UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Identification of structural determinants involved in ComR activation and XIP specificity in salivarius streptococci

Author: Joaquin Guzman Espinola
Supervisors: Prs. Hols P. and Soumillion P.
Daily supervision: Dr. Ledesma García L.
Readers: Prs. Chaumont F. and Rezsohazy R.
Academic years: 2016-2017 and 2022-2023
Master in Biochemistry and Molecular and Cell Biology

Remerciements

Je tiens tout d'abord à remercier le Professeur Pascal Hols et la Docteure Laura Ledesma García qui m'ont accueilli dans le laboratoire du Louvain Institute of Biomolecular Science and Technology (LIBST) pour y réaliser mon projet de thèse de master. Chacun d'eux aura su faire preuve de pédagogie, de patience et de compréhension infinies à mon égard. Je les remercie chaleureusement pour leurs enseignements qui m'auront fait grandir tant d'un point de vue scientifique que d'un point de vue personnel. A ce titre, je tiens également à saluer tous les professeurs de l'UCL qui ont participé à ma formation, et plus particulièrement, les membres du jury de l'Ecole de Biologie qui m'ont témoigné leur confiance à un moment où je doutais de moi.

Je souhaite également saluer mes camarades de la promotion 2015-2017, l'équipe iGEM 2016 et toutes les belles âmes que j'ai eu la chance de croiser pendant mes années d'études. Parmi tous, je souhaite tout particulièrement remercier Antoine Godeaux et Olivier Suys pour leur amitié.

Ce travail n'aurait pas été possible sans le soutien inconditionnel de ma famille et de mes proches. Je les en remercie du fond du cœur.

Enfin, je dédie mes dernières pensées à toutes les personnes qui ont su me soutenir dans des moments plus éprouvants, avec la même facilité qu'ils ont partagé des moments de joie à mes côtés.

Abstract

Natural competence is the genetically specified ability of some bacteria to take up extracellular DNA in order to provide an evolutionary advantage or as a means of genome plasticity/repair. It is a social behavior extensively studied in the *Streptococcus* genus, in which this transient and energy-intensive physiological process is linked, *inter alia*, to biofilm formation and bacteriocin production. In most species of this genus, competence development is tightly regulated by the pheromone quorum-sensing system ComRS, which activates the expression of the central regulator of competence, the sigma factor X (also known as σ^X/ComX/SigX), responsible for the expression of late competence genes, including the transformasome and the recombination machinery.

ComS is the precursor of the competence pheromone peptide. After secretion and maturation, the matured ComS form (named XIP) is reinternalized and directly binds the transcriptional regulator ComR inducing its activation and enabling it in turn to stimulate the expression of *comS* and *comX* genes, marking the onset of competence development. The transcriptional regulator ComR belongs to the RRNPP superfamily of cytoplasmic sensors. It is characterized by an N-terminal helix-turn-helix (HTH) DNA-binding domain and a C-terminal alpha-helical domain containing 5 tetratricopeptide repeats (TPR) which interacts with the XIP pheromone required for its activation. Depending on the species, ComR proteins display strict or more permissive specificity according to their ability to be activated by heterologous peptides. Thus, some ComRs are able to recognize peptides of various sequences, while others can only be activated by their cognate peptides.

The aim of this master's thesis project is to identify structural determinants involved in ComR activation and XIP specificity in two closely-related species of the salivarius group of streptococci, i.e. *Streptococcus thermophilus* (Sth), a model organism for the study of the ComRS system, and a species of medical interest, *Streptococcus vestibularis* (Sve), a commensal species that has been shown to adopt a pathogenic behavior opportunistically. In this project, we report that $ComR_{Sve}$ is fully functional in *S. thermophilus*, but that the two highly similar ComRS systems do not cross-talk. We also report concurrently that *S. vestibularis* F0396 is competent under laboratory conditions and that its competence system is functional when XIP_{Sve} is provided in the culture medium. Finally, we generated $ComR_{Sth}$ mutants (ComR*) with altered XIP recognition capacity, including some capable of moderate cross-activation with XIP_{Sve}. These results highlighted the importance of different regions of the ComR TPR domain involved in XIP specificity and its binding-induced activation mechanism, notably loop L6 and alpha-helices α7 and α12 of TPR-1 and TPR-4, respectively. Those results pave the way towards a detailed molecular understanding of ComR-XIP interactions responsible for ComR activation and XIP selectivity.

Résumé

La compétence naturelle est la capacité déterminée génétiquement de certaines bactéries d'assimiler de l'ADN extracellulaire afin de bénéficier d'un avantage évolutif ou comme moyen de contribuer à la plasticité du génome, ou à sa réparation. Il s'agit d'un comportement social abondamment étudié dans le genre *Streptococcus*, dans lequel ce processus physiologique transitoire et énergivore est notamment associé à la formation de biofilms et à la production de bactériocines. Chez la plupart des espèces de ce genre, le développement de la compétence est étroitement régulé par le système de quorum sensing par phéromone ComRS, qui active l'expression du régulateur central de la compétence, le facteur sigma X (également connu sous le nom de σ^X/ComX/SigX), responsable de l'expression des gènes tardifs de la compétence, comprenant le transformasome et la machinerie de recombinaison génétique.

ComS est le précurseur de la phéromone peptidique de la compétence. Après sécrétion et maturation, la forme mature de ComS (appelée XIP) est réinternalisée et se lie directement au régulateur transcriptionnel ComR, induisant son activation et lui permettant de stimuler à son tour l'expression des gènes *comS* et *comX*, marquant ainsi le début du développement de la compétence. Le régulateur transcriptionnel ComR appartient à la superfamille RRNPP de senseurs cytoplasmiques. Il est caractérisé par un domaine N-terminal de liaison à l'ADN de type helix-turn-helix (HTH) et un domaine C-terminal alpha-hélicoïdal contenant 5 répétitions tetratricopeptidiques (TPR) qui interagit avec la phéromone XIP nécessaire à son activation. Selon les espèces, les protéines ComR présentent une spécificité stricte ou plus permissive en fonction de leur capacité à être activées par des peptides hétérologues. Ainsi, certains ComRs sont capables de reconnaître des peptides de séquences variées, tandis que d'autres ne peuvent être activés que par leurs propres peptides.

Le but de ce projet de thèse de master est d'identifier les déterminants structuraux impliqués dans l'activation de ComR et sa spécificité pour XIP chez deux espèces de streptocoques phylogénétiquement proches appartenant au groupe salivarius, à savoir *Streptococcus thermophilus* (Sth), un organisme modèle pour l'étude du système ComRS, et une espèce d'intérêt médical, *Streptococcus vestibularis* (Sve), une espèce commensale dont il a été démontré qu'elle peut adopter un comportement pathogène de manière opportuniste. Dans ce projet, nous démontrons que $ComR_{Sve}$ est entièrement fonctionnel dans *S. thermophilus*, mais que les deux systèmes ComRS très similaires ne communiquent pas entre eux. Nous rapportons également que *S. vestibularis* F0396 est compétent dans des conditions de laboratoire et que son système de compétence est fonctionnel lorsque XIP_{Sve} est disponible dans le milieu de culture. Enfin, nous présentons la création de mutants $ComR_{Sth}$ (ComR*) avec une capacité de reconnaissance pour XIP altérée, dont notamment certains capables d'une activation croisée modérée avec XIP_{Sve}. Ces résultats ont mis en évidence l'importance de différentes régions du domaine TPR de ComR impliquées dans la spécificité pour XIP et dans son mécanisme d'activation phéromone-dépendant, en particulier la boucle L6 et les hélices alpha α7 et α12 des motifs TPR-1 et TPR-4 respectivement. Ces résultats ouvrent la voie à une compréhension fine au niveau moléculaire des interactions ComR-XIP responsables de l'activation de ComR et de la sélectivité pour XIP.

Abbreviations

aa	Amino acids
ABC	ATP binding cassette
AHL	Acylhomoserine lactone
AI	Auto-inducer
AIP	Autoinducing peptide
Amp	Ampicillin
bp	Base pair
cat	Chloramphenicol acetyltransferase gene
CDM	Chemically defined medium
CFU	Colony-forming unit
Cm	Chloramphenicol
ComR*	ComR mutant
C-ter	Carboxyl-terminus
DNA	Deoxyribonucleic acid
eDNA	Extracellular deoxyribonucleic acid
ssDNA	Single-stranded deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
tDNA	Transforming deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DU-pilus	DNA-uptake pilus
DUS	DNA uptake sequence
Ecom-box	Early com genes box
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
FDA	U.S. Food and drug administration
glu	Glucose
GRAS	Generally regarded as safe
HGT	Horizontal gene transfer
HRP	Horseradish peroxidase
HTH	Helix-turn-helix
LB	Lysogeny broth
N-ter	Amine-terminus
OD_{600}	Optical density at 600 nm wavelength
O/N	Overnight
PCR	Polymerase chain reaction
OE-PCR	Overlap extension PCR
PG	Peptidoglycan
QS	Quorum sensing
RLU	Relative light unit
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ssa	*Streptococcus salivarius*
Sth	*Streptococcus thermophilus*
Strp	Streptomycin
Sve	*Streptococcus vestibularis*
TEMED	Tetramethylethylenediamine
TPR	Tetratricopeptide repeat
US$	United States dollar
USS	Uptake signal sequence
UV	Ultraviolet
wt	Wild-type
W/WO	With and without
XIP	σ^X/ComX/SigX-inducing peptide

Table of Contents

Introduction

1 Plasticity of the bacterial genome

Populating Earth for about 4 billion years, bacteria still represent the second kingdom of life in terms of biomass abundance after plants (1). Although single-celled organisms harbor a limited genome size, their evolutionary success is largely due to their ability to adapt quickly to various environments and lifestyles. This property is based on short-lived cell cycles allowing rapid modification of their genome, but also on their ability to acquire sequences from closely related organisms through horizontal gene transfer (HGT) (2). Unlike vertical gene transfer where the main evolutionary mechanism is based on mutations of existing genetic material, horizontal gene transfer allows potentially rapid access to novel sequences such as entire genetic systems coding for new metabolic pathways, antibiotic resistance genes or virulence factors. In this regard, bacterial genome plasticity both represents the set of external genetic resources accessible to a particular bacterial population to evolve, and its ability to integrate them in its genome (3). HGTs are detectable by multiple ways (2). These events produce a scattered phylogenetic distribution of the acquired trait making it often absent from closely related taxa. Furthermore, the overall G+C content of the acquired sequence will differ from the species' average and several clues such as a plasmid transfer origin can also be visible. Although HGTs can trouble phylogenetic relationships for some acquired sequences, lateral transfers do not impact the classification of organisms at the genomic scale (4). The three classical modes of HGTs are conjugation, transduction and natural transformation (see Fig. 1), but other mechanisms exist such as gene transfer agents (5), DNA-containing membrane vesicles (6) or intercellular nanotubes (7).

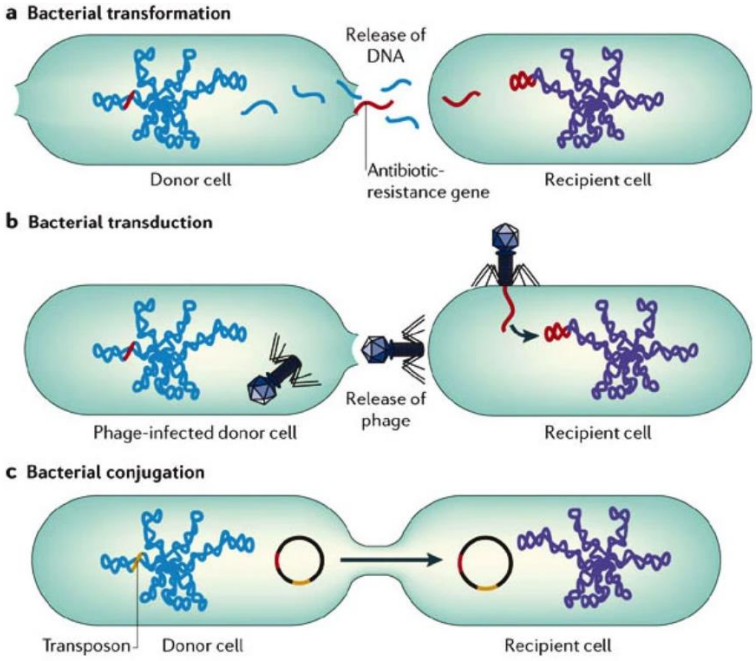


Figure 1 Classical modes of horizontal gene transfer in bacteria. Example of an antibiotic-resistance gene transfer. (A) Transformation is the uptake of extracellular DNA followed by its integration into the genome. (B) Transduction is a DNA transfer between two bacteria mediated by a bacteriophage. (C) Conjugation is a transfer of DNA between two bacteria mediated by cell-to-cell junctions and a pore. This mechanism is very common in plasmid exchange. Adapted from Furuya et al. (2006) (59).

1.1 Horizontal gene transfer barriers

In a study of 88 prokaryotic genomes, the percentage of laterally transferred genes was estimated to vary from 0 to 22% in bacteria (8) strengthening the claim that HGT is a major driving force of evolution in bacterial species. However, depending on the species and mechanisms considered, several constraints will limit the uptake, protection and stabilization steps of horizontal DNA transfers (Fig. 2), so that certain mechanisms and sequences will not be accessible to all bacterial species (9).

Theoretically, microorganisms must be part of the same exchange community to exchange genes (10). These communities are defined by several internal and environmental factors. The most significant are genome size, genome G+C composition, carbon utilization and oxygen tolerance, followed by temperature (10). In addition to these factors, there are specific barriers related to certain modes of transfer or established by particular species (9). Conjugation is for example limited by the phenomenon of surface exclusion which prevents the conjugative transfer of plasmids into a bacterium that already contains genes coding for a related transfer apparatus (9). Moreover, each plasmid is characterized by its host range which corresponds to its compatibility with the host replication system. During natural transformation, the stabilization step of a sequence in the genome relies on the presence of a homologous region in the genome. On the other hand, any transferred DNA can be targeted by endonucleases of restriction-modification systems of a cell if its sequence does not have the same chemical signatures such as methylations (11). Finally, from a practical point of view, spatiotemporal barriers also exist. Thus, a theoretically possible HGT between two species may not take place if they do not share the same habitat (10).

HGTs mechanisms are metabolically costly processes and potentially harmful if the transferred sequence has a deleterious effect on the fitness of the bacterium (9). For this reason, their expression is strictly regulated in such a way that it only occurs in optimal conditions according to the lifestyle of the bacterium.

In this master thesis, we will focus on natural transformation and the physiological process that enables it, natural competence.

