentent, par des pics de positions et d'intensités variables, tous les vecteurs intramoléculaires et intermoléculaires de la molécule. Le programme recherche le maximum de corrélation entre la carte de Patterson du modèle et du cristal en faisant varier l'orientation et la position du modèle. On détermine ainsi la meilleure matrice de superposition (rotation + translation).

Lorsqu'une solution est trouvée, des phases sont calculées à partir des coordonnées atomiques du modèle correctement positionné dans la nouvelle maille et deux cartes de densité sont calculées :

- La carte $2F_o-F_c$ qui représente la localisation des atomes du modèle au sein de la maille
- La carte de différence F_o-F_c , qui permet de montrer les erreurs dans le modèle sous forme de pics positifs (en vert) si des atomes sont manquants et négatifs (en rouge) s'il y a des atomes dans le modèle qui ne correspondent à aucune densité.

L'analyse visuelle de ces deux cartes ainsi que différents scores fournis par le programme de remplacement moléculaire permettent de reconstruire le modèle de notre protéine.

6.3. Affinement du modèle atomique

L'affinement des différentes structures résolues au cours de cette étude a été réalisé grâce au logiciel PHENIX-Refine (ADAMS *et al.* 2010). Cette étape permet non seulement d'affiner les coordonnées atomiques du modèle, mais aussi pour chaque atome, si la résolution le permet, son facteur de température (facteur B) qui représente le mouvement d'un atome autour de sa position d'équilibre et le taux d'occupation de la position.

A chaque cycle d'affinement, on vérifie manuellement que le modèle coïncide bien avec la carte de densité électronique. Pour cela, j'ai utilisé le logiciel graphique COOT (Crystallographic Oriented Object Toolkit) (EMSLEY & COWTAN 2004) qui permet de superposer les deux cartes de densité électronique (2Fo-Fc) et (Fo-Fc) avec le modèle atomique, et qui propose de nombreuses fonctionnalités permettant d'optimiser manuellement le modèle en fonction des cartes de densité électronique.

Le modèle corrigé et si besoin complété manuellement est réinjecté dans le programme PHENIX-refine qui calcul à chaque cycle de cette procédure itérative, deux facteurs d'accord appelés R_{free} et R_{work} :

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

Ce facteur mesure la différence entre les modules des facteurs de structure calculés à partir des positions atomiques du modèle et les modules des facteurs de structure mesurés expérimentalement. Ce facteur R doit être d'autant plus bas que la résolution est bonne et que les atomes peuvent être placés précisément. Pour l'affinement du modèle et du calcul des cartes on soustrait un petit pourcentage de taches du jeu de données. Les F_{calc} correspondants ne sont donc pas biaisés par le modèle. Le R_{work} est calculé sur les données utilisées pour l'affinement alors que le R_{free} est calculé sur la sous-partie extraite de l'affinement. Ainsi, si les deux facteurs R_{free} et R_{work} diminuent au cours des cycles d'affinement, on peut en déduire que notre modèle s'améliore. Lorsque ceux-ci ont atteint des valeurs assez basses (de l'ordre de 20% à haute résolution) et qu'elles stagnent, on estime que l'affinement est terminé.

Tout au long du processus de reconstruction et d'affinement, le logiciel MOLPROBITY (DAVIS *et al.* 2004) implanté dans PHENIX permet de mesurer différents paramètres stéréochimiques que chaque atome au sein de notre modèle doit respecter :

- Le diagramme de Ramachandran
- Les rotamères préférentiels
- Les clashes
- La chiralité
- La planarité
- La corrélation entre le modèle et la densité électronique

Ainsi tout atome ne respectant pas l'un de ces paramètres est signalé et il est possible via le logiciel COOT de le reconstruire.

6.3.1. Dépôt de la structure dans la PDB

Enfin, j'ai participé au dépôt des structures dans la Protein Data Bank (PDB). Après validation de la qualité du jeu de données et du modèle sur le PDB-Validation Server on indique toutes les caractéristiques de la structure (nom, fonction, origine de la ou les molécules, structure quaternaire, conditions de cristallisation etc...). Ces indications permettront d'établir l'entête du fichier PDB. La comparaison des statistiques de notre modèle aux structures déjà déposées dans la banque ayant une résolution semblable à la nôtre, ce qui permet de juger de la qualité de notre modèle.

V. Références bibliographique

- ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. (2010): PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**(Pt 2): 213–221. <https://doi.org/10.1107/S0907444909052925>
- AGAISSE, H., GOMINET, M., ØKSTAD, O. A., KOLSTØ, A.-B. & LERECLUS, D. (1999): PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Molecular Microbiology* **32**(5): 1043–1053. <https://doi.org/10.1046/j.1365-2958.1999.01419.x>
- ALI, L., GORAYA, M. U., ARAFAT, Y., AJMAL, M., CHEN, J.-L. & YU, D. (2017): Molecular Mechanism of Quorum-Sensing in *Enterococcus faecalis*: Its Role in Virulence and Therapeutic Approaches. *International Journal of Molecular Sciences* **18**(5). <https://doi.org/10.3390/ijms18050960>
- ARAVIND, L., ANANTHARAMAN, V., BALAJI, S., BABU, M. & IYER, L. (2005): The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* **29**(2): 231–262. <https://doi.org/10.1016/j.femsre.2004.12.008>
- AUCHTUNG, J. M., LEE, C. A., MONSON, R. E., LEHMAN, A. P. & GROSSMAN, A. D. (2005): Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proceedings of the National Academy of Sciences of the United States of America* **102**(35): 12554–12559. <https://doi.org/10.1073/pnas.0505835102>
- BAKER, M. D. & NEIDITCH, M. B. (2011): Structural Basis of Response Regulator Inhibition by a Bacterial Anti-Activator Protein. STOCK, A. (ed). – *PLoS Biology* **9**(12): e1001226. <https://doi.org/10.1371/journal.pbio.1001226>
- BASSLER, B. L. & LOSICK, R. (2006): Bacterially speaking. *Cell* **125**(2): 237–246. <https://doi.org/10.1016/j.cell.2006.04.001>
- BLATCH, G. L. & LÄSSLE, M. (1999): The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* **21**(11): 932–939. [https://doi.org/10.1002/\(SICI\)1521-1878\(199911\)21:11<932::AID-BIES5>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N)
- BOUILLAUT, L., PERCHAT, S., AROLD, S., ZORRILLA, S., SLAMTI, L., HENRY, C., GOHAR, M., DECLERCK, N. & LERECLUS, D. (2008): Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. *Nucleic Acids Research* **36**(11): 3791–3801. <https://doi.org/10.1093/nar/gkn149>