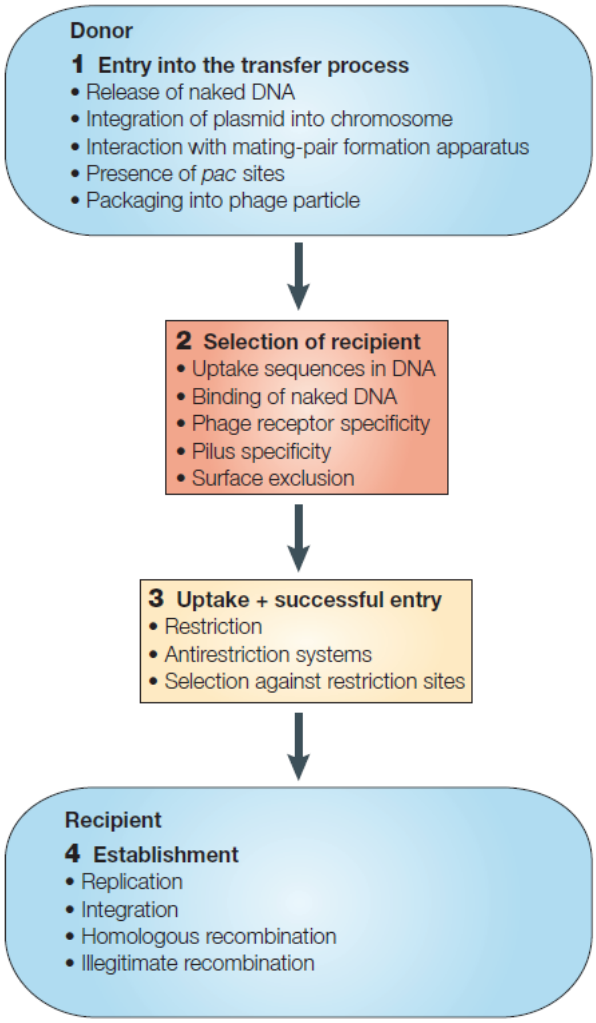


Figure 2 The process of horizontal gene transfer. A summary of the various steps that DNA undergoes during horizontal gene transfer, from the source to the recipient cell. Adapted from Thomas et al. (2005) (9).

2 Natural competence for transformation

Natural transformation is a type of DNA horizontal transfer between a recipient cell and extracellular DNA. It is a parasexual process, meaning it is a form of reproduction that allows genetic recombination without meiosis and gamete fusion (12). Where we refer to the gene transfer by "transformation", the physiological process that supports it is called "natural competence". In other words, when a bacterium reaches the competence state, it is then able to perform transformation. The term "natural" is opposed here to similar artificially induced properties as when referring to electrocompetent or heat shock competent cells for example.

2.1 Natural competence discovery

Historically, the first natural transformation event was reported in 1928 by Griffith (13). His experiment (Fig. 3) consisted in using two different strains of *Streptococcus pneumoniae* to inoculate mice. Both strains were pathogenic but differed morphologically by their smooth appearance for one and rough for the other. Rodents inoculated with the smooth strain died while those inoculated with the rough strain showed no symptoms. However, mice inoculated with a co-culture of rough cells with heat-killed smooth cells

died too. Thus, there was an exchange of properties between the two strains, which he named the "transformation principle". The rough strain became smooth and virulent as a result. Today, current knowledge allows us to understand that a HGT occurred during co-culture. *S. pneumoniae* is known to be a competent species and the smooth appearance of one of the strains was due to the presence of a capsule surrounding the bacteria that allowed them to resist against the mouse's immune system and become virulent. In addition to illustrating that natural transformation is a phenomenon that leads to the spread of virulence factors among pathogenic strains, this observation was also pioneering in the understanding that DNA was the vehicle used for the storage and transfer of biological information (14).

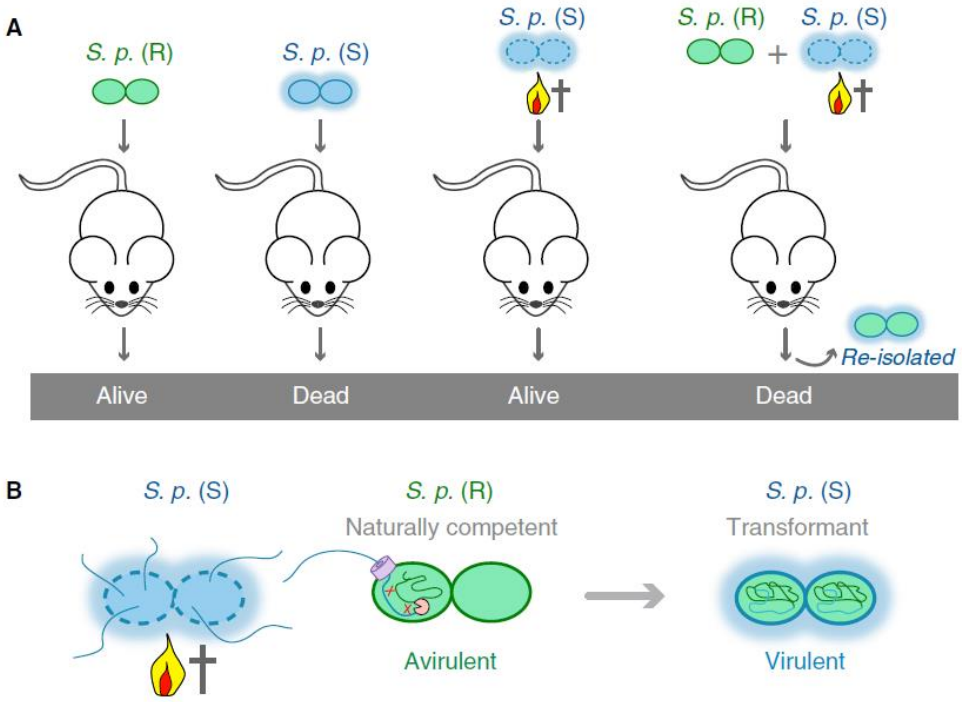


Figure 3 The discovery of natural competence for transformation via "the transformation principle" by Frederick Griffith. (A) Inoculation of avirulent (rough; R) or virulent (smooth; S) *S. pneumoniae* resulted in the survival or death of mice, respectively. Heat-inactivated (cross and flame) virulent strains were unable to kill the rodent, unless they were co-injected with the live avirulent strain. (B) Modern explanation of Griffith's "transformation principle". The live avirulent strain of *S. pneumoniae* acquired virulence traits from the heat-killed virulent isolate through its ability to uptake extracellular DNA i.e. natural competence. Adapted from Blokesch M. (2016) (15).

2.2 The natural transformation process

Transformation differs from transduction and conjugation in being a process mediated solely by the recipient cell. All the genes necessary for its functioning are encoded entirely in its genome and both their replication and expression are independent of extrachromosomal mobile genetic elements. It is thus particularly well adapted to the environment and to the physiological needs of the species in question (8).

The natural transformation process can be divided into three major steps: DNA uptake and processing, protection of internalized DNA, and finally, stabilization through integration into the genome or plasmid recircularization (see Fig. 4).

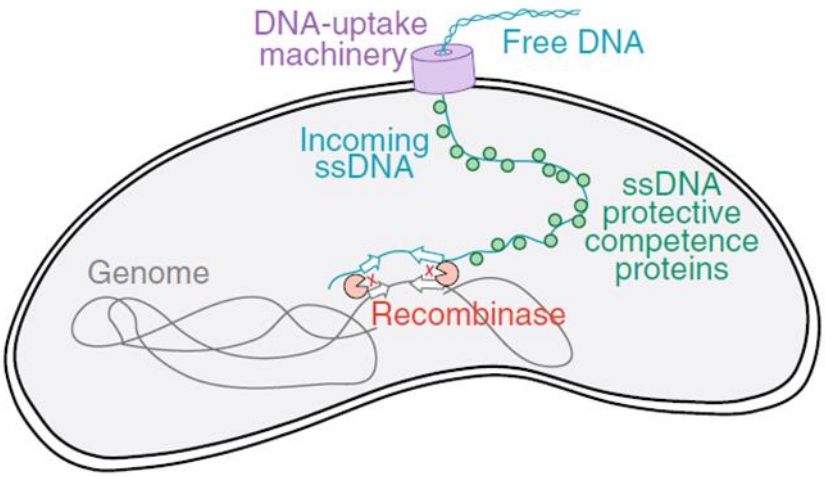


Figure 4 Overview of the transformation process. Free extracellular DNA is transferred to the cytosol by the DNA uptake apparatus and processed into single-stranded DNA (ssDNA). The ssDNA is then protected from degradation by specialized competence proteins and will be stabilized by integration into the bacterial chromosome through the action of a recombinase. Adapted from Blokesch M. (2016) (15).

2.3 Prevalence of natural competence in bacteria

Natural competence for transformation is widespread in Gram-negative and Gram-positive bacteria, and at least in some Archaea (8,12). However, despite phylogenetic predictions based on the presence of homologous genes essential for transformation, effective transformation was only documented in about 80 bacterial species in 2014 (12). It is likely that many more species are naturally competent, but the specific conditions and inducers have either not yet been elucidated or are not reproducible in the laboratory, to date. For instance, *Escherichia coli*, one of the most studied bacteria, contains all the required genes for transformation, but its competence has never been demonstrated. The general transient nature of competence activation and the lack of identification of a central competence regulator represent often obstacles to its observation. It is also possible that several homologous genes have drifted genetically and that this assertion is overstated. Transformation is indeed thought to be an ancient trait inherited from a common ancestor (12).

For more information on the phylogenetic distribution of naturally transformable bacterial species, we refer to the article by Johnsborg et al. (2007) (8) and to the review by Johnston et al. (2014) (12).

2.4 Natural competence regulation

While the proteins of the DNA uptake apparatus are conserved in both Gram-negative and Gram-positive organisms, regulatory systems vary greatly, having co-evolved to match the lifestyle of each bacterial species (12).

2.4.1 Quorum-sensing-dependent regulation of competence

The development of the competence state is a complex and energetically expensive process that is adopted collectively at opportune times, by a bacterial population to reduce its cost and increase its efficiency. As many processes with these characteristics, its regulation is coupled to a quorum sensing (QS) system (16). A QS system is a cell-cell communication process based on the production and detection of secreted signal molecules, collectively known as auto-inducers (AIs), that accumulate in the extracellular environment. Their concentration reflects the number of local producing cells. Gram-

positive and Gram-negative bacteria use different types of QS systems and signaling molecules. Gram-positive bacteria use peptides, called autoinducing peptides (AIPs), or peptides pheromones, while Gram-negative bacteria use small molecules having their own specific synthesis pathway instead, such as acylhomoserine lactones (AHL) for instance (Fig. 5).

QS systems share the same overall mechanism. At low cell density, the secreted auto-inducer (AI) molecules diffuse away and their local concentration is too low to be detected. The extracellular concentration of AIs increases linearly as the bacterial population grows. At a given cell density, this concentration reaches a species-specific threshold allowing it to be detected directly by membrane receptors or, after internalization, by cytoplasmic receptors. This will trigger an increase in AI production, thus creating an auto-amplification loop allowing an exponential AI accumulation supporting the expression of responding genes to which the QS is linked, but also the propagation of the response to other neighboring cells. Bacteria can thereby synchronize their behavior and link it to a specific cell density. Other examples of QS-regulated processes are bioluminescence, sporulation, biofilm formation, or the production of virulence factors by pathogenic bacteria (16).

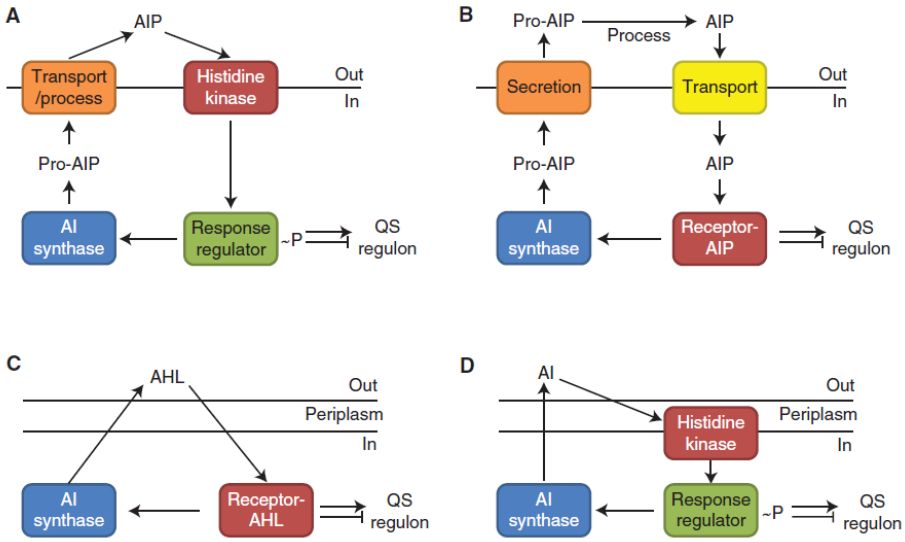


Figure 5 Overview of the different types quorum sensing (QS) circuits in bacteria. In Gram-positive bacteria, autoinducing peptides (AIP) are secreted/transported and are either sensed externally by a two-component signaling system involving a histidine kinase (A), or internally, after reimportation, by an AIP-binding transcription factor (B). In Gram-negative bacteria, small QS molecules, acylhomoserine lactones (AHL) or other autoinducers (AI), freely diffuse in and out and are sensed internally (C) or externally (D) (17). Adapted from Rutherford et al. (2012) (16).

Besides being a form of intra-species cell-cell communication, QS systems may allow bacteria of one species to possibly detect signal molecules of other species (18). This crosstalk may allow them to assess the presence of other bacteria in their environment and to react accordingly. Therefore, we can consider that QS systems represent a means of recognition of self and non-self in bacteria, which can even lead to the discrimination of slightly different genotypes within the same species. Strains that produce the same type of QS peptides are called “pherotypes” and some pherotypes cannot communicate with each other. In sociomicrobiology, quorum-sensing is considered to be a social trait, i.e. a trait that affects both the individual performing a behavior and a recipient (19). It is important to note, however, that not all signals have the same intentionality. There is indeed a difference between signalling systems and coercion (20). It is therefore essential

to consider the whole context in which QS systems take place, and the same applies for the development of natural competence.

2.4.2 Natural competence inducers

Usually, the competence state is limited in time, linked to a specific cell density by a QS system, and its expression occurs only when specific cellular or environmental conditions are met. There is no universal inducer of competence, but some known inducers include high cell density, antibiotic stress, DNA damage, absence of preferred carbon sources, chitinous surfaces, and nutrient limitation (15). Furthermore, some species are constitutively competent in laboratory conditions, such as *Neisseria gonorrhoeae* or *Helicobacter pylori*. Hypothetically, *H. pylori* would periodically alternate between phases of competence and non-competence (21).