- BROOKES, E., VACHETTE, P., ROCCO, M. & PÉREZ, J. (2016): US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of Applied Crystallography* **49**(Pt 5): 1827–1841. <https://doi.org/10.1107/S1600576716011201>
- CARAPETIS, J. R., STEER, A. C., MULHOLLAND, E. K. & WEBER, M. (2005): The global burden of group A streptococcal diseases. *The Lancet Infectious Diseases* **5**(11): 685–694. [https://doi.org/10.1016/S1473-3099\(05\)70267-X](https://doi.org/10.1016/S1473-3099(05)70267-X)
- CHANDLER, J. R. & DUNNY, G. M. (2004): Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides* **25**(9): 1377–1388. <https://doi.org/10.1016/j.peptides.2003.10.020>
- CHOI, J., SHIN, D., KIM, M., PARK, J., LIM, S. & RYU, S. (2012): LsrR-Mediated Quorum Sensing Controls Invasiveness of Salmonella typhimurium by Regulating SPI-1 and Flagella Genes. *PLoS ONE* **7**(5): e37059. <https://doi.org/10.1371/journal.pone.0037059>
- CHOI, S. H. & GREENBERG, E. P. (1991): The C-terminal region of the Vibrio fischeri LuxR protein contains an inducer-independent lux gene activating domain. *Proceedings of the National Academy of Sciences of the United States of America* **88**(24): 11115–11119. <https://doi.org/10.1073/pnas.88.24.11115>
- CHURCHILL, M. E. A. & CHEN, L. (2011): Structural Basis of Acyl-homoserine Lactone-Dependent Signaling. *Chemical Reviews* **111**(1): 68–85. <https://doi.org/10.1021/cr1000817>
- CLARKE, J. K. ON THE BACTERIAL FACTOR IN THE ÆTIOLOGY OF DENTAL CARIES. 9.
- COOK, L. C. & FEDERLE, M. J. (2014): Peptide pheromone signaling in Streptococcus and Enterococcus. *FEMS microbiology reviews* **38**(3): 473–492. <https://doi.org/10.1111/1574-6976.12046>
- CORE, L. & PEREGO, M. (2003): TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in Bacillus subtilis. *Molecular Microbiology* **49**(6): 1509–1522. <https://doi.org/10.1046/j.1365-2958.2003.03659.x>
- DAVIS, I. W., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. (2004): MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research* **32**(Issue suppl_2): W615–W619. <https://doi.org/10.1093/nar/gkh398>
- DAY, W. A. & MAURELLI, A. T. (2001): Shigella flexneri LuxS Quorum-Sensing System Modulates virB Expression but Is Not Essential for Virulence. *Infection and Immunity* **69**(1): 15–23. <https://doi.org/10.1128/IAI.69.1.15-23.2001>
- DECLERCK, N., BOUILLAUT, L., CHAIX, D., RUGANI, N., SLAMTI, L., HOH, F., LERECLUS, D. & AROLD, S. T. (2007): Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **104**(47): 18490–18495. <https://doi.org/10.1073/pnas.0704501104>

- DEFOIRDT, T., BOON, N. & BOSSIER, P. (2010): Can Bacteria Evolve Resistance to Quorum Sensing Disruption? *PLoS Pathogens* **6**(7): e1000989. <https://doi.org/10.1371/journal.ppat.1000989>
- DESAI, K., MASHBURN-WARREN, L., FEDERLE, M. J. & MORRISON, D. A. (2012): Development of Competence for Genetic Transformation of *Streptococcus mutans* in a Chemically Defined Medium. *Journal of Bacteriology* **194**(15): 3774–3780. <https://doi.org/10.1128/JB.00337-12>
- DIGGLE, S. P., GARDNER, A., WEST, S. A. & GRIFFIN, A. S. (2007): Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**(1483): 1241–1249. <https://doi.org/10.1098/rstb.2007.2049>
- DOU, Y., SONG, F., GUO, F., ZHOU, Z., ZHU, C., XIANG, J. & HUAN, J. (2017): *Acinetobacter baumannii* quorum-sensing signalling molecule induces the expression of drug-resistance genes. *Molecular Medicine Reports* **15**(6): 4061–4068. <https://doi.org/10.3892/mmr.2017.6528>
- DUBOIS, T., FAEGRI, K., PERCHAT, S., LEMY, C., BUISSON, C., NIELSEN-LEROUX, C., GOHAR, M., JACQUES, P., RAMARAO, N., KOLSTØ, A.-B. & LERECLUS, D. (2012): Necrotrophism Is a Quorum-Sensing-Regulated Lifestyle in *Bacillus thuringiensis*. KOEHLER, T.M. (ed). – *PLoS Pathogens* **8**(4): e1002629. <https://doi.org/10.1371/journal.ppat.1002629>
- EMSLEY, P. & COWTAN, K. (2004): *Coot*: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **60**(12): 2126–2132. <https://doi.org/10.1107/S0907444904019158>
- FALSETTA, M. L., STEICHEN, C. T., MCEWAN, A. G., CHO, C., KETTERER, M., SHAO, J., HUNT, J., JENNINGS, M. P. & APICELLA, M. A. (2011): The Composition and Metabolic Phenotype of *Neisseria gonorrhoeae* Biofilms. *Frontiers in Microbiology* **2**(75). <https://doi.org/10.3389/fmicb.2011.00075>
- FAST, W. & TIPTON, P. A. (2012): The enzymes of bacterial census and censorship. *Trends in Biochemical Sciences* **37**(1): 7–14. <https://doi.org/10.1016/j.tibs.2011.10.001>
- FEDERLE, M. J. (2009): Autoinducer-2-Based Chemical Communication in Bacteria: Complexities of Interspecies Signaling. *Contributions to microbiology* **16**: 18–32. <https://doi.org/10.1159/000219371>
- FETZNER, S. (2015): Quorum quenching enzymes. *Journal of Biotechnology* **201**: 2–14. <https://doi.org/10.1016/j.jbiotec.2014.09.001>
- FLEUCHOT, B., GITTON, C., GUILLOT, A., VIDIC, J., NICOLAS, P., BESSET, C., FONTAINE, L., HOLS, P., LEBLOND-BOURGET, N., MONNET, V. & GARDAN, R. (2011): Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Molecular Microbiology* **80**(4): 1102–1119. <https://doi.org/10.1111/j.1365-2958.2011.07633.x>
- FONTAINE, L., BOUTRY, C., FRAHAN, M. H. de, DELPLACE, B., FREMAUX, C., HORVATH, P., BOYAVAL, P. & HOLS, P. (2010): A Novel Pheromone Quorum-Sensing System Controls the Development of Natural Competence in *Streptococcus thermophilus* and