2.4.3 Central regulators of natural competence

Central regulators of competence are defined as proteins that directly control the expression of "*com* regulons" (12), i.e. sets of genes that are expressed in a coordinated manner at the initiation of competence. These regulators include alternative sigma factors, transcription activators or co-regulators that act as a gateway between the different biotic and abiotic signals inducing the activation of competence and the expression of genes involved in the broader physiological response. The end of competence depends on a shut-off system, in charge, amongst other things, of ending the positive feedback loop in the QS system and thus signal propagation to neighboring cells. Depending on the species, the entire bacterial population will become competent, or only a part. In *Bacillus subtilis* for example, only a subpopulation of 10-20% of cells become competent. This phenomenon is known as bimodal expression (22). All these variations are adaptation marks of the mechanisms of competence to the different lifestyles of bacteria.

Finally, it should be noted that means to interfere in the regulatory networks of competence are a key subject in medical research. Not only is transformation an important vector for the propagation of pathogenic traits, but presumably, antimicrobial molecules that affect bacterial behavior would exert less selection pressure on their targets. Such treatments represent a promising alternative to antibiotics and are believed to generate resistances at a slower pace (23).

2.4.4 Roles and significance of natural competence

The role of DNA uptake conferred by competence is still a matter of debate in the scientific literature. In many species, the activation of natural competence represents only a part of the physiological response collectively engaged. For example, in *B. subtilis*, the genes required for transformation represent only a quarter of approximately a hundred genes under the control of the central regulator ComK (24). Similarly, one should not draw hasty conclusions between a condition identified as an inducer of competence and the overall meaning of the triggering of the competence state. For instance, a nutritional role was suggested for competence in *Haemophilus influenzae* after a correlation was found between its onset and the depletion of purine pools, but these changes can equally be interpreted as replication stress (12,25).

The most obvious advantage of DNA uptake is the genetic plasticity that transformation confers by providing new genetic material and hence potentially new functions or traits. In certain species, natural competence is activated when DNA damage occurs or during

exposition to certain antibiotics. It would then have a role in DNA repair. This deduction was made in particular in *S. pneumoniae* where competence development increases its resistance to UV radiation and other DNA-damaging agents (26). *S. pneumoniae* lacks a SOS response to mitigate DNA damage. Instead, overexpression of RecA recombinase during competence in this species combined with transferred extracellular DNA would substitute this lacking (27). Moreover, since homologous recombination promotes integration of sequences with high sequence identity, competence is seen as a maintenance mechanism to purge the genome of deleterious mutations or selfish genetic elements for example (15). Another role often mentioned for competence is a nutritional role (12,28). The internalized DNA segment would either be used directly as a source of carbon or degraded and the resulting nucleotides would be recycled in the synthesis of new nucleic acids. This hypothesis is however highly controversial. The assembly of the transformation machinery requires a high energetic cost whereas specific transporters for the import of nucleotides already exist. Likewise, the transient and highly regulated nature of competence seems counterintuitive with a nutritional function responding to constant needs (12).

On the counterparts, transformation can also lead to the insertion of genetic material that is deleterious to the fitness of the bacteria. Some mathematical models theorize that the extracellular DNA pool from dead bacteria contain a bias of sequences that are poorly adapted to the environment (28). This "bad gene effect" is however limited by various species-specific selection mechanisms. The high energy cost of the transformation machinery assembly is another limitation. The competence state is additionally often associated with growth slowdown or arrest (7). This step is seen as crucial for transformation, which proves the importance of a tight regulation system.

2.5 Extracellular DNA capture and integration

DNA uptake is the first step of transformation. This term defines the entry of extracellular DNA (eDNA), also called transforming DNA (tDNA), into the periplasm in Gram-negative bacteria, or into the cytoplasm in Gram-positive bacteria (7,12). After internalization, the tDNA will be protected from cell nucleases and prepared for integration into the chromosome.

2.5.1 Origin and availability of eDNA in nature

Extracellular DNA (eDNA) is widespread in natural environments, from saline or freshwater aquatic environments to sediments and soils. It is present in substantial quantities ranging from a few $ng.L^{-1}$ to several $\mu g.L^{-1}$ in marine environments and up to several dozen of $\mu g.g^{-1}$ in soils. Its shelf life can vary from a few hours to hundreds of years (29).

Accumulation of eDNA in the environment originates from several sources. It may derive from the active release of living cells, the passive release of dead cells, or from viruses (29). The quality of the eDNA available for transformation varies greatly depending on its source and its state of degradation. As integration is based on homologous recombination, it is expected that the integration of a sequence coming from a highly degraded molecule and/or from a genetically distant species will be more difficult and less useful for processes based on its coding capacity, in contrast to fresh genetic material coming from a phylogenetically closer cell (7).

In many species, there are active mechanisms of DNA release. This is the case in *N. gonorrhoeae*, where eDNA is released without cell lysis, (30) by a type IV secretion system

(31). In *S. pneumoniae* and *Vibrio cholerae*, respectively, lytic mechanisms of fratricide (a subpopulation within the same species) and non-kin predation (inter-species) are tightly linked to the competent state (32,33). Such behaviors provide genetic material for transformation.

2.5.2 tDNA uptake and specificity

As aforementioned, the proteins required for the assembly of the DNA uptake apparatus, sometimes referred to as the "transformasome" (34), are remarkably well conserved in both Gram-positive and Gram-negative bacteria (12) (Fig. 6). With the exception of *H. pylori* (35), both types of bacteria possess filamentous structures, referred here as DNA-uptake pili (DU-pilus) are related to type IV pili and the type II secretion systems (7). These DU-pili transport the double-stranded tDNA to the cell surface. This step involves crossing the peptidoglycan cell wall in Gram-positive cells or the outer membrane in Gram-negative cells. The DNA interactions of this DU-pilus and the mechanism by which it would retract or fold to accomplish this role are poorly understood at this time. The DU-pilus then introduces the tDNA to a DNA binding protein, ComEA, which mediates the uptake through the membrane channel protein ComEC (7). Simultaneously to its passage to the cytosol, one of the strands of the tDNA is degraded. The nuclease responsible for this activity is generally not identified, except in *S. pneumoniae* where it is attributed to EndA (36). The single-stranded tDNA then enters the cytosol in a 3'-5' orientation (7,37).

DNA uptake by competent bacteria is documented to occur at a rate of up to 100 bp per second in *S. pneumoniae* (9,36). Both circular and linear double-stranded DNA molecules can be transformed. However, it is commonly accepted that the uptake of plasmids is less efficient (9). DNA uptake is not sequence-specific in most species. Yet, in some Gram-negative species such as *H. influenzae* (38) or in the genus *Neisseria* (39), the presence of certain sequence motifs called DUS (DNA uptake sequences) or USS (Uptake signal sequences) are essential for the uptake to occur (40). This example illustrates further the sociological links between speciation and natural competence.

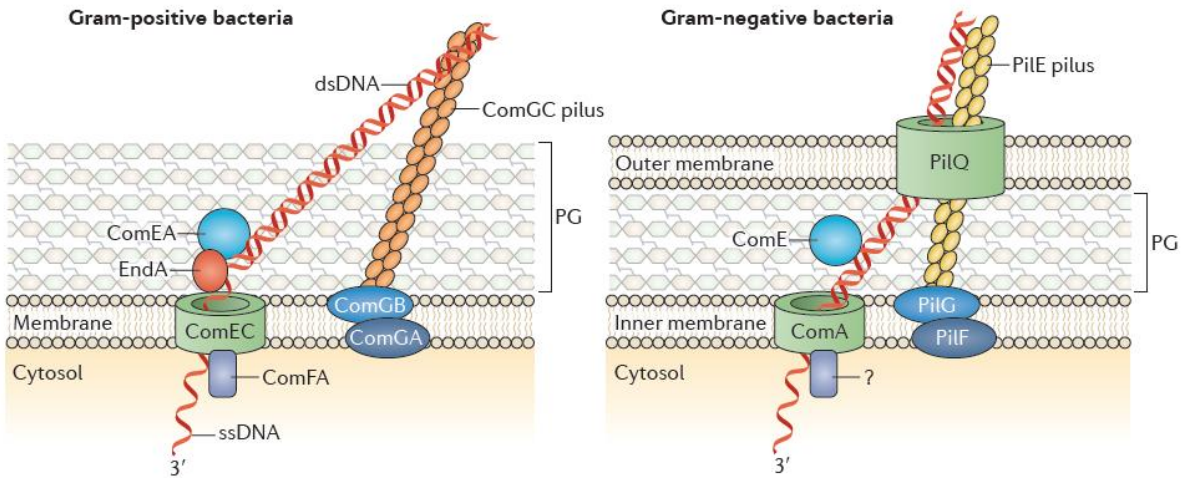


Figure 6 The DNA uptake step during natural transformation. In Gram-positive bacteria, the DU-pilus is composed mainly of ComGC subunits.Presumably, ComGA, ComGB and ComFA are involved in ATP-dependent assembly and retraction of the pilus in Firmicutes. On the other hand, ComEA is thought to be anchored to the membrane(7). In Gram-negative bacteria, the DU-pilus is mainly composed of PilE subunits and crosses the outer membrane through the PilQ secretin channel. Other proteins are also required for the elongation and retraction mechanism of the pilus,such as PilG and PilF. ComE is free in the periplasm. Adapted from Johnston et al. (2014) (12).

2.5.3 tDNA protection and stabilization

After entering the cytosol, the tDNA will be protected from intracellular nucleases by interacting with a series of cytoplasmic proteins. To be stabilized, the single-stranded tDNA must recombine with the chromosome or, in the case of a translocated plasmid without sequence homology, it can recirculate into a self-replicating plasmid (9).

Integration of the tDNA into the genome requires sequence homology between the tDNA and the bacterial chromosome. This step is carried out by the RecA recombinase which performs homology search and strand exchange to form a DNA heteroduplex between the genomic DNA and the tDNA (Fig. 7) (12). This heteroduplex can be a fully homologous double stranded recombination intermediate or can have a non-homologous region that does not match the genomic sequence in front of it and will form a loop. In this case, it will lead to additive integration which does not respect the size and functionality of the replaced region. In order to work, homologous recombination requires, in most cases, the presence of regions of 25 to 200 bp of high sequence similarity and up to 25% divergence (9). Besides, integration can also lead to illegitimate recombination event that, although not frequent, will extend the range of potential genetic exchanges (9). A final mechanism limiting the acquisition of DNA from other species is the action of its restriction-modification systems which can specifically degrade foreign DNA depending on its methylation state (12). As with all evolutionary processes, transformation-induced mutations will be conserved if they are beneficial or neutral for the fitness of the bacterium, while they will be eliminated if deleterious.

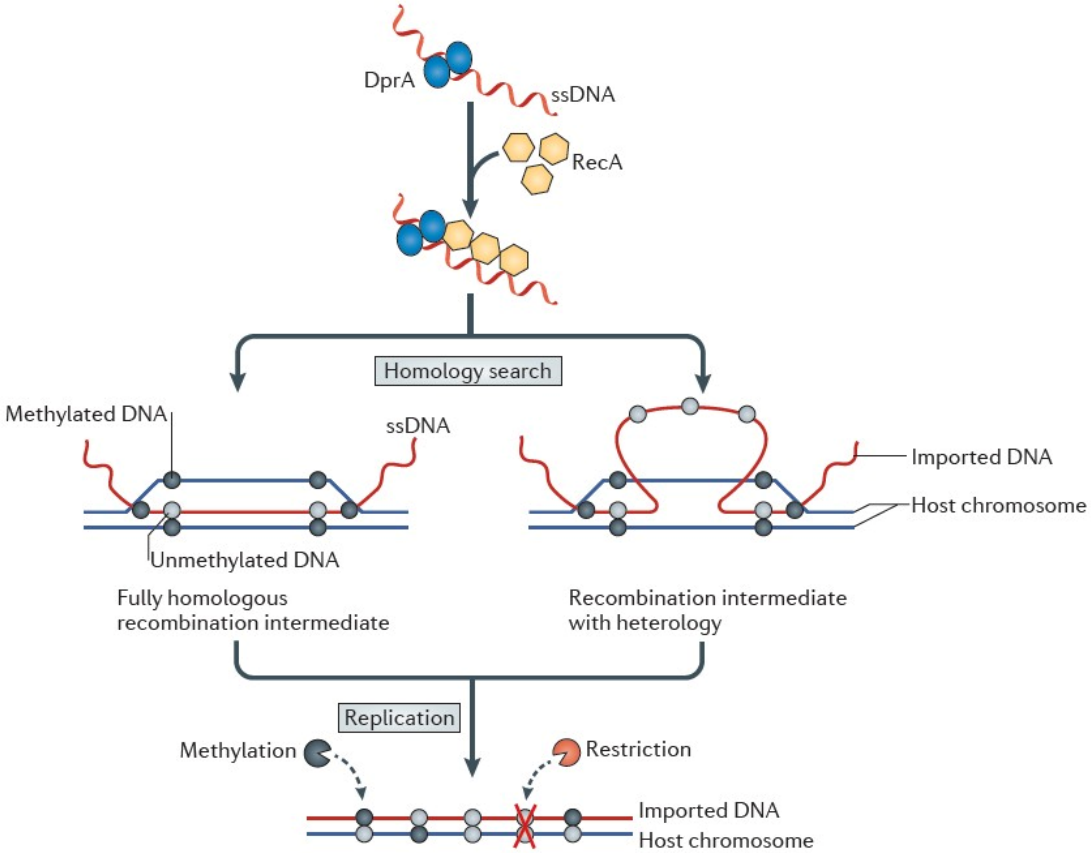


Figure 7 Stabilization of the tDNA by integration in the bacterial chromosome. Internalized ssDNA is presumably bound by DprA (DNA processing protein A) which recruits RecA. RecA polymerizes on ssDNA and will perform a homology search. On the left path, the tDNA is fully homologous to a chromosomic region, while on the right path, it is partially heterologous. Finally, the restriction-modification system may specifically limit natural transformation by degrading imported DNA whose methylation profile does not match the one of the recipient cell. Adapted from Johnston et al (12).

The regulatory systems of competence being extremely varied and complex, we will focus in the remainder of this work, on the regulatory system in streptococci and more particularly on the ComRS system.

3 Characteristics and origin of the genus *Streptococcus*

Streptococcus is a genus that belongs to the family *Streptococcaceae*, within the order *Lactobacillales*, in the phylum *Bacillota*, also named *Firmicutes* (41). The genus *Streptococcus* (from the ancient Greek "streptós", i.e. "having the form of a twisted chain") (42) consists of catalase-negative, Gram-positive cocci arranged in pairs and linear chains, aerotolerant, and unable to grow at 10°C, at pH 9.6 or in 6.5% NaCl broth (43–45).