- Streptococcus salivarius*. *Journal of Bacteriology* **192**(5): 1444–1454.
<https://doi.org/10.1128/JB.01251-09>
- FONTAINE, L., DANDOY, D., BOUTRY, C., DELPLACE, B., FRAHAN, M. H. DE, FREMAUX, C., HORVATH, P., BOYAVAL, P. & HOLS, P. (2010): Development of a Versatile Procedure Based on Natural Transformation for Marker-Free Targeted Genetic Modification in *Streptococcus thermophilus*. *Applied and Environmental Microbiology* **76**(23): 7870–7877. <https://doi.org/10.1128/AEM.01671-10>
- FONTAINE, L., GOFFIN, P., DUBOUT, H., DELPLACE, B., BAULARD, A., LECAT-GUILLET, N., CHAMBELLON, E., GARDAN, R. & HOLS, P. (2013): Mechanism of competence activation by the ComRS signalling system in streptococci: ComRS-dependent activation of competence. *Molecular Microbiology* **87**(6): 1113–1132. <https://doi.org/10.1111/mmi.12157>
- FRANKE, D. & SVERGUN, D. I. (2009): *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**(2): 342–346. <https://doi.org/10.1107/S0021889809000338>
- FUQUA, C. & GREENBERG, E. P. (2002): Listening in on bacteria: acyl-homoserine lactone signalling. *Nature Reviews Molecular Cell Biology* **3**(9): 685–695. <https://doi.org/10.1038/nrm907>
- FUQUA, W. C. & WINANS, S. C. (1994): A LuxR-LuxI type regulatory system activates *Agrobacterium Ti* plasmid conjugal transfer in the presence of a plant tumor metabolite. *Journal of Bacteriology* **176**(10): 2796–2806.
- GALLEGO DEL SOL, F. & MARINA, A. (2013): Structural Basis of Rap Phosphatase Inhibition by Phr Peptides. LAUB, M.T. (ed). – *PLoS Biology* **11**(3): e1001511. <https://doi.org/10.1371/journal.pbio.1001511>
- GARAULT, P., BARS, D. L., BESSET, C. & MONNET, V. (2002): Three Oligopeptide-binding Proteins Are Involved in the Oligopeptide Transport of *Streptococcus thermophilus*. *Journal of Biological Chemistry* **277**(1): 32–39. <https://doi.org/10.1074/jbc.M107002200>
- GARDAN, R., BESSET, C., GITTON, C., GUILLOT, A., FONTAINE, L., HOLS, P. & MONNET, V. (2013): Extracellular Life Cycle of ComS, the Competence-Stimulating Peptide of *Streptococcus thermophilus*. *Journal of Bacteriology* **195**(8): 1845–1855. <https://doi.org/10.1128/JB.02196-12>
- GOHAR, M., FAEGRI, K., PERCHAT, S., RAVNUM, S., ØKSTAD, O. A., GOMINET, M., KOLSTØ, A.-B. & LERECLUS, D. (2008): The PlcR Virulence Regulon of *Bacillus cereus*. YANG, C.-H. (ed). – *PLoS ONE* **3**(7): e2793. <https://doi.org/10.1371/journal.pone.0002793>
- GRAY, B., HALL, P. & GRESHAM, H. (2013): Targeting agr- and agr-Like Quorum Sensing Systems for Development of Common Therapeutics to Treat Multiple Gram-Positive Bacterial Infections. *Sensors (Basel, Switzerland)* **13**(4): 5130–5166. <https://doi.org/10.3390/s130405130>
- GRENHA, R., SLAMTI, L., NICAISE, M., REFES, Y., LERECLUS, D. & NESSLER, S. (2013): Structural basis for the activation mechanism of the PlcR virulence regulator by the

- quorum-sensing signal peptide PapR. *Proceedings of the National Academy of Sciences of the United States of America* **110**(3): 1047–1052. <https://doi.org/10.1073/pnas.1213770110>
- GROSS, R., ARICÒ, B. & RAPPUOLI, R. (1989): Families of bacterial signal-transducing proteins. *Molecular Microbiology* **3**(11): 1661–1667. <https://doi.org/10.1111/j.1365-2958.1989.tb00152.x>
- HAUDECOEUR, E. & FAURE, D. (2010): A fine control of quorum-sensing communication in *Agrobacterium tumefaciens*. *Communicative & Integrative Biology* **3**(2): 84–88. <https://doi.org/10.4161/cib.3.2.10429>
- HÅVARSTEIN, L. S., COOMARASWAMY, G. & MORRISON, D. A. (1995): An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America* **92**(24): 11140–11144. <https://doi.org/10.1073/pnas.92.24.11140>
- HOOSHANGI, S. & BENTLEY, W. E. (2008): From unicellular properties to multicellular behavior: bacteria quorum sensing circuitry and applications. *Current Opinion in Biotechnology* **19**(6): 550–555. <https://doi.org/10.1016/j.copbio.2008.10.007>
- HUILLET, E., TEMPELAARS, M. H., ANDRÉ-LEROUX, G., WANAPAIAN, P., BRIDOUX, L., MAKHZAMI, S., PANBANGRED, W., MARTIN-VERSTRAETE, I., ABEE, T. & LERECLUS, D. (2012): PlcRa, a New Quorum-Sensing Regulator from *Bacillus cereus*, Plays a Role in Oxidative Stress Responses and Cysteine Metabolism in Stationary Phase. *PLOS ONE* **7**(12): e51047. <https://doi.org/10.1371/journal.pone.0051047>
- Ji, G., BEAVIS, R. C. & NOVICK, R. P. (1995): Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proceedings of the National Academy of Sciences of the United States of America* **92**(26): 12055–12059. <https://doi.org/10.1073/pnas.92.26.12055>
- KABSCH, W. (2010): XDS. *Acta Crystallographica Section D: Biological Crystallography* **66**(Pt 2): 125–132. <https://doi.org/10.1107/S0907444909047337>
- KANTARDJIEFF, K. A. & RUPP, B. (2003): Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science: A Publication of the Protein Society* **12**(9): 1865–1871. <https://doi.org/10.1110/ps.0350503>
- KATO, N., MOROHOSHI, T., NOZAWA, T., MATSUMOTO, H. & IKEDA, T. (2006): Control of Gram-Negative Bacterial Quorum Sensing with Cyclodextrin Immobilized Cellulose Ether Gel. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **56**(1–2): 55–59. <https://doi.org/10.1007/s10847-006-9060-y>
- KHAN, R., RUKKE, H. V., FILHO, A. P. R., FIMLAND, G., ARNTZEN, M. Ø., THIEDE, B. & PETERSEN, F. C. (2012): Extracellular Identification of a Processed Type II ComR/ComS Pheromone of *Streptococcus mutans*. *Journal of Bacteriology* **194**(15): 3781–3788. <https://doi.org/10.1128/JB.00624-12>