The definition of the genus *Streptococcus* has been extensively modified throughout history. Traditionally, the member species of this genus were differentiated by qualitative tests with little discriminative value, such as growth tests on different substrates, in anaerobia or not, or tolerance tests to different conditions: pH, temperature, salinity, etc. (43). Microbial cultures were also spread on blood agar plates to observe the presence of haemolysis, a distinctive feature of several pathogenic streptococci. The use of serological tests, popularized by Lancefield in 1933, was also common (46). Later on, these approaches were mostly replaced by biochemical phenotypic tests (see Fig. 8). Although practical in clinical microbiology for identifying bacteria representing an immediate threat to patient health, these methods do not allow rigorous classification of species whose groupings, sometimes still used today, contain many exceptions. Finally, since the end of the 20th century, the classification of bacteria evolved from a phenotypic classification to a polyphasic classification (47) based mainly on a phylogenetic approach relying primarily on comparative analyses of 16S rRNA gene sequences (48).

The genus *Streptococcus* includes several species of interest in various fields ranging from clinical microbiology to the dairy products industry. Several species of streptococci are thus agents associated with various contagious human diseases or are opportunistic germs responsible for diseases in immunodeficient individuals. *Streptococcus pyogenes*, a beta-haemolytic *Streptococcus* is the most pathogenic species in the genus. It is often associated with pharyngitis, impetigo in young children and other severe invasive diseases. Among the alpha-haemolytic species is *S. pneumoniae*, the main causative agent of community-acquired pneumonia and a point of concern for the prevalence of multidrug resistance (43). The old but still used term "viridans streptococci" (from Latin "viridis" meaning "green") groups together 26 alpha-haemolytic and gamma-haemolytic species based on phenotypic criteria (43,49). It includes five subgroups of mainly opportunistic commensal streptococci: anginosus, mitis, salivarius, mutans and sanguinis. These species are also sometimes called "oral streptococci", because they are the dominant bacterial inhabitants in the oropharynx, although the native niche of many of these species is not oral.

TABLE 3. Identification of non-beta-hemolytic gram-positive cocci in chains[a]

Species or group	Antigen	Opt	BS	BE	Na	Pyr	Esc	Vp	Man	Mel	Sbl	Tre	St	Dx	Origin
S. pneumoniae	pn	+	+	−	−	−	v	−	−	+	−	v	−	−	Human
S. equinus (*S. bovis*)	D	−	−	+	−	−	+	+	−	−	−	v	−	−	Equine, bovine
S. gallolyticus (*S. bovis* I)	D	−	−	+	−	−	+	+	+	+	−	+	+	+	Human, koala bovine
S. pasteurianus (*S. bovis* II.2)	D	−	−	+	−	−	+	+	−	+	−	+	−	−	
S. infantarius (*S. bovis* II/1)	D(v)	−	−	−	−	−	v	+	−	+	−	−	+	−	Human, bovine
S. lutetiensis	D(v)	−	−	+	−	−	+	+	−	−	−	−	v	−	
S. suis[b]	Type 1–35 (R,S,T)	−	−	−	−	−	+	−	−	−	−	+	+	−	Swine, human
Viridans streptococci	A, C, G, F, none	−	−	−	−	−	v	v	v	v	v	v	v	v	Human
Other streptococci and genera	Unknown	−	−	v	v	v	v	v	v	v	v	v	v	v	Animal, human

[a] Abbreviations: pn, pneumococcal typing antiserum or Omni serum; letters, Lancefield group antigen; Opt, optochin; BS, bile solubility; BE, bile-esculin reaction; Na, growth in 6.5% NaCl broth; Pyr, pyrrolidonylarylamidase reaction; Esc, hyrolysis of esculin; Vp, Voges-Proskauer reaction; Man, Mel, Sbl, Tre, acidification of mannitol, melibiose, sorbitol, and trehalose broths; respectively; St, hydrolysis of starch; Dx, production of extracellular polysaccharide. See footnote in Table 2 for positive and negative reactions.

[b] Note that the only way to differentiate *S. suis* from viridans streptococci is by serologic typing.

Figure 8 Identification table of streptococci based on phenotypic characteristics. Adapted from Facklam R. (2002) (43).

4 Natural competence in streptococci

The genus *Streptococcus* includes a large number of Gram-positive species identified or predicted as competent. It includes *S. pneumoniae*, the species where the transformation phenomenon was observed for the first time, and a model organism for natural transformation and competence regulation in Gram-positive cocci.

The ability of naturally competent streptococci to exchange genetic material between closely-related species is a matter of concern in the medical field, particularly regarding the propagation of virulence factors and antibiotic resistances. The term "supragenome" (8) is commonly used to define a collective gene pool containing allelic variants and genes that are absent from the genome of a species, but accessible by HGT. Access to a supragenome would allow bacteria to circumvent occasionally the limited size of their genome. The worldwide increase in penicillin resistance in *S. pneumoniae* strains observed in hospitals in the 1990s would thus be the result of HGTs with other streptococci indigenous to the upper respiratory tract microflora, namely *Streptococcus mitis* and *Streptococcus oralis* (50).

The regulatory system of competence in streptococci is no exception regarding its complexity. In *S. pneumoniae*, it consists of a large network of genes (its *com* regulon incudes approximately 80 genes) (51) coupled to several QS systems.

4.1 Competence regulation in streptococci

4.1.1 σ^X and the early and late competence genes

In the genera *Streptococcus* and *Bacillus*, the set of genes whose expression is coupled to natural competence, the “*com* regulon” can be divided into two categories (8), the "early genes" which group genes responsible for sensing conditions favorable to the development of the competent state, signal transduction, and communication of this signal to neighboring cells, and the “late genes” responsible for the expression of the transformasome and the recombination machinery as well as other functions associated with the broader physiological response (Fig. 9) (7) At the intersection of these two groups, a central regulator ensures the link between these two phases of natural competence development. In *S. pneumoniae*, and probably in all streptococci, it is the alternative sigma factor σ^X which, by forming a complex with the RNA polymerase, will

target a specific motif, named the "σ^X-box", in the promoters of all late genes of the "*com* regulon". These genes include 14 essential late competence genes encoding the transformasome and additional species-specific genes responsible for competence-associated processes. The amount of σ^X is the main factor responsible for the entry into the competence state. Its abundance is proximally regulated at the transcriptional level by a QS system, and at the post-translational level by its degradation or stabilization (34). Alternatively, "early competence genes" can also be defined as being upstream of σ^X in the competence signaling pathway and "late competence genes" as being downstream.

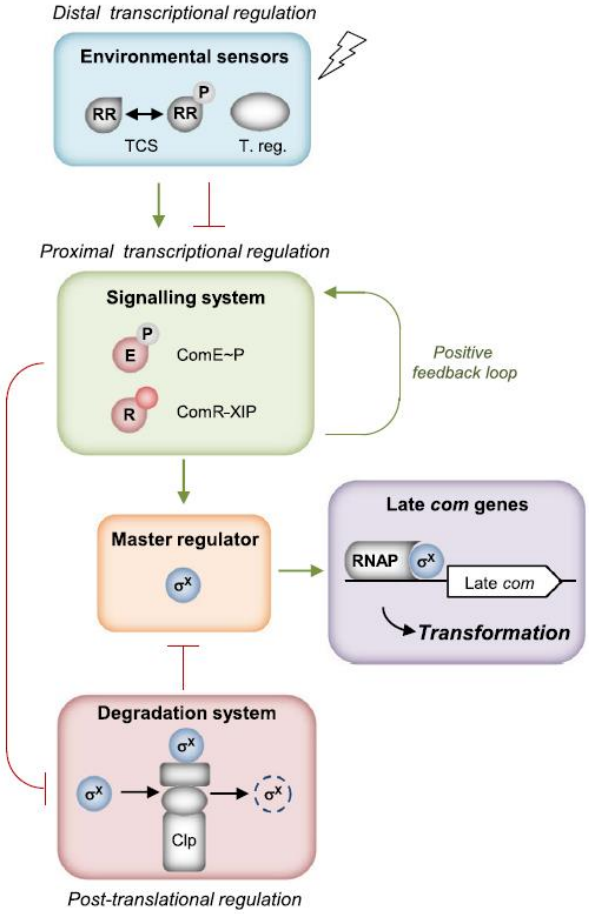


Figure 9 Overview of competence regulation in streptococci. In streptococci, competence development is activated by the master regulator σ^X whose abundance is regulated at the transcriptional level by a peptide pheromone QS signaling system, the ComCDE system or the ComRS system depending on the species, and at the post-transcriptional level by the proteolytic system Clp. ComE~P is the active phosphorylated form of the transcription factor upregulating σ^X expression in the ComCDE system. Similarly, ComR-XIP is a complex which consists of the transcription factor ComR bound to the pheromone XIP, in the ComRS system. The signaling system is auto-activated (positive feedback loop) and could also presumably inhibit the action of the Clp degradation system, thus stabilizing σ^X. The activity of the proximal system is also tuned by distal transcriptional regulatory systems which sense several environmental signals. These systems might include two-component systems (TCS) and transcriptional regulators (T. Reg) for example. Green arrows mean transcription activation while red lines suggest inhibition mechanisms. Adapted from Fontaine et al. (2015) (34).

4.1.2 Prevalence of the ComCDE and ComRS systems in streptococci

Two types of competence cell-cell signalling systems are described in the literature for streptococci. These are the ComCDE system (mainly studied in *S. pneumoniae*) and the ComRS system for which model organisms are *Streptococcus thermophilus* and *Streptococcus mutans*. The ComCDE system controls σ^X expression in the mitis and anginosus groups of streptococci, while the ComRS system is found in the salivarius, bovis, mutans, suis and pyogenic groups (Fig. 10) (34,52).

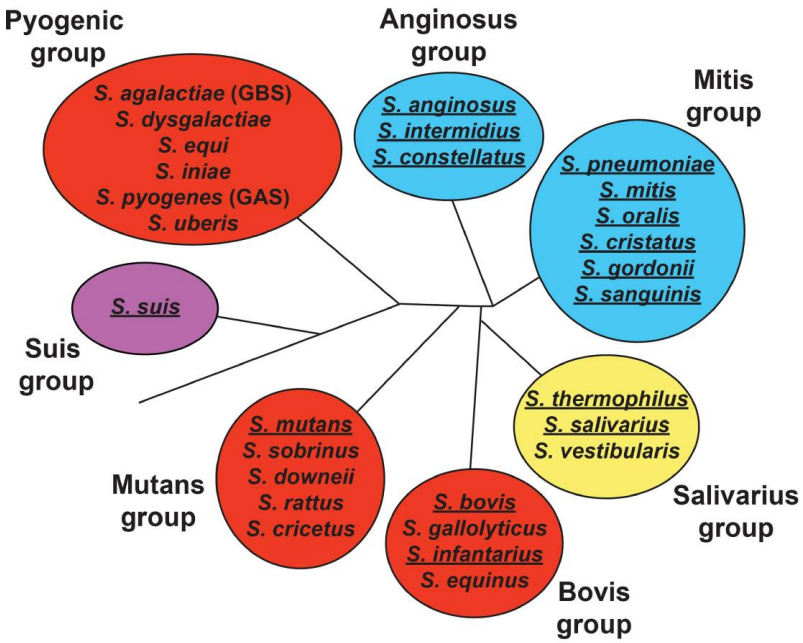


Figure 10 Distribution of the ComCDE and ComRS competence regulation systems in streptococci. ComCDE is present in the *mitis* and *anginosus* groups of streptococci (displayed in blue). Type I ComRS pathway is found in the *salivarius* group (in yellow). The type II ComRS system is found in the *pyogenic*, *mutans* and *bovis* groups (in red), while *S. suis* is the only species belonging to the type III ComRS system (in violet). Species for which natural competence has been experimentally demonstrated, according to the state of knowledge in 2016, are underlined. Adapted from Shanker et al. (2016) (52).

4.1.3 Competence regulation by the ComRS system

The ComRS system is a σ^X regulatory pathway in streptococci discovered in 2010 by Fontaine et al. in *S. thermophilus*. In the ComRS system, proximal regulation of the expression of *comX*, the gene encoding the alternative sigma factor σ^X, is driven directly by the transcription factor ComR, an intracellular receptor, and XIP, a peptide pheromone belonging to the family of auto-inducers (53).

At the beginning of the cycle (Fig. 11), ComS, a peptide of about 20 amino acids with no apparent secreting signature (53,54), was shown to be secreted through the ABC transporter PptAB in *S. thermophilus* and *Streptococcus salivarius* (55). Concomitantly to its secretion, it will be matured by proteolytic cleavage by species-specific peptidases, such as the membrane-anchored protease Eep, in the salivarius group of streptococci (34). Its released C-terminal part corresponds to its active form, called "XIP" for SigX-inducing peptide, SigX being a synonym of the sigma factor X, "σ^X". As cell density increases, the extracellular concentration of XIP increases too until a critical concentration is reached. Once this threshold is exceeded, XIP peptides will preferentially bind to a specific substrate-binding lipoprotein named OppA or AmiA on the cell surface (53,56), which will allow their re-import into the cytoplasm by way of generalized transporters of the oligopeptide permease Opp/Ami family. XIP will then interact with the transcriptional activator ComR with which it will form a homodimeric complex composed of two XIP molecules and two ComR receptors (57). In this active form, the ComR-XIP complex binds to a DNA motif, called the "ComR-box", present in the promoter of *comS* and *comX*, thus stimulating their expression. The increase in *comS* transcription from this point onwards will create a positive feedback loop on its own expression. The expression of *comR* remains unchanged because its promoter does not contain a ComR-box. Instead, *comR* expression is the target of distal competence regulatory systems (57). In the salivarius group of streptococci, the CovRS system, an environmental sensor, directly

repress *comR* expression (58). Finally, the stimulation of σ^X production will trigger the expression of late competence genes responsible for the synthesis of the transformasome and consequently the onset of the competent state.

The shutdown of competence regulated by the ComRS system is not entirely understood. In *S. salivarius*, the oligoendopeptidase PepF has been identified as a major actor of competence shut-off by degrading XIP pheromones (Fig. 11) (59). PepF expression is upregulated by ComX, enabling it to repress ComR·XIP formation in the late phase of competence, whereas it acts as a locking mechanism at basal expression, in the early phase of competence (59). In both *S. thermophilus* and *S. mutans*, the MecA-ClpCP proteolytic complex degrading σ^X acts as a negative regulator of competence development, but its action consists more in preventing the onset of competence under subpermissive conditions than in stopping it once the response has been initiated (60). The competence regulation mechanism by the ComRS system is conserved in all streptococcal species of the groups previously-mentioned, even if some differences exist, especially regarding the pheromones sequences and their maturation process, or the mechanisms involved in the competence shut-off system (34).