- KIM, M. K., ZHAO, A., WANG, A., BROWN, Z. Z., MUIR, T. W., STONE, H. A. & BASSLER, B. L. (2017): Surface-Attached Molecules Control *Staphylococcus aureus* Quorum Sensing and Biofilm Development. *Nature microbiology* **2**: 17080. <https://doi.org/10.1038/nmicrobiol.2017.80>
- KOZLOWICZ, B. K., SHI, K., GU, Z.-Y., OHLENDORF, D. H., EARHART, C. A. & DUNNY, G. M. (2006): Molecular basis for control of conjugation by bacterial pheromone and inhibitor peptides. *Molecular microbiology* **62**(4): 958–969. <https://doi.org/10.1111/j.1365-2958.2006.05434.x>
- KRISSINEL, E. & HENRICK, K. (2007): Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **372**(3): 774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- KROL, E. & BECKER, A. (2014): Rhizobial homologs of the fatty acid transporter FadL facilitate perception of long-chain acyl-homoserine lactone signals. *Proceedings of the National Academy of Sciences of the United States of America* **111**(29): 10702–10707. <https://doi.org/10.1073/pnas.1404929111>
- KUAN, G., DASSA, E., SAURIN, W., HOFNUNG, M. & SAIER, M. H. (1995): Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases. *Research in Microbiology* **146**(4): 271–278. [https://doi.org/10.1016/0923-2508\(96\)81050-3](https://doi.org/10.1016/0923-2508(96)81050-3)
- LANDMAN, C. (2017): Implications des N-acyl homosérine lactones, molécules du quorum sensing dans les maladies inflammatoires chroniques intestinales. Microbiologie et Parasitologie. Université Pierre et Marie Curie - Paris VI, France : Paris.
- LAU, Y. Y., SULAIMAN, J., CHEN, J. W., YIN, W.-F. & CHAN, K.-G. (2013): Quorum Sensing Activity of *Enterobacter asburiae* Isolated from Lettuce Leaves. *Sensors (Basel, Switzerland)* **13**(10): 14189–14199. <https://doi.org/10.3390/s131014189>
- LE BERRE, R., FAURE, K., NGUYEN, S., PIERRE, M., ADER, F. & GUERY, B. (2006): Quorum sensing : une nouvelle cible thérapeutique pour *Pseudomonas aeruginosa*. *Médecine et Maladies Infectieuses* **36**(7): 349–357. <https://doi.org/10.1016/j.medmal.2006.01.008>
- LERECLUS, D., AGAISSE, H., GOMINET, M., SALAMITOU, S. & SANCHIS, V. (1996): Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. *Journal of Bacteriology* **178**(10): 2749–2756. <https://doi.org/10.1128/jb.178.10.2749-2756.1996>
- LUPP, C. & RUBY, E. G. (2005): *Vibrio fischeri* Uses Two Quorum-Sensing Systems for the Regulation of Early and Late Colonization Factors. *Journal of Bacteriology* **187**(11): 3620–3629. <https://doi.org/10.1128/JB.187.11.3620-3629.2005>
- LUPP, C., URBANOWSKI, M., GREENBERG, E. P. & RUBY, E. G. (2003): The *Vibrio fischeri* quorum-sensing systems *ain* and *lux* sequentially induce luminescence gene expression and are important for persistence in the squid host. *Molecular Microbiology* **50**: 319–331. <https://doi.org/10.1046/j.1365-2958.2003.t01-1-03585.x>

- LYTE, M. (2016): Microbial Endocrinology in the Pathogenesis of Infectious Disease. Pp. 137–168. *In*: KUDVA, I.T., CORNICK, N.A., PLUMMER, P.J., ZHANG, Q., NICHOLSON, T.L., BANNANTINE, J.P., & BELLAIRE, B.H. (eds): *Virulence Mechanisms of Bacterial Pathogens*, Fifth Edition. 5th ed. American Society of Microbiology, USA : Washington DC.
- MAMISHI, S., MORADKHANI, S., MAHMOUDI, S., HOSSEINPOUR-SADEGHI, R. & POURAKBARI, B. (2014): Penicillin-Resistant trend of *Streptococcus pneumoniae* in Asia: A systematic review. *Iranian Journal of Microbiology* **6**(4): 198–210.
- MARKS, L. R., MASHBURN-WARREN, L., FEDERLE, M. J. & HAKANSSON, A. P. (2014): *Streptococcus pyogenes* biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. *The Journal of Infectious Diseases* **210**(1): 25–34. <https://doi.org/10.1093/infdis/jiu058>
- MARTIN, B., QUENTIN, Y., FICHANT, G. & CLAVERYS, J.-P. (2006): Independent evolution of competence regulatory cascades in streptococci? *Trends in Microbiology* **14**(8): 339–345. <https://doi.org/10.1016/j.tim.2006.06.007>
- MASHBURN-WARREN, L., MORRISON, D. A. & FEDERLE, M. J. (2010): A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. *Molecular Microbiology* **78**(3): 589–606. <https://doi.org/10.1111/j.1365-2958.2010.07361.x>
- MCCOY, A. J. (2006): Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography* **63**(Pt 1): 32–41. <https://doi.org/10.1107/S0907444906045975>
- MIGNOLET, J., CERCKEL, G., DAMOCZI, J., LEDESMA-GARCIA, L., SASS, A., COENYE, T., NESSLER, S. & HOLS, P. (2019): Subtle selectivity in a pheromone sensor triumvirate desynchronizes competence and predation in a human gut commensal. BRAKHAGE, A.A., STORZ, G., & FEDERLE, M. (eds). – *eLife* **8**: e47139. <https://doi.org/10.7554/eLife.47139>
- MIGNOLET, J., FONTAINE, L., SASS, A., NANNAN, C., MAHILLON, J., COENYE, T. & HOLS, P. (2018): Circuitry Rewiring Directly Couples Competence to Predation in the Gut Dweller *Streptococcus salivarius*. *Cell Reports* **22**(7): 1627–1638. <https://doi.org/10.1016/j.celrep.2018.01.055>
- MILLER, M. B. & BASSLER, B. L. (2001): QUORUM SENSING IN BACTERIA. 36.
- MION, S., REMY, B., PLENER, L., CHABRIERE, É. & DAUDE, D. (2019): Quorum sensing et quorum quenching : Comment bloquer la communication des bactéries pour inhiber leur virulence ? *médecine/sciences* **35**(1): 31–38. <https://doi.org/10.1051/medsci/2018310>
- MONNET, V., JUILLARD, V. & GARDAN, R. (2014): Peptide conversations in Gram-positive bacteria. *Critical Reviews in Microbiology* 339–351. <https://doi.org/10.3109/1040841X.2014.948804>
- MORRISON, D. A., GUÉDON, E. & RENAULT, P. (2013): Competence for Natural Genetic Transformation in the *Streptococcus bovis* Group *Streptococci* *S. infantarius* and *S.*

- macedonicus. *Journal of Bacteriology* **195**(11): 2612–2620. <https://doi.org/10.1128/JB.00230-13>
- NEALSON, K. H. & HASTINGS, J. W. (1979): Bacterial bioluminescence: its control and ecological significance. *Microbiological Reviews* **43**(4): 496–518.
- NG, W.-L. & BASSLER, B. L. (2009): Bacterial Quorum-Sensing Network Architectures. *Annual review of genetics* **43**: 197–222. <https://doi.org/10.1146/annurev-genet-102108-134304>
- NISHIGUCHI, K., NAGATA, K., TANOKURA, M., SONOMOTO, K. & NAKAYAMA, J. (2009): Structure-Activity Relationship of Gelatinase Biosynthesis-Activating Pheromone of *Enterococcus faecalis*. *Journal of Bacteriology* **191**(2): 641–650. <https://doi.org/10.1128/JB.01029-08>
- OGURA, M. & FUJITA, Y. (2007): *Bacillus subtilis rapD*, a direct target of transcription repression by RghR, negatively regulates *srfA* expression. *FEMS Microbiology Letters* **268**(1): 73–80. <https://doi.org/10.1111/j.1574-6968.2006.00559.x>
- OGURA, M., SHIMANE, K., ASAI, K., OGASAWARA, N. & TANAKA, T. (2003): Binding of response regulator DegU to the *aprE* promoter is inhibited by RapG, which is counteracted by extracellular PhrG in *Bacillus subtilis*. *Molecular Microbiology* **49**(6): 1685–1697. <https://doi.org/10.1046/j.1365-2958.2003.03665.x>
- PARASHAR, V., JEFFREY, P. D. & NEIDITCH, M. B. (2013): Conformational Change-Induced Repeat Domain Expansion Regulates Rap Phosphatase Quorum-Sensing Signal Receptors. *PLoS Biology* **11**(3): e1001512. <https://doi.org/10.1371/journal.pbio.1001512>
- PARASHAR, V., MIROUZE, N., DUBNAU, D. A. & NEIDITCH, M. B. (2011): Structural Basis of Response Regulator Dephosphorylation by Rap Phosphatases. *PLoS Biology* **9**(2): e1000589. <https://doi.org/10.1371/journal.pbio.1000589>
- PARK, J., JAGASIA, R., KAUFMANN, G. F., MATHISON, J. C., RUIZ, D. I., MOSS, J. A., MEIJLER, M. M., ULEVITCH, R. J. & JANDA, K. D. (2007): Infection Control by Antibody Disruption of Bacterial Quorum Sensing Signaling. *Chemistry & Biology* **14**(10): 1119–1127. <https://doi.org/10.1016/j.chembiol.2007.08.013>
- PATTERSON, A. L. (2015): A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Kristallographie - Crystalline Materials* **90**(1–6): 517–542. <https://doi.org/10.1524/zkri.1935.90.1.517>
- PEARSON, J. P., VAN DELDEN, C. & IGLEWSKI, B. H. (1999): Active Efflux and Diffusion Are Involved in Transport of *Pseudomonas aeruginosa* Cell-to-Cell Signals. *Journal of Bacteriology* **181**(4): 1203–1210.
- PELLEGRINI-CALACE, M. & THORNTON, J. M. (2005): Detecting DNA-binding helix–turn–helix structural motifs using sequence and structure information. *Nucleic Acids Research* **33**(7): 2129–2140. <https://doi.org/10.1093/nar/gki349>
- PERCHAT, S., DUBOIS, T., ZOUHIR, S., GOMINET, M., PONCET, S., LEMY, C., AUMONT-NICAISE, M., DEUTSCHER, J., GOHAR, M., NESSLER, S. & LERECLUS, D. (2011): A cell–cell

- communication system regulates protease production during sporulation in bacteria of the *Bacillus cereus* group. *Molecular Microbiology* **82**(3): 619–633. <https://doi.org/10.1111/j.1365-2958.2011.07839.x>
- PERCHAT, S., TALAGAS, A., PONCET, S., LAZAR, N., LI DE LA SIERRA-GALLAY, I., GOHAR, M., LERECLUS, D. & NESSLER, S. (2016): How Quorum Sensing Connects Sporulation to Necrotrophism in *Bacillus thuringiensis*. *PLoS Pathogens* **12**(8): e1005779. <https://doi.org/10.1371/journal.ppat.1005779>
- PERCHAT, S., TALAGAS, A., ZOUHIR, S., PONCET, S., BOUILLAUT, L., NESSLER, S. & LERECLUS, D. (2016): NprR, a moonlighting quorum sensor shifting from a phosphatase activity to a transcriptional activator. *Microbial Cell* **3**(11): 573–575. <https://doi.org/10.15698/mic2016.11.542>
- PEREGO, M. & BRANNIGAN, J. A. (2001): Pentapeptide regulation of aspartyl-phosphate phosphatases. *Peptides* **22**(10): 1541–1547. [https://doi.org/10.1016/S0196-9781\(01\)00490-9](https://doi.org/10.1016/S0196-9781(01)00490-9)
- PEREGO, M. & HOCH, J. A. (1996): Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **93**(4): 1549–1553. <https://doi.org/10.1073/pnas.93.4.1549>
- PEREZ-PASCUAL, D., MONNET, V. & GARDAN, R. (2016): Bacterial Cell–Cell Communication in the Host via RRNPP Peptide-Binding Regulators. *Frontiers in Microbiology* **7**: 706. <https://doi.org/10.3389/fmicb.2016.00706>
- PLUMMER, P. J. (2012): LuxS and quorum-sensing in *Campylobacter*. *Frontiers in Cellular and Infection Microbiology* **2**: 22. <https://doi.org/10.3389/fcimb.2012.00022>
- POTTATHIL, M. & LAZZAZERA, B. A. (2003): The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Frontiers in Bioscience: A Journal and Virtual Library* **8**: d32-45. <https://doi.org/10.2741/913>
- PUTNAM, C. D., HAMMEL, M., HURA, G. L. & TAINER, J. A. (2007): X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**(3): 191–285. <https://doi.org/10.1017/S0033583507004635>
- RADER, B. A., WREDEN, C., HICKS, K. G., SWEENEY, E. G., OTTEMANN, K. M. & GUILLEMIN, K. (2011): *Helicobacter pylori* perceives the quorum-sensing molecule AI-2 as a chemorepellent via the chemoreceptor TlpB. *Microbiology* **157**(Pt 9): 2445–2455. <https://doi.org/10.1099/mic.0.049353-0>
- RAFFA, R. B., IANNUZZO, J. R., LEVINE, D. R., SAEID, K. K., SCHWARTZ, R. C., SUCIC, N. T., TERLECKYJ, O. D. & YOUNG, J. M. (2005): Bacterial Communication (“Quorum Sensing”) via Ligands and Receptors: A Novel Pharmacologic Target for the Design of Antibiotic Drugs. *Journal of Pharmacology and Experimental Therapeutics* **312**(2): 417–423. <https://doi.org/10.1124/jpet.104.075150>