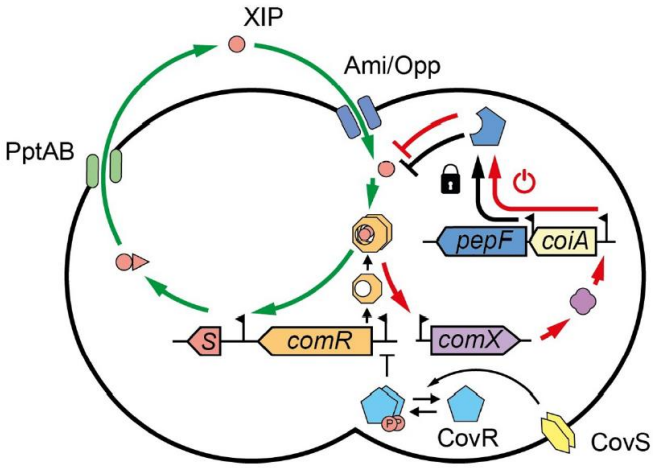


Figure 11 O Overview of the ComRS signaling system controlling *comX* expression in *S. salivarius*. The ComS propeptide (in red) is matured by species-specific proteases concomitantly to its secretion through the ABC transporter PptAB. XIP, the resulting processed form is re-imported into the cytosol by the oligopeptide transporter Ami/Opp where it interacts with the transcriptional regulator ComR (in orange). The resulting ComR·XIP complex activates the transcription of *comS* (creating an auto-inducing loop, green arrows) and *comX* encoding the alternative sigma factor σ^X. In association with RNA polymerase, σ^X allows the expression of late *com* genes and initiates natural competence. Simultaneously, σ^X increases expression of the endopeptidase PepF involved in the competence shut-off mechanism by degrading XIP peptides (red arrows). At basal expression levels, PepF acts rather as a locking system preventing competence development (black arrow in bold). The *comR* gene is not part of the ComRS regulon. Instead, *comR* expression is the target of distal regulatory system, such as the CovRS environmental sensor system which acts as an inhibitor of its expression. Adapted from Knoops et al. (2022) (59).

The continuation of this work will be dedicated to the study of the ComRS system of *S. thermophilus* and to a lesser extent, to the one of *S. vestibularis*.

5 The salivarius group of streptococci

The salivarius group is one of five categories of non-beta-hemolytic viridans streptococci formed on the basis of similar phenotypic characteristics. It contains three genetically related species: *S. salivarius*, *Streptococcus vestibularius* and *S. thermophilus* (43,49). *S. salivarius* and S. *vestibularis* are commensal species that can be opportunistically associated with human diseases, while *S. thermophilus* is a non-pathogenic bacterium used extensively in the dairy industry.

5.1 *Streptococcus thermophilus*

S. thermophilus was first described by Orla-Jensen et al. in 1919. For some time, it was considered a subspecies of *S. salivarius*, but finally received full species status in 1991 (61). Its original habitat is bovine mammary mucosa and raw milk. *S. thermophilus* is a thermophilic lactic acid bacterium and is the only species of the genus *Streptococcus* to have received the "Generally Regarded As Safe" (GRAS) status from the U.S. Food and Drug Administration (FDA) (62).

It is the second most important industrial dairy starter after *L. lactis* with an estimated market value of 40 billion US$ in 2008 (63). Often paired with *Lactobacillus delbrueckii* subsp*. bulgaricus* or *Lactobacillus helveticus* in dairy products (62), it is generally associated with fermentation processes related to the production of cheese and yoghurt where its ability to resist temperatures of 45°C is valued (44). In dairy starters, it is mainly responsible for the rapid acidification of milk during lactic fermentation and therefore has an important role in food biopreservation. In addition to its capacity to produce lactic acid, *S. thermophilus* confers gustative and textural properties to dairy products, notably through its production of acetaldehyde and exopolysaccharides. Its probiotic action is also praised (63).

S. thermophilus possesses a relatively small genome size (~1.8 Mbps vs. ~2.3 Mbps for *S. salivarius* and its ORFeome contains ~10% pseudogenes, revealing an evolutionary reduction of its genome associated to its specialization to a nutrient-rich milk environment (49,62). Among the lost functions are genes responsible for carbohydrate metabolism and virulence. On the counterparts, *S. thermophilus* has a metabolism specialized in lactose utilization and in the import of extracellular peptides and amino acids. Comparative genomic analyses also reveal a low polymorphism between different strains of *S. thermophilus*, which would suggest that *S. thermophilus* is a recently emerged species. Many of the functions unique to *S. thermophilus* have no counterparts in the genus *Streptococcus* and are thought to result from HGTs from other lactic acid bacteria.

The natural competence of *S. thermophilus* was discovered in 2009 by Gardan et al. (64). *S. thermophilus* is able to spontaneously develop competence on a chemically defined medium (CDM) containing free amino acids but no oligopeptides, and in milk, its ecological niche (56).

5.2 *Streptococcus vestibularis*

S. vestibularis is a commensal species that has recently been included in the viridans group of streptococci (65). It was first isolated from the vestibular mucosa of the human oral cavity and therefore shares the same environmental niche as many streptococci and staphylococci. This species is poorly described in the scientific literature even though it can adopt a pathogenic behavior opportunistically. For instance, *S. vestibularis* was reported to cause infectious endocarditis (66) in association with *S. oralis*, and dental caries (67).

S. vestibularis is phylogenetically closer to *S. salivarius* than to *S. thermophilus* (49). Its competence was also observed on CDM, but only after artificial induction with its corresponding XIP pheromone (68).

6 ComRS regulation in *S. thermophilus*

6.1 Kinetics of competence induction in *S. thermophilus*

The kinetics of competence induction in the ComRS system (Fig. 12) vary greatly depending on the species considered and are mainly derived from experiments on CDM (34). *S. thermophilus* LMD-9 is spontaneously transformable on CDM. Similar to *S. pneumoniae*, the expression of early *com* genes in *S. thermophilus* is transient and starts at the early exponential phase over a wide range of initial cell densities. For this reason, the triggering of the competent state in the ComRS system can be described as a timing device rather than a quorum sensing in a strict sense (56).

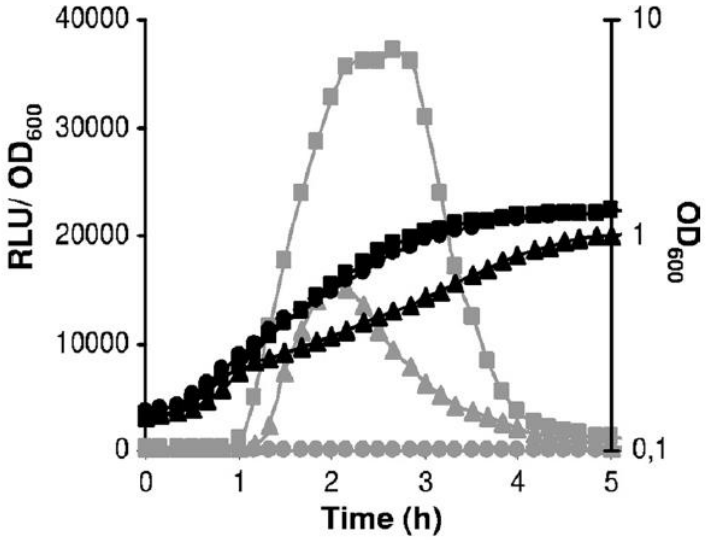


Figure 12 Kinetics of *comX* induction in *S. thermophilus* LMD-9 inferred from *in vivo* luciferase experiments. Growth (OD_{600}) and luciferase activities (RLU/OD_{600}) of *S.th* reporter strains in CDM over time (h). In gray, RLU/OD_{600} represents the luminescence emitted by the isogenic reporter strains weighted by their cell density. In black, OD_{600} represents the cell density of the culture. In each of the three strains, a transcriptional fusion between the *comX* promoter (P_{comX}) and the luciferase genes *luxAB* of *Photorabdus luminescens* is chromosomally inserted. Triangles represent a strain whose growth and transformability are close to *wild type*. Upon induction of *comX*, the growth of this strain is slowed down, demonstrating the energy cost of natural competence. Squares represent a strain whose gene encoding the central regulator of competence (*comX*) is deleted. Its longer and more intense luminescence is explained by the partial absence of a post-transcriptional shut-off mechanism related to the absence of *comX*. It is not transformable and its growth is not affected. Circles represent a strain with a deleted Ami/Opp oligopeptide transporter essential for transformation. No luminescence is emitted because the luciferase reporter gene is not expressed. Like the previous strain, it is not transformable and its growth is not affected. Adapted from Fontaine et al. (2010) (53).

6.2 Genetic organization of the ComRS core regulon in *S. thermophilus*

In *S. thermophilus*, as in other streptococci with a ComRS system, *comR* and *comS* form a cluster with *comS* always downstream of *comR* (34,69). The genes whose expression is directly regulated by the ComR·XIP complex form the ComRS core regulon. In *S. thermophilus* strain LMD-9, it consists of 17 genes with functions related to competence regulation and bacteriocin immunity and production (see Fig. 13) (57). Since *comR* transcription is not under the control of the ComR·XIP complex, it is not part of this regulon. *S. thermophilus* and *S. mutans* have only one copy of *comX* while other streptococci may have two or three.

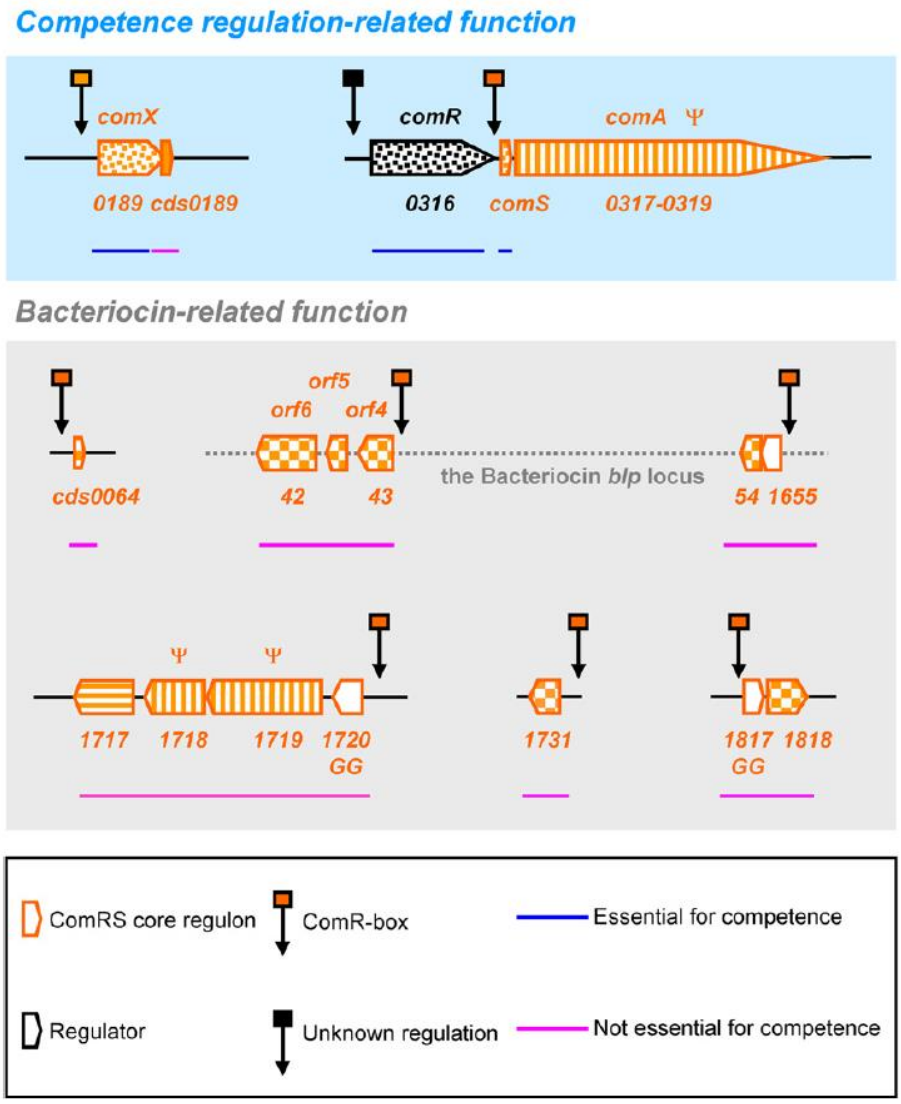


Figure 13 Genetic organization of the ComRS core regulon in *S. thermophilus* LMD-9. The genes composing the ComRS core regulon are grouped here according to their (predicted) physiological function. The light blue frame surrounds genes having a competence regulation-related function. The grey frame surrounds genes with a function related to bacteriocin immunity and production. The ComRS core regulon is composed of 17 genes colored in orange (*comR* not being part of it, it appears here in black). The presence of a ComR-box in the promoter region is represented with an orange box and a vertical arrow. The function of the encoded proteins is represented by the gene filling patterns: full coloration, no predicted function; no pattern, (putative) bacteriocin precursor; squares, (putative) bacteriocin immunity peptide; dots, regulatory function; vertical lines, (putative) transport and maturation of double-glycine (GG) peptides; horizontal lines, (putative) bacteriocin modification protein. Solid lines below the genes indicate whether their deletion inhibits transformability (blue line) or not (pink line). The grey dashed line represents the bacteriocin *blp* locus. Numbers below genes correspond to the STER locus tag. Ψ indicates a pseudogene. Adapted from Fontaine et al. (2013) (57)

The motif recognized by the ComR·XIP complex in the promoters of the core ComRS regulon genes is called the ComR-box (Fig. 14) or Ecom-box (for early com genes). The consensus sequence of the 8 ComR-boxes of *S. thermophilus* is described below. The ComR-box is a 20 bp palindromic motif characterized by a conserved inner GACA/TGTC inverted repeat essential for transcription (57).

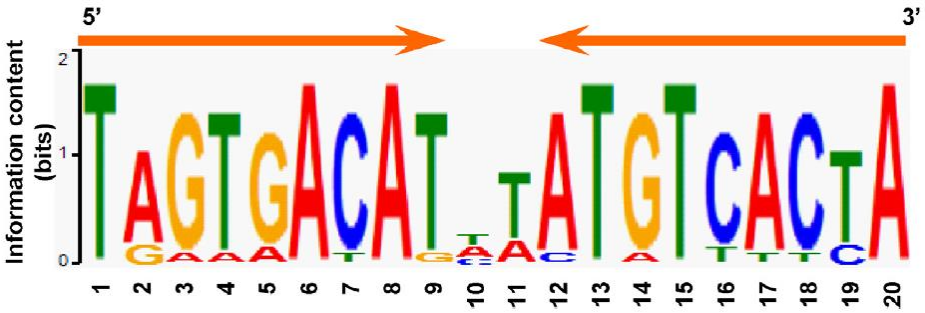


Figure 14 Logo representation of the position weight matrix for the 20 bp ComR-box motif based on the eight promoters of the ComRS core regulon of *S. thermophilus* LMD-9. Adapted from Fontaine et al. (2013) (57).