- RASKO, D. A. & SPERANDIO, V. (2010): Anti-virulence strategies to combat bacteria-mediated disease. *Nature Reviews Drug Discovery* **9**(2): 117–128. <https://doi.org/10.1038/nrd3013>
- REDFIELD, R. (2002): Is quorum sensing a side effect of diffusion sensing? *Trends in Microbiology* **10**(8): 365–370. [https://doi.org/10.1016/S0966-842X\(02\)02400-9](https://doi.org/10.1016/S0966-842X(02)02400-9)
- RELIGA, T. L., JOHNSON, C. M., VU, D. M., BREWER, S. H., DYER, R. B. & FERSHT, A. R. (2007): The helix–turn–helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proceedings of the National Academy of Sciences of the United States of America* **104**(22): 9272–9277. <https://doi.org/10.1073/pnas.0703434104>
- REZZONICO, F. & DUFFY, B. (2008): Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for luxS in most bacteria. *BMC Microbiology* **8**: 154. <https://doi.org/10.1186/1471-2180-8-154>
- ROCHA-ESTRADA, J., ACEVES-DIEZ, A. E., GUARNEROS, G. & TORRE, M. DE LA. (2010): The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Applied Microbiology and Biotechnology* **87**(3): 913–923. <https://doi.org/10.1007/s00253-010-2651-y>
- ROY, V., ADAMS, B. L. & BENTLEY, W. E. (2011): Developing next generation antimicrobials by intercepting AI-2 mediated quorum sensing. *Enzyme and Microbial Technology* **49**(2): 113–123. <https://doi.org/10.1016/j.enzmictec.2011.06.001>
- RUDENKO, O., THUREAU, A. & PEREZ, J. (2019): Evolutionary Refinement of the 3D Structure of Multi-domain Protein Complexes from Small Angle X-ray Scattering Data. Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '19. ACM, New York, NY, USA.
- RUIMY, R. & ANDREMONT, A. (2004): Quorum-sensing chez *Pseudomonas aeruginosa* : mécanisme moléculaire, impact clinique, et inhibition. *Réanimation* **13**(3): 176–184. <https://doi.org/10.1016/j.reaurg.2004.02.003>
- RUTHERFORD, S. T. & BASSLER, B. L. (2012): Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. *Cold Spring Harbor Perspectives in Medicine* **2**(11): a012427. <https://doi.org/10.1101/cshperspect.a012427>
- SALMOND, G. P. C., BYCROFT, B. W., STEWART, G. S. a. B. & WILLIAMS, P. (1995): The bacterial ‘enigma’: cracking the code of cell–cell communication. *Molecular Microbiology* **16**(4): 615–624. <https://doi.org/10.1111/j.1365-2958.1995.tb02424.x>
- SCHAUDER, S., SHOKAT, K., SURETTE, M. G. & BASSLER, B. L. (2001): The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Molecular Microbiology* **41**(2): 463–476. <https://doi.org/10.1046/j.1365-2958.2001.02532.x>
- SCHUSTER, M., SEXTON, D. J., DIGGLE, S. P. & GREENBERG, E. P. (2013): Acyl-Homoserine Lactone Quorum Sensing: From Evolution to Application. *Annual Review of Microbiology* **67**(1): 43–63. <https://doi.org/10.1146/annurev-micro-092412-155635>

- SCUTERA, S., ZUCCA, M. & SAVOIA, D. (2014): Novel approaches for the design and discovery of quorum-sensing inhibitors. *Expert Opinion on Drug Discovery* **9**(4): 353–366. <https://doi.org/10.1517/17460441.2014.894974>
- SHANKER, E., MORRISON, D. A., TALAGAS, A., NESSLER, S., FEDERLE, M. J. & PREHNA, G. (2016): Pheromone Recognition and Selectivity by ComR Proteins among Streptococcus Species. ZHANG, G. (ed). – *PLOS Pathogens* **12**(12): e1005979. <https://doi.org/10.1371/journal.ppat.1005979>
- SHI, K., BROWN, C. K., GU, Z.-Y., KOZLOWICZ, B. K., DUNNY, G. M., OHLENDORF, D. H. & EARHART, C. A. (2005): Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in Enterococcus faecalis. *Proceedings of the National Academy of Sciences of the United States of America* **102**(51): 18596–18601. <https://doi.org/10.1073/pnas.0506163102>
- SIFRI, C. D. (2008): Healthcare Epidemiology: Quorum Sensing: Bacteria Talk Sense. *Clinical Infectious Diseases* **47**(8): 1070–1076. <https://doi.org/10.1086/592072>
- SIKORSKI, R. S., BOGUSKI, M. S., GOEBL, M. & HIETER, P. (1990): A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* **60**(2): 307–317. [https://doi.org/10.1016/0092-8674\(90\)90745-Z](https://doi.org/10.1016/0092-8674(90)90745-Z)
- SLAMTI, L. & LERECLUS, D. (2002): A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the Bacillus cereus group. *The EMBO journal* **21**(17): 4550–4559. <https://doi.org/10.1093/emboj/cdf450>
- SLAMTI, L. & LERECLUS, D. (2005): Specificity and Polymorphism of the PlcR-PapR Quorum-Sensing System in the Bacillus cereus Group. *Journal of Bacteriology* **187**(3): 1182–1187. <https://doi.org/10.1128/JB.187.3.1182-1187.2005>
- SOLOMON, J. M., LAZAZZERA, B. A. & GROSSMAN, A. D. (1996): Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in Bacillus subtilis. *Genes & Development* **10**(16): 2014–2024. <https://doi.org/10.1101/gad.10.16.2014>
- STEINER, E., SCOTT, J., MINTON, N. P. & WINZER, K. (2012): An agr Quorum Sensing System That Regulates Granulose Formation and Sporulation in Clostridium acetobutylicum. *Applied and Environmental Microbiology* **78**(4): 1113–1122. <https://doi.org/10.1128/AEM.06376-11>
- SVERGUN, D. I. (1992): Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**(4): 495–503. <https://doi.org/10.1107/S0021889892001663>
- SVERGUN, D. I., PETOUKHOV, M. V. & KOCH, M. H. J. (2001): Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophysical Journal* **80**(6): 2946–2953. [https://doi.org/10.1016/S0006-3495\(01\)76260-1](https://doi.org/10.1016/S0006-3495(01)76260-1)
- SWORDS, W. E. (2012): Quorum signaling and sensing by nontypeable Haemophilus influenzae. *Frontiers in Cellular and Infection Microbiology* **2**: 100. <https://doi.org/10.3389/fcimb.2012.00100>

- TALAGAS, A., FONTAINE, L., LEDESMA-GARCA, L., MIGNOLET, J., SIERRA-GALLAY, I. L. de la, LAZAR, N., AUMONT-NICAISE, M., FEDERLE, M. J., PREHNA, G., HOLS, P. & NESSLER, S. (2016): Structural Insights into Streptococcal Competence Regulation by the Cell-to-Cell Communication System ComRS. *PLOS Pathogens* **12**(12): e1005980. <https://doi.org/10.1371/journal.ppat.1005980>
- TAN, L., LI, S. R., JIANG, B., HU, X. M. & LI, S. (2018): Therapeutic Targeting of the Staphylococcus aureus Accessory Gene Regulator (agr) System. *Frontiers in Microbiology* **9**: 55. <https://doi.org/10.3389/fmicb.2018.00055>
- TANG, K. & ZHANG, X.-H. (2014): Quorum Quenching Agents: Resources for Antivirulence Therapy. *Marine Drugs* **12**(6): 3245–3282. <https://doi.org/10.3390/md12063245>
- THOENDEL, M., KAVANAUGH, J. S., FLACK, C. E. & HORSWILL, A. R. (2011): Peptide Signaling in the Staphylococci. *Chemical Reviews* **111**(1): 117–151. <https://doi.org/10.1021/cr100370n>
- VADAKKAN, K., CHOUDHURY, A. A., GUNASEKARAN, R., HEMAPRIYA, J. & VIJAYANAND, S. (2018): Quorum sensing intervened bacterial signaling: Pursuit of its cognizance and repression. *Journal of Genetic Engineering & Biotechnology* **16**(2): 239–252. <https://doi.org/10.1016/j.jgeb.2018.07.001>
- VAGIN, A. & TEPLYAKOV, A. (2010): Molecular replacement with *MOLREP*. *Acta Crystallographica Section D Biological Crystallography* **66**(1): 22–25. <https://doi.org/10.1107/S09074444909042589>
- VERMA, S. C. & MIYASHIRO, T. (2013): Quorum Sensing in the Squid-Vibrio Symbiosis. *International Journal of Molecular Sciences* **14**(8): 16386–16401. <https://doi.org/10.3390/ijms140816386>
- WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. (2018): SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**(W1): W296–W303. <https://doi.org/10.1093/nar/gky427>
- WATERS, C. M. & BASSLER, B. L. (2005): QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology* **21**(1): 319–346. <https://doi.org/10.1146/annurev.cellbio.21.012704.131001>
- WENDERSKA, I. B., LUKENDA, N., CORDOVA, M., MAGARVEY, N., CVITKOVITCH, D. G. & SENADHEERA, D. B. (2012): A novel function for the competence inducing peptide, XIP, as a cell death effector of Streptococcus mutans. *FEMS microbiology letters* **336**(2): 104–112. <https://doi.org/10.1111/j.1574-6968.2012.02660.x>
- WHOLEY, W.-Y., KOCHAN, T. J., STORCK, D. N. & DAWID, S. (2016): Coordinated Bacteriocin Expression and Competence in Streptococcus pneumoniae Contributes to Genetic Adaptation through Neighbor Predation. *PLoS Pathogens* **12**(2): e1005413. <https://doi.org/10.1371/journal.ppat.1005413>
- WILLIAMS, P., WINZER, K., CHAN, W. C. & CÁMARA, M. (2007): Look who's talking: communication and quorum sensing in the bacterial world. *Philosophical Transactions*

- of the Royal Society B: Biological Sciences* **362**(1483): 1119–1134.
<https://doi.org/10.1098/rstb.2007.2039>
- WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISINEL, E. B., LESLIE, A. G. W., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. (2011): Overview of the CCP 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**(4): 235–242. <https://doi.org/10.1107/S0907444910045749>
- WINTJENS, R. & ROOMAN, M. (1996): Structural Classification of HTH DNA-binding Domains and Protein – DNA Interaction Modes. *Journal of Molecular Biology* **262**(2): 294–313.
<https://doi.org/10.1006/jmbi.1996.0514>
- WYNENDAELE, E., BRONSELAER, A., NIELANDT, J., D’HONDT, M., STALMANS, S., BRACKE, N., VERBEKE, F., VAN DE WIELE, C., DE TRÉ, G. & DE SPIEGELEER, B. (2013): Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. *Nucleic Acids Research* **41**(D1): D655–D659.
<https://doi.org/10.1093/nar/gks1137>
- YEHUDA, A., SLAMTI, L., BOCHNIK-TAMIR, R., MALACH, E., LERECLUS, D. & HAYOUKA, Z. (2018): Turning off *Bacillus cereus* quorum sensing system with peptidic analogs. *Chemical Communications (Cambridge, England)* **54**(70): 9777–9780.
<https://doi.org/10.1039/c8cc05496g>
- YEHUDA, A., SLAMTI, L., MALACH, E., LERECLUS, D. & HAYOUKA, Z. (2019): Elucidating the Hot Spot Residues of Quorum Sensing Peptidic Autoinducer PapR by Multiple Amino Acid Replacements. *Frontiers in Microbiology* **10**: 1246.
<https://doi.org/10.3389/fmicb.2019.01246>
- ZACCARIA, E., BAARLEN, P. VAN, GREEFF, A. DE, MORRISON, D. A., SMITH, H. & WELLS, J. M. (2014): Control of competence for DNA transformation in *Streptococcus suis* by genetically transferable phenotypes. *PloS One* **9**(6): e99394.
<https://doi.org/10.1371/journal.pone.0099394>
- ZETZMANN, M., SÁNCHEZ-KOPPER, A., WAIDMANN, M. S., BLOMBACH, B. & RIEDEL, C. U. (2016): Identification of the agr Peptide of *Listeria monocytogenes*. *Frontiers in Microbiology* **7**: 989. <https://doi.org/10.3389/fmicb.2016.00989>
- ZHU, J. & KAUFMANN, G. F. (2013): Quo vadis quorum quenching? *Current Opinion in Pharmacology* **13**(5): 688–698. <https://doi.org/10.1016/j.coph.2013.07.003>
- ZOUHIR, S., PERCHAT, S., NICAISE, M., PEREZ, J., GUIMARAES, B., LERECLUS, D. & NESSLER, S. (2013): Peptide-binding dependent conformational changes regulate the transcriptional activity of the quorum-sensor NprR. *Nucleic Acids Research* **41**(16): 7920–7933. <https://doi.org/10.1093/nar/gkt546>