6.3 The pheromone peptide XIP of *S. thermophilus*

The linear peptide pheromone XIP is the matured form of the ComS precursor after secretion and proteolytic cleavage. After secretion, XIP_{Sth} do not diffuse freely in the extracellular medium and remain close to the cell surface from where they propagate the signal from cell to cell by close contacts. However, it was possible to retrieve sufficient XIPs from supernatants of mutated strains overexpressing *comS* to analyze its sequence by mass spectrometry (56).

The natural form of XIP from *S. thermophilus* LMD-9 corresponds to the undecapeptide $ComS_{14-24}$: IAILPYFAGCL. Nevertheless, a minimal form with the same competence induction properties and corresponding to the octapeptide $ComS_{17-24}$ is used in experiments by convenience.

The full sequence of the ComS precursor (24 amino acids) of *S. thermophilus* strain LMD-9 is shown below with the native mature form XIP underlined and the minimal form of XIP in bold.

$ComS_{Sth}$: MKTLKIFVLFSLLIAI**LPYFAGCL**

ComRS systems are classified into 3 categories according to the sequence of their XIP pheromones. Type I XIPs are characterized by the presence of a (V/L)P (F/Y)F motif and their absence of charged residues. They are found in the salivarius group of streptococci to which *S. thermophilus* and *S. vestibularis* belong. Type II XIPs are characterized by the presence of a WW motif and the presence of one or two charged acidic and/or basic residues. They are found in the bovis, pyogenes and mutans groups (54). Finally, the XIPs of *S. suis* strains are the only ones not characterized by the presence of two consecutive aromatic residues. They have a WG (T/K)W motif instead and are therefore classified as type III XIPs (34,70).

XIP_{Sth} belongs to type I ComRS systems. According to the definition, it is characterized by the presence of a leucine at the N-terminal position (L_{17}), a proline (P_{18}), two contiguous aromatic amino acids, Y_{19} and F_{20}, and does not possess charged residues. It also has a cysteine (C_{23}) that must remain unmodified to maintain its activity and a branched-chain amino acid at its C-terminal end, L_{24}. The residues identified by an alanine-scan as the most critical for its activity are Y_{19}, F_{20} and L_{24} (Fig. 15) (57).

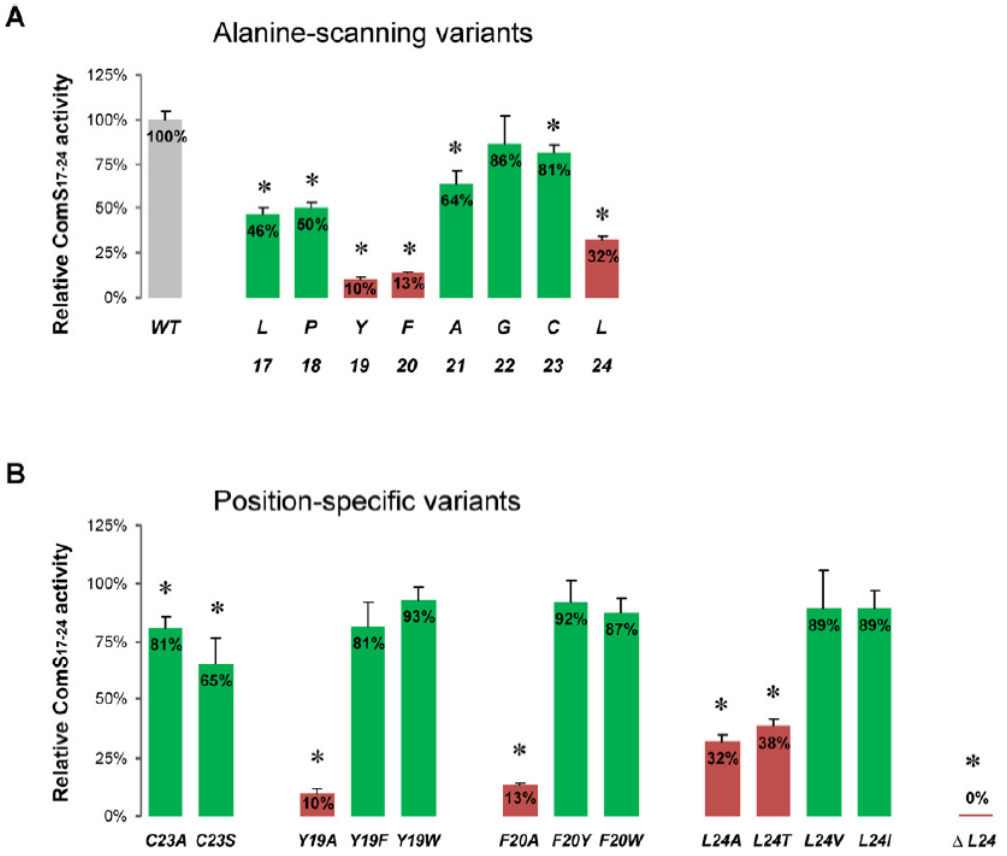


Figure 15 Importance of $ComS_{17-24}$ (XIP) residues for activation of ComR in *S. thermophilus* LMD-9. Relative ComS activity represents the percentage of luminescence emitted by a reporter strain containing a luciferase gene whose expression is under the control of the ComR-XIP complex, upon addition of a mutated $ComS_{17-24}$ variant compared to the addition of wild type $ComS_{17-24}$. (A) Impact of the replacement of each amino acid residue of $ComS_{17-24}$ with alanine, with the exception of A21, which was substituted by a glycine. (B) Impact of the replacement of the three previously identified $ComS_{17-24}$ critical amino acid residues for ComR activation, Y_{19}, F_{20} and L_{24}, as well as C_{23}. Results are shown as the mean of three independent replicates. Error bars represent the standard deviation. The asterisk '*' indicates a significant difference ($P < 0.05$) compared to $ComS_{17–24}$ (Student's *t*-test). The colors of the histogram bars indicate whether ComR is able (in green) to bind to a DNA probe containing the binding motif of the ComR-XIP complex (ComR-box), or not (in red), in electrophoretic mobility shift assays (EMSA) upon addition of the $ComS_{17-24}$ variant. Adapted from Fontaine et al. (2013) (57).

6.4 The pheromone peptide XIP of *S. vestibularis*

The natural form of XIP from *S. vestibularis* is not known. A minimal form able to artificially induce competence in the laboratory corresponds to the octapeptide $ComS_{15-22}$.

The full sequence of the ComS precursor (22 amino acids) of *S. vestibularis* strain F0396 XIP is shown below with the minimal form in bold.

$ComS_{Sve}$: MKNLKKFLVLLIAA**VPFFMIYY**

XIP_{Sve} also belongs to the ComRS type I system. It has a N-terminal valine and a proline (V_{15} and P_{16}), two contiguous aromatic residues, F_{17} and F_{18}, and has no charged residues in its theoretical processed form. Unlike XIP_{Sth}, it does not have a C-terminal branched-chained amino acid but has an isoleucine (I_{20}) near its end.

6.5 Structure of *S. thermophilus* ComR

ComR is a cytoplasmic positive regulator of natural competence in streptococci, activated by the pheromone peptide XIP. With a length of 299 amino acids in *S. thermophilus*, it is characterized by a N-terminal helix-turn-helix (HTH)-DNA-binding domain and a C-terminal peptide-binding domain featuring tetratricopeptide repeats (TPR). A tetratricopeptide repeat is a degenerated structural motif of 34 residues mediating protein-protein or protein-peptide interactions (71). In ComR and related cytoplasmic peptide sensors, each TPR motif is formed by the combination of 2 antiparallel α-helices (72). The HTH (66 aa) and TPR (224 aa) domains are composed of 5 and 11 α-helices respectively (see Fig. 16), connected by a 9-aa linker. The α-helices of the C-terminal domain are grouped in 5 pairs forming each one a TPR, while the last helix (α16) is referred as the capping helix, also called "CAP" (72).

The QS regulators PlcR from *Bacillus thuringensis* and PrgX from *Enterococcus faecalis* are the best ComR homologs based on structural predictions (34,53,54). They belong to the pheromone-responsive regulators RRNPP superfamily (73). ComR is a member of the RRNPP, but there is currently a debate as to whether ComR is the representant of a new family of regulators or whether it should be considered as a Rgg-like regulator (69,72,73).

Interestingly, DNA-binding HTH domains and XIP-binding TPR domains evolved at different paces within the ComR orthologues of different species. The HTH domains are more conserved than the TPR domains which have presumably undergone a selective evolutionary pressure and have co-evolved with their cognate XIP peptide pheromones. The phylogenetic distribution of ComR thus follows the types of ComRS systems defined above on the basis of the XIP sequence (34).

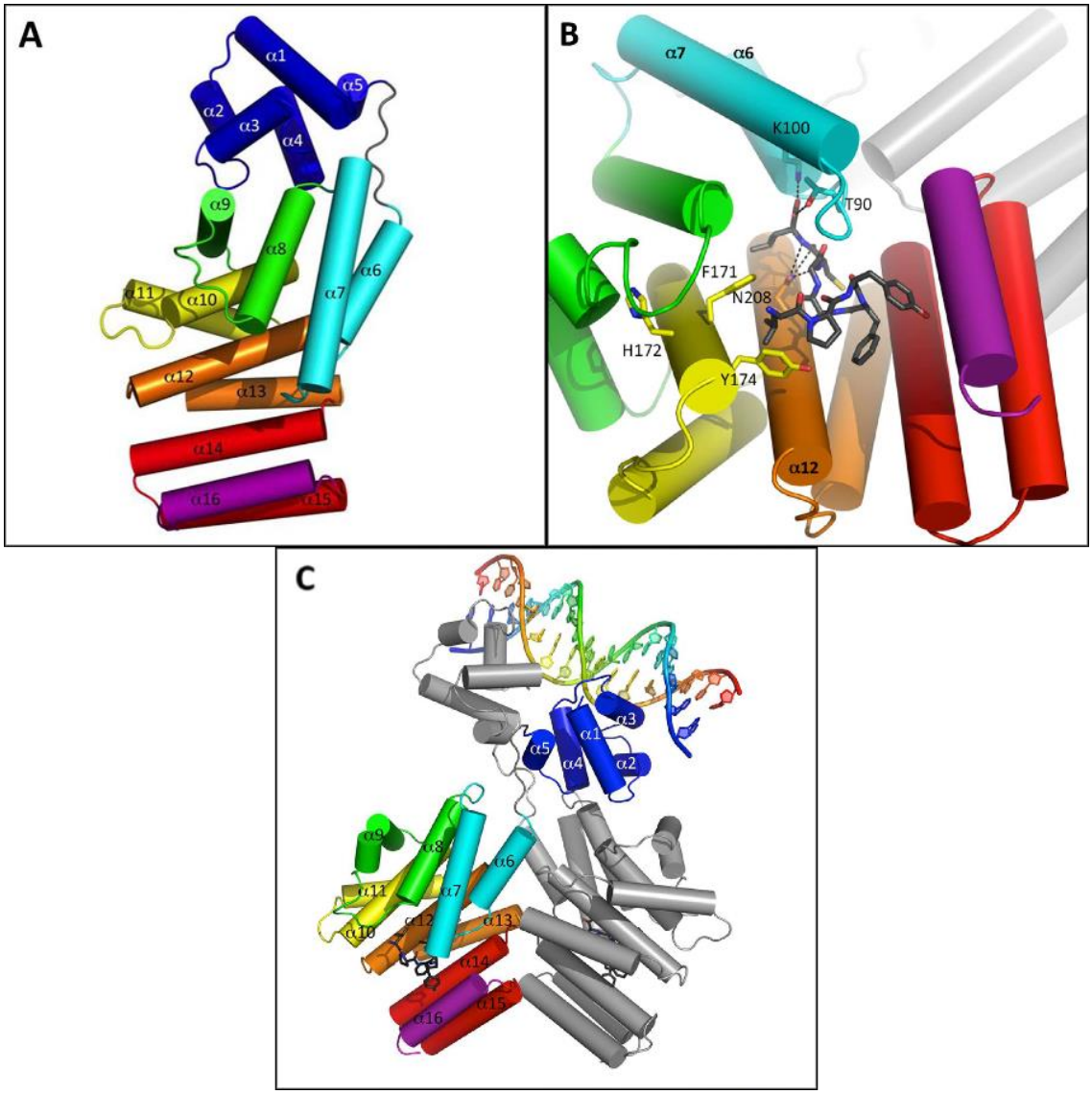


Figure 16 3D structures of ComR and XIP of *S. thermophilus*. (A) Overview of ComR in the apo conformation. ComR is shown as cartoon colored by spectrum from TPR-1 in cyan to TPR-5 in red with the additional CAP helix α16 in purple and the HTH-domain in blue. The α-helices are labelled. (B) Close view of XIP bound to ComR, inside the binding pocket. ComR is represented as in (A). The bound peptide is shown as sticks colored by atom type. Key residues of ComR in its interaction with the bound XIP peptide are highlighted in sticks. Hydrogen bonding between K_{100}, T_{90} and N_{208} and XIP appear as dashed lines. The second chain of the ComR dimer appearing in the background is shown as cartoon colored in gray. (C) Overview of the ComR/XIP/DNA ternary complex. ComR is represented as in (A). The second chain of the ComR dimer is shown as cartoon colored in gray. The 20-bp DNA fragment from the *comX* promoter is shown as cartoon colored by spectrum, from 5' in blue to 3' in red. Bound XIP peptides are represented as black sticks. Adapted from Talagas et al. (2016) (72).

6.6 ComR activation by the XIP peptide

The activation mechanism of the transcriptional regulator ComR (Fig. 17) is strictly dependent on the binding of the XIP peptide and is conserved among ComR orthologs (52,72). The apo form of ComR is a monomer whose DNA-binding-HTH domain is sequestered by the TPR domain. Upon binding of XIP in the binding pocket, the TPR domain will undergo a drastic conformational change marked in particular by the breaking of α9 to adopt a more closed structure. This conformational change will release the HTH domain and allow the dimerization of ComR, which is essential for its activation. The ComR-XIP complex is then in its active form and the free HTH-domains specifically recognize the pseudo-palindromic DNA sequence of the ComR-box and its conserved inner GACA/TGTC inverted repeat to form the ternary ComR-XIP-DNA complex. During

this step, the flexibility of the HTH domains of the ComR-XIP dimer allows their reorientation and a domain swapping compatible with DNA binding. The domain-swapped form of the ternary complex is stabilized both by interactions of the HTHs with the inverted repeats of the ComR-box, and by cross interactions between the HTH and TPR domains of the two monomers. The conformation of the ternary complex is thus more compact than the initial apo form.