Titre : Déterminants structuraux de la régulation de la compétence par le système de communication ComRS chez les streptocoques

Mots clés : analyse structure-fonction ; cristallographie des protéines ; mécanisme moléculaire ; communication bactérienne ; maladies infectieuses ; cibles thérapeutiques

Résumé : Les bactéries utilisent la communication intercellulaire pour se coordonner afin de réguler des processus bactériens majeurs comme la virulence, la sporulation, la compétence ou la formation de biofilms en fonction de la densité cellulaire. Ce processus repose sur la sécrétion de petites molécules signal appelées phéromones. Chez les bactéries à gram positif, ces phéromones sont de petits peptides qui peuvent être, soit reconnus au niveau de la membrane externe par des systèmes à deux composants dits systèmes indirects, soit être re-intériorisés afin de se fixer directement sur un régulateur transcriptionnel cytoplasmique. Dans la recherche de nouveaux agents antimicrobiens, le potentiel thérapeutique d'inhibiteurs ciblant les systèmes de communication intercellulaire est bien étudié chez les bactéries à gram négatif qui utilisent des homoserine lactones comme phéromones. Quelques études s'intéressent aux systèmes indirects de bactéries pathogènes à gram positif mais le potentiel des systèmes directs, plus récemment identifiés, n'a pas encore été exploré. Les récepteurs cytoplasmiques directement régulés par des peptides signal re-intériorisés forment la famille des RNPP. Ces récepteurs sont caractérisés par un domaine de fixation du peptide composé d'une répétition de motifs en hélice α 61537; appelés Tetratricopeptide Repeats (TPR). La plupart de ces récepteurs sont des régulateurs transcriptionnels contenant un domaine N-terminal de fixation de l'ADN de type Hélice-Tour-Hélice (HTH). Les études précédentes ont montré que, malgré une structure conservée, les modes de régulation des différents membres de la famille RNPP suivent des mécanismes moléculaires distincts. L'un de ces systèmes directs le mieux caractérisé est le système ComRS qui régule la compétence chez les streptocoques. La compétence permet aux bactéries d'internaliser des fragments d'ADN exogènes pour l'acquisition de nouveaux phénotypes tels que la résistance aux antibiotiques ou la virulence. Dans le groupe salivarius, il a été montré que les récepteurs ComR de deux espèces très proches, *S. thermophilus* et *S. vestibularis*, ne sont pas capables de coordonner leur état de compétence par échange de leurs peptides ComS respectifs. L'objectif de ma thèse a été d'étudier les déterminants structuraux de la spécificité du système ComRS. J'ai produit, purifié et cristallisé le récepteur ComR de *S. vestibularis* afin d'en déterminer la structure cristalline, seul et en présence de son peptide signal ComS. La comparaison de ces structures avec celles précédemment résolues chez *S. thermophilus*, conjointement à une étude fonctionnelle par mutagenèse dirigée réalisée chez nos collaborateurs (P. Hols, UCLouvain, Belgique), a permis d'aller plus loin dans la compréhension du mécanisme de régulation de ComR mais aussi d'identifier les résidus responsables de la spécificité du système ComRS. En parallèle, j'ai également initié la caractérisation structurale de deux paralogues de ComR chez *S. salivarius*, ScuR et SarF, qui ne sont pas activés par ComS et ne reconnaissent pas les mêmes cibles ADN que ComR malgré des séquences très conservées. Une analyse par SEC-MALS et SAXS m'a permis de montrer que ScuR semble suivre un mécanisme similaire à celui de ComR alors que SarF se comporte différemment en solution. J'ai proposé un modèle par homologie pour ScuR et cristallisé SarF. La résolution de sa structure est en cours. Cette étude permet donc de mieux comprendre la régulation de la compétence chez les streptocoques et ouvre la voie à d'éventuelles applications biotechnologiques ou biomédicales.

Title : Structural determinants of competence regulation by the communication system ComRS in streptococci

Keywords : structure-function analysis ; protein crystallography ; molecular mechanism ; bacterial communication ; infectious diseases ; therapeutic targets

Abstract : Bacteria use intercellular communication to coordinate and regulate major bacterial processes such as virulence, sporulation, competence or biofilm formation, as a function of cell density. This process relies on the secretion of small signal molecules called pheromones. In gram-positive bacteria, these pheromones are small peptides that can be either recognized at the outer membrane by two-component systems called indirect systems, or can be re-internalized to directly interact with a cytoplasmic transcriptional regulator. In the search for new antimicrobial agents, the therapeutic potential of inhibitors targeting intercellular communication systems is well studied in gram-negative bacteria that use homoserine lactones as pheromones. Some studies focus on the indirect systems of gram-positive pathogenic bacteria, but the potential of the more recently identified direct systems has not yet been explored. The cytoplasmic receptors directly regulated by re-internalized signal peptides form the RNPP family. These receptors are characterized by a peptide binding domain consisting of repeats of α 61537; helical motifs called Tetratricopeptide Repeats (TPR). Most of these receptors are transcriptional regulators containing an N-terminal DNA binding domain of the helix-turn-helix (HTH) type. Previous studies have shown that, despite a conserved structure, the modes of regulation of the different members of the RNPP family follow distinct molecular mechanisms. One of the best characterized direct systems is the ComRS system that regulates competence in streptococci. Competence allows bacteria to internalize exogenous DNA fragments for the acquisition of new phenotypes such as antibiotic resistance or virulence. In the salivarius group, it has been shown that the ComR receptors of two closely related species, *S. thermophilus* and *S. vestibularis*, are not able to coordinate their state of competence by exchange of their respective ComS peptides. The aim of my thesis was to study the structural determinants of the specificity of the ComRS system. I produced, purified and crystallized the ComR receptor from *S. vestibularis* in order to determine its crystal structure, alone and in the presence of its ComS signal peptide. The comparison of these structures with those previously solved with ComR from *S. thermophilus*, together with a functional study by directed mutagenesis performed by our collaborators (P. Hols, UCLouvain, Belgium), allowed us to go further in the understanding of the regulatory mechanism of ComR but also to identify residues responsible for the specificity of the ComRS system. In parallel, I also initiated the structural characterization of two ComR paralogs from *S. salivarius*, ScuR and SarF, which are not activated by ComS and do not recognize the same DNA targets as ComR despite highly conserved sequences. SEC-MALS and SAXS analyses allowed me to show that ScuR seems to follow a mechanism similar to that of ComR whereas SarF behaves differently in solution. I proposed a homology model for ScuR and crystallized SarF. The resolution of its structure is in progress. This study therefore provides a better understanding of the regulation of competence in streptococci and opens the way to potential biotechnological or biomedical applications.