Sequestration of the DNA-binding domain and its release under strict control of peptide binding is a novel control mechanism for a transcriptional regulator. HTH domain sequestration is not found in Rgg proteins, and most members of the RRNPP superfamily already exist as domain-swapped dimers in their apo form. This mode of activation is well suited to the function of regulating natural competence, which is an energy-intensive and potentially deleterious physiological process.

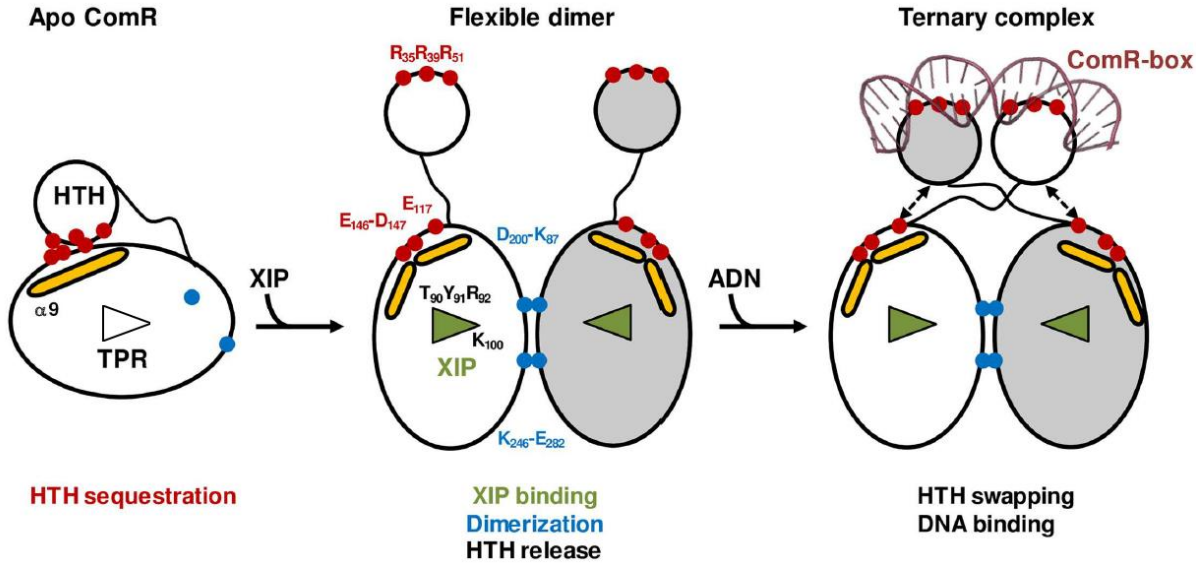


Figure 17 Mechanism of ComR activation by the XIP peptide. From left to right. Representation of the inactive form of ComR in its apo conformation. The HTH-DNA-binding domain (small circle) is sequestered by the TPR domain (large oval). Key residues for HTH sequestration are indicated in red. After binding of the XIP peptide (green triangle) to ComR in the TPR binding pocket (empty triangle), the TPR domain changes conformation, in particular the α9 "breaks", which releases the HTH domain and leads to the dimerization of ComR at the TPR domains. Key residues for ComR dimerization are indicated in blue. The flexible dimer form corresponds to the active form of ComR. Finally, the ComR dimer binds at the HTH domains to a 20 bp pseudo-palindromic DNA sequence, the ComR-box, to form a ComR-XIP-DNA ternary complex. Upon DNA binding, the HTH domains of the two monomers swap and cross interactions between the HTH and TPR domains of the two proteins (indicated by two double arrows) contract the structure of the complex. Adapted from Talagas et al. (2016) (72).

6.7 Specificity of ComR activation by heterologous peptides

Consistent with the coevolution of ComR's XIP-binding TPR domains with their cognate peptides, ComRs display at least some specificity towards their own peptide. However, it may be interesting for a species to be able to detect heterologous pheromones to link natural competence and other species-specific functions (such as bacteriocin production) (53,54) to the presence of other species (57). It has been reported that the ComRs of certain species show greater permissiveness in recognizing peptides with different sequences. The ComR of *S. thermophilus* LMD-9 is able to recognize several mutant XIPs derived from an alanine scan with a certain degree of flexibility (see Fig. 15 (57)). Cross-species activation has been observed between $ComR_{Sth}$ from *S. thermophilus* LMD-9 and $ComR_{Ssa}$ from *S. salivarius* SK126 (53). Similarly, competence was restored in a *S. thermophilus* ComS-deficient strain by the addition of casein-derived peptides in a milk-based medium mimicking its native habitat (57).

In a cross-talk analysis carried out on type II and type III ComRS systems, the ComRs analyzed were classified as strict, intermediate or promiscuous, according to their ability to recognize XIPs from other species (52). Although ComRs from the same species fall into the same permissiveness category, differences in selectivity exist within the same species, demonstrating that the specificity of heterologous peptide recognition is not strictly correlated with phylogenetic relationships.

From a structural point of view, Shanker et al. (2016) (52) hypothesize that some of the most conserved ComR pocket residues are required for XIP binding, while some of the more variable pocket aa are responsible for ComR specificity and further activation. This concept is shared by Talagas et al. (2016) in their structure-function analysis of $ComR_{Sth}$ and its activation mechanism by its cognate XIP (72). The differences in selectivity of the different ComRS systems have a physiological significance linked to their ecological niche that is difficult to interpret. In addition to ComR selectivity, it would also be interesting to look at XIP detectability.

Objectives

Natural competence is the genetically specified ability of some bacteria to take up extracellular DNA. Internalized DNA can then be integrated into the genome by homologous recombination to provide an evolutionary advantage, as a means of genome plasticity/repair, or marginally, to be processed in a nutritional role. It is a widespread process in Gram-positive and Gram-negative bacteria which has been intensively studied, particularly in the *Streptococcus* genus. In this genus, development of competence is an energy-intensive transient process based on finely-tuned mechanisms and linked, for instance, to bacteriocin production or biofilm formation (74,75).

The ComRS system is a pheromone quorum-sensing (QS) system found in streptococci of the salivarius, mutans, pyogenes and bovis groups (12). It controls the development of competence by activating the expression of the central regulator of competence, the sigma factor X (also known as σ^X/ComX/SigX), responsible for the expression of late competence genes, including the transformasome and the recombination machinery (34). ComS is the precursor of the competence pheromone peptide. After secretion and maturation, the matured ComS form (named XIP) is reinternalized and directly interacts with the transcriptional regulator ComR that induces the formation of a dimeric active complex (ComR·XIP). When the concentration of XIP reaches a certain threshold, the concentration of ComR·XIP allows it to bind a 20-bp motif (known as the ComR-box) present, *inter alia*, in the promoter regions of *comS* and *comX*, activating their expression. Upregulation of *comS* creates a positive feedback loop typical of QS systems, while expression of *comX* marks the onset of competence development for transformation (53).

ComR is a transcriptional activator of natural competence belonging to the RRNPP superfamily of cytoplasmic regulators. It is characterized by an N-terminal helix-turn-helix (HTH) DNA-binding domain and a C-terminal alpha-helical domain containing 5 tetratricopeptide repeats (TPR) which interacts with the XIP pheromone required for its activation (72). On the basis of phylogenetic studies, a different evolutionary pressure is observed on HTH domains, conserved between orthologous ComR, and TPR domains, more diverse in regard to their sequences, which have co-evolved with their XIP. Consistent with this finding, ComRs display specificity for their cognate peptides, but they are also more or less permissive depending on the species, and some are capable of recognizing heterologous peptides of varied sequences (52,57). A better understanding of the structural determinants responsible for ComR selectivity would allow us to better assess the extent of ComRS involvement in interspecies crosstalk among streptococci. Moreover, given the interconnections between natural competence and other physiological processes of high medical importance such as biofilm formation, virulence and antibiotic resistance, the study of regulatory systems for competence development is a promising area for developing new antimicrobial agents.

The first aim of this master thesis project is to identify key structural determinants of ComR involved in the selectivity and activation mechanisms by XIP in *Streptococcus thermophilus*, a model species of the salivarius group of streptococci. The second aim of this project is to obtain $ComR_{Sth}$ mutants (ComR*) with altered XIP recognition capacity. For this purpose, the main strategy is focused on the design of a ComR mutant capable to

induce competence with either XIP of *S. thermophilus*, and/or the one of *Streptococcus vestibularis*, a closely related species also belonging to the salivarius group.
To investigate the recognition capacity of ComR for both XIPs experimentally, their functionality will be assayed *in vivo* by chromosomally substituting $ComR_{Sth}$ wt with ComR* in a *S. thermophilus* LMD-9 strain carrying a luciferase reporter system under the control of a ComRS-dependent promoter. Complementarily, the production of ComR* and their potential toxicity will be verified by western blot and SDS-PAGE. On the other hand, the necessity of ComR interaction with XIP for DNA binding, as well as the formation of the ComR·XIP·DNA complex will be observed *in vitro* via electrophoretic mobility shift assays (EMSA) performed on purified recombinant ComR* proteins expressed in *E. coli*. Finally, to further dissect the effects of ComR mutations on its ability to recognize XIP, luciferase experiments will also be carried out with hybrid XIPs featuring parts of XIP_{Sth} and XIP_{Sve}.

Material and methods

1 Bacterial strains

Streptococcus thermophilus (*S.th.*) LMD-9

LMD-9 wt: genome sequenced strain isolated from dairy products that is part of the ATCC (76).

CB001 - LMD-9 *blp::P_{comS}-luxAB*: strain where part of the bacteriocins *blp* locus is replaced by the *Photorabdus luminescens luxAB* reporter genes under the control of P_{comS} (L. Fontaine, (53)).

LF118 - LMD-9 Δ*comS::P32-cat blp::P_{comS}-luxAB*: strain where *comS* is deleted and replaced by a chloramphenicol resistance cassette, and part of the *blp* locus is replaced by the P_{comS} *luxAB reporter system*. Reference strain for luciferase reporter assays (L. Fontaine, (72)).

LF118 *$comR_{Sth}$::$comR_{Sve}$*: LF118 strain in which the $comR_{Sth}$ gene has been replaced by $ComR_{Sve}$ (this work).

LF118 *$comR_{Sth\text{-}TPRSve}$* : Strain LF118 where $comR_{Sth}$ is replaced by a hybrid *comR* whose HTH domain and linker are those of *S.th*, and whose TPR domain is that of *S.ve* (this work).

LF118 *$comR_{Sth\text{-}linkerSve\text{-}TPRSve}$*: Strain LF118 where $comR_{Sth}$ is replaced by a hybrid *comR* whose HTH domain is that of *S.th.*, and the TPR domain and linker are those of *S.ve*. strain (this work).

LF118 *$comR_{F171Y}$*: Point mutant of LF118 in which $comR_{Sth}$' F171 has been replaced by a Y (this work).

LF118 *$comR_{W197P}$*: Point mutant of LF118 in which $comR_{Sth}$' W197 has been replaced by a P (this work).

LF118 *$comR_{R92G,P94K}$*: Point mutant of LF118 in which $comR_{Sth}$' R92 has been replaced by a G, and P94 by a K (this work).

LF118 *$comR_{V201A,V205A}$*: Point mutant of LF118 in which the V residues in positions 201 and 205 have been replaced by A residues (this work).

LF118 *$comR_{R92G,P94K,W197P,V201A,V205A}$*: Point mutant of LF118 that combines mutations from the three previous strains in $comR_{Sth}$: R92 by a G, P94 by a K, W197 by a P, and V201 and V205 by A residues (this work).

LF118 *comR*$_{\alpha 6,\ \alpha 7\ Sve}$: strain LF118 in which the sequence encoding the first two alpha helices of the TPR domain (TPR-1) of *comR*$_{Sth}$ has been replaced by that of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 7\ Sve}$: strain LF118 in which the sequence encoding the second alpha helix of the *comR*$_{Sth}$' TPR domain has been replaced by that of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 12\ Sve}$*:* strain LF118 in which the sequence encoding the 7th alpha helix of the *comR*$_{Sth}$' TPR domain has been replaced by that of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 7,\ \alpha 12\ Sve}$: strain LF118 in which the sequences encoding the 2nd and 7th alpha helices of the *comR*$_{Sth}$' TPR domain have been replaced by those of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 12,\ \alpha 13\ Sve}$: strain LF118 in which the sequence encoding the 7th and 8th alpha helices of the TPR domain (TPR4) of *comR*$_{Sth}$ has been replaced by that of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 12,\ \alpha 13,\ \alpha 14\ Sve}$: strain LF118 in which the sequence encoding the 7th, 8th and 9th alpha helices of the *comR*$_{Sth}$' TPR domain has been replaced by that of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 6,\ \alpha 7,\ \alpha 12,\ \alpha 13\ Sve}$: strain LF118 in which the sequences encoding alpha helices numbers 1, 2, 7 and 8 of the TPR domain (TPR-1 and TPR-4) of *comR*$_{Sth}$ have been replaced by those of *S. vestibularis* (this work).

LF118 *comR*$_{\alpha 6,\ \alpha 7,\ \alpha 12,\ \alpha 13,\ \alpha 14\ Sve}$: strain LF118 in which the sequences encoding alpha helices numbers 1, 2, 7, 8 and 9 of the *comR*$_{Sth}$' TPR domain have been replaced by those of *S. vestibularis* (this work).

***Streptococcus vestibularis* F0396 *wt*:** wild-type strain. Genome sequence submitted by the J. Craig Venter Institute (77).

***Escherichia coli* TOP10™:** laboratory strain commercialized by Thermo Fisher (78) Ideal for high-efficiency cloning and plasmid propagation. Also used in this project for recombinant protein expression of ComR proteins. Chromosomal genotype (79) : *mcrA, Δ (mrr-hsdRMS-mcrBC), Phi80lacZ (del)M15, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL (SmR), endA1, nupG.* Streptomycin resistant.

2 Plasmids

pBAD-HisA: an expression vector designed for regulated, dose-dependent recombinant protein expression and purification in *E. coli* (80). This 4,102 bp high-copy number plasmid enables L-arabinose-inducible and glucose-repressible protein expression via its *araC* regulon. This plasmid has a 6xHis tag and an ampicillin resistance gene (β-lactamase).

pBAD-*comR*[1]-streptag: expression vector for the ComR proteins studied during this project. This vector is the result of a restriction-ligation between NcoI and EcoRI-restricted pBAD-HisA, and a PCR-amplified *comR* with a C-terminal streptag sequence and floating tails compatible with BspHI and EcoRI restriction sites (this work).

3 Primers

The sequences of the primers used in this project and their purpose are listed in the table below. Primers were ordered lyophilized and resuspended in nuclease-free water to form 100 µM stock solution, which were aliquoted in 50 ml 10 µM working solutions.

Name	Sequence (5'-3')	Use description
LF UpIntComRmut-Fw	TTCTGTTTTAGGAACGATTT TGCTTACAGTTGC	Amplification of the region upstream of the *comRS$_{Sth}$* locus.
LF UpIntComRmut_Rv	ACGATTCTTCATTTCATAAT ATTCGTCAGGAATGG	Amplification of the region upstream of the *comRS$_{Sth}$* locus
LF DNIntComRmut_FW	CAGGAAAATTGGCAGATGGT TTATAGAAATG	Amplification of the region downstream of the *comRS$_{Sth}$* locus
LF DNIntComRmut_RV	AAATCATCAATAATAGCAGT ATTGACCTGACTATTTGC	Amplification of the region downstream of the *comRS$_{Sth}$* locus
LF ChSTER0316A	TAAGAGTGCTATTGGTGTTC TCTTGC	Diagnostic PCR to check the presence of the insert after transformation
LF ChSTER0316B	TCATGGAATTTCACCTCAAT TTCTTGC	Diagnostic PCR to check the presence of the insert after transformation
LL-27 Rv pBAD ComR	CTGTTTTATCAGACCGCTTC	Sequencing primer at the end of comR in the pBAD plasmid
LL-28 Fw pBAD ComR	CCATAAGATTAGCGGATCCTA	Sequencing primer at the beginning of comR in the pBAD plasmid
LL-30 Fw.I.comRSth	TTGAACTTAAAAGACAGCAT TGG	Sequencing primer at the start of *comR$_{Sth}$*
LL-31 Rv.End.comRSth	CTATAAACCATCTGCCAATTTT	Sequencing primer at the end of *comR$_{Sth}$*
LL-63 streptag+STOP+EcoRI	GGTGCTGGTTGGAGTCACCC GCAGTTCGAGAAATAGAATTCG	Forward primer at the end of the streptag sequence in the pBAD-HisA plasmid used for Gibson Assembly cloning
LL-75 Fw.9ah-end.Sth.Ssa	AACAAGAAAGAGGCTGCAG	Amplification of the part downstream the 14th α-helix of the TPR domain of *ComR$_{Sth}$* for overlap PCR
LL-76 Fw.8ah-end.Sth	TATAGCTACAAGCCTAGTGT TTTTG	Amplification of the part downstream the 13th α-helix of the TPR domain of *ComR$_{Sth}$* for overlap PCR
AG1-pBAD-ComRsvest-Fw	GGTCATGAGCATAAAAGACA GCATTGGAC	Tailed primer at the start of *comR$_{Sve}$* including a BspHI restriction site and a ATG start codon for cloning in the pBAD vector
AG2-pBAD-ComRsv-Rv + streptag	GGGAATTCTATTTCTCGAAC TGCGGGTGACTCCAACCAGC	Tailed primer at the end of *comR$_{Sve}$* including an EcoRI restriction site, a

[1] The following *comR* genes were cloned in pBAD vectors: *comR$_{Sth}$*, *comR$_{Sve}$*, *comR$_{F171Y}$*, *comR$_{W197P}$*, *comR$_{R92G,P94K}$*, *comR$_{V201A,V205A}$*, *comR$_{Sth\text{-}TPRSve}$*, *comR$_{Sth\text{-}linkerSve\text{-}TPRSve}$* et *comR$_{α6,\ α7\ Sve}$*.

	ACCCAAACCATCTGCCGCTTTCTC	streptag sequence and a stop codon for cloning in the pBAD vector
AG3-ComRsthF171Y	CTATTATTTCTATCATCTTT ATGGGAGAAAACAG	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "F171Y"punctual mutant in pBAD
AG4-Rv-ComRsth-F171Y	GATGATAGAAATAATAGTCA ATGACTAATAAATC	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "F171Y"punctual mutant in pBAD
AG5-Fw-ComRsth-W197P	GGAAATTCCAACAGATGATG TTTACAATATTGTTT	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "W197P"punctual mutant in pBAD
AG6-ComRsth-W197P	CATCATCTGTTGGAATTTCC TGATTTAATACTC	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "W197P"punctual mutant in pBAD
AG7-ComRsthR92G,P94K.Fw	CAACGTACGGTAACAAAGAC AGAATAAAGTCTAAAC	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "R92G, P94K"punctual mutant in pBAD
AG8-ComRsthR92G,P94K.Rev	CTGTCTTTGTTACCGTACGT TGGAAACTTAATC	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "R92G, P94K"punctual mutant in pBAD
AG9-ComRsthV201A,V201A.Fw	CAGATGATGCTTACAATATT GCTTTATTTAATGATTTG	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "V201A, V205A"punctual mutant in pBAD
AG10-ComRsthV201A,V201A.Rev	CATTAAATAAAGCAATATTG TAAGCATCATCTGTCCAAATTTC	Primer for QuickChange site-directed mutagenesis of *comR$_{Sth}$* "V201A, V205A"punctual mutant in pBAD
JG1-UP.ComRSth+I.ComRSves.Fw	TTGATATAAAGGAGATTCTC TTGAGCATAAAAGACAGCATTG	Forward primer at the start of *comR$_{Sv}$* with a region homologous to *S.th*, to replace *comR$_{Sth}$* by *comR$_{Sv}$* in *S.th*
JG2b-End.ComRSves+DN.ComRSth.Rev	CATTTATGTCACCACCATTT TTACAAACCATCTGCCGC	Reverse primer at the end of *comR$_{Sve}$* with a region homologous to *S.th*, to replace *comR$_{Sth}$* by *comR$_{Sv}$* in *S.th.*
JG3-DN start ComRSth.Fw	AAATGGTGGTGACATAAATG	Forward primer to amplify the region downstream *comR* in *S.th.* for overlap PCR
JG4-UP end ComRSth.Rv	GAGAATCTCCTTTATATCAAAAC	Reverse primer to amplify the region upstream *comR* in *S.th.* for overlap PCR
JG5 RV HTH + Linker S Th	AGGAATGGTAATATTGTCTTG	Reverse primer at the end of the linker region between the HTH and the TPR domain in *comR$_{Sth}$*
JG6-FW end Linker S Th + XIP S Vest	CAAGACAATATTACCATTCC TGACACTTACTATGAGATG	Forward primer including the end of the linker part of *comR$_{Sth}$* and the following sequence in *S.ve.*
JG7b-RV HTH S Th	TCCTAAACGTTTAGCAATATAC	Reverse primer at the end of the *comR$_{Sth}$* HTH domain
JG8-FW end HTH S Th + Linker S Vest	GTATATTGCTAAACGTTTAG GAAAAAGCATGGCTGACTTATTG	Forward primer at the end of the *comR$_{Sth}$* HTH domain and the start of the *comR$_{Sv}$* linker
JG9-Fw.EndCm.Revlox71.EndComRvestCompl	TACATTCCCTTTAGTAACGT GAAACATAAATGTCACTACGCAG	Primer in the *S.ve* genome at the end of *comR* to test its natural competence ability by inserting a chloramphenicol resistance gene
JG10-Rv.InCm.Revlox66.UPComRvestCompl	CCTTATGGGATTTATCTTCC TTATACTTTTTTTATAGTTAATCG	Primer in the *S.ve* genome downstream *comR* to test its natural

		competence ability by inserting a chloramphenicol resistance gene
JG11-Fw.UP.ComRvestCompl	TAGGCTTACATGACCACTAG	Primer in the *S.ve* genome downstream *comR* to test its natural competence ability by inserting a chloramphenicol resistance gene
JG12-RV Sth comR left 2 ahelix	TGGAAACTTAATCAAACGATTC	Reverse primer before the start of the "large" 7th α-helix of the TPR domain of $comR_{Sth}$. Used in overlap PCR
JG13-FW Sth comR left 2 ahelix+Svest start	GAATCGTTTGATTAAGTTTC CAACGTACGGAGATAAGGAGAG	Forward primer to replace the "large" 7th α-helix of $comR_{Sth}$ by the corresponding one of *S.ve*
JG14-RV Svest 2 ahelix end + Sth right	GTTAATAATTCTTCTTCTGG AAGAATATCAAAGAATTGATTATA C	Reverse primer to replace the "large" 7th α-helix of $comR_{Sth}$ by the corresponding one of *S.ve*
JG15-FW.Sth right 2.ahelix	CCAGAAGAAGAATTATTAAC TTTAG	Forward primer after the "large" 7th α-helix of the TPR domain of $comR_{Sth}$. Used in overlap PCR
JG16-RV.Sth ComR left 7.ahelix	AATTTCCTGATTTAATACTCTC	Reverse primer before the start of the "large" 12th α-helix of the TPR domain of $comR_{Sth}$. Used in overlap PCR
JG17-FW Sth left 7.ahelix+Svst start	GAGTATTAAATCAGGAAATT CCAACGGATGATGCTTATAATATT G	Forward primer to replace the "large" 12th α-helix of $ComR_{Sth}$ by the corresponding one of *S.ve*
JG18-RV Svest 7.ahelix end+Sth right	GTTAAGAAGTCTGAGAAGGA CTCAAGAGAAATCTTTAACC	Reverse primer to replace the "large" 12th α-helix of $comR_{Sth}$ by the corresponding one of *S.ve*
JG19-FW Sth right 7.ahelix	TCCTTCTCAGACTTCTTAACAG	Forward primer after the "large" 12th α-helix of the TPR domain of $comR_{Sth}$. Used in overlap PCR
JG20-RV ComRSth 8ax Svest End+Sth right	CAAAAACACTAGGCTTGTAG CTATAAAACTGTGATTTTTC TATGACAGC	Reverse primer to replace the 12th and 13th α-helices of $comR_{Sth}$ by the corresponding ones of *S.ve*
JG21b-RV ComRSth9AxSvtEnd+SthRig	CTGCAGCCTCTTTCTTGTTC CCATTATGTATAAGCTCATACTTA	Reverse primer to replace the 12th, 13th and 14th α-helices of $comR_{Sth}$ by the corresponding ones of *S.ve*

Table 1 Primers used in this project

4 XIP peptides

The sequences of the synthetic XIP peptides used in this project are listed below. Based on published results on XIP_{Sth} (56), octapeptides corresponding to the C-terminal part of ComS *wt* from *S. thermophilus* and *S. vestibularis* were used in this project as an alternative. In addition to the synthetic XIPs of these two strains, three hybrid peptides combining the wild-type XIP sequences of both streptococci were also tested. Ordered as lyophilized powder, aliquots at a concentration of 100 µM were prepared in RNAse-free water. These peptides were used at various concentrations of 50 nM, 100 nM, 1 µM and 5 µM depending on the experimental protocols.

XIP_{Sth} ($ComS_{Sth\ 17\text{-}24}$): LPYFAGCL
MM : 883.08 $g.mol^{-1}$

XIP_{Sve} ($ComS_{Sve\ 15-22}$): VPFFMIYY
MM : 1079.33 g.mol^{-1}

Hybrid peptide n°1 : LPFFAGYY
MM : 997.13 g.mol^{-1}

Hybrid peptide n°2: LPYFAGYY
MM: 993.13 g.mol^{-1}

5 Growth conditions

Escherichia coli

E. coli cells were cultured at 37°C in liquid Lysogeny Broth (LB) under shaking and in aerobic conditions for plasmid cloning and recombinant protein expression. *E. coli* cells were also grown in solid LB medium after electroporation and could be stocked in this way at 4°C for several weeks. For long term conservation, *E. coli* were kept at -80°C in 10% DMSO LB solutions. Streptomycin at 50 µg.ml^{-1} was added to all *E. coli* cultures and ampicillin at 200 µg.ml^{-1} was added to cultures containing pBAD plasmids.

Streptococcus thermophilus et *Streptococcus vestibularis*

Streptococcal strains used in this project were cultured at 37°C in chemically defined medium (CDM) or in M17 medium, containing glucose or lactose at 1% w/w in anaerobic conditions, depending on the experiments. For transformation experiments, 1 ml of CDM glu was inoculated in a 1.5 ml closed Eppendorf tube with the desired strain and grown O/N. After transformation, cells were spread on M17 glu/lac Petri dishes, and incubated in anaerobic conditions using Becton Dickinson™ (BD) Gaspack™ system (81). Isolated streptococcal colonies could be stocked on M17 Petri dishes at 4°C for several weeks. For long term conservation, streptococcal cell cultures were kept at -80°C in 10% v/v DMSO M17 glu solutions. Chloramphenicol at 5 µg.ml^{-1} was added to streptococcal cultures containing the Δ*comS::P32-cat* chloramphenicol resistance marker.

5.1 Lysogeny Broth (LB) medium

Lysogeny Broth, also known as Luria-Bertani medium, is a nutrient-rich growth medium commonly used for bacteria. Its recipe was first published by G. Bertani in 1951 (82) and consist of an aqueous solution of 1% w/v tryptone, 0.5% w/v yeast extract and 0.5% w/v NaCl. In this project agar was further added at 2% w/v to obtain solid medium when required.

5.2 M17 medium

M17 Broth is a commercially available dehydrated medium specifically suited for lactic streptococci (83). Its formulation is as follows: Pancreatic digest of Casein 5 g.L^{-1}, Soy Peptone 5g.L^{-1}, Beef extract 5 g.L^{-1}, Yeast extract 2.5 g.L^{-1}, Ascorbic Acid 0.5 g.L^{-1}, Magnesium Sulfate 0.25 g.L^{-1}, Disodium-beta-glycerophosphate 19 g.L^{-1}. In our experiments, 1% w/v glucose monohydrate or lactose was also added. Agar at 1.5% w/v was added to obtain solid medium when required.

5.3 CDM medium

A chemically defined medium is a synthetic medium in which all the chemicals used and their quantities are known (84). In opposition to undefined media, a chemically defined medium is free of animal or plant-derived components.

In this project, the term CDM refers to a specific medium shown to be experimentally permissive to natural transformation in *S. thermophilus* (64).

The CDM solution is prepared from 5 different solutions.
For 1 liter, mix:

- 860 ml of growth buffer
- 10 ml of vitamin solution 100x
- 10 ml of metal solution 100x
- 20 ml of DNA precursors solution 50x
- 100 ml of amino acid solution 10x

The formulation of each solution is described in the following table. After mixing, the final CDM solution is brought to pH 6.6, sterilized through 0.22 µm filtration and aliquoted in 50 ml falcon tubes wrapped in aluminum foil as a protection against light. CDM aliquots were kept at -20°C and thawed up to maximum two times.

Component	Quantity (g)
Growth buffer (860 ml)	
Glucose monohydrate ($C_6H_{12}O_6$)	10
Sodium acetate (CH_3COONa)	1
Triammonium citrate ($C_6H_{17}N_3O_7$)	0.6
Potassium phosphate (KH_2PO_4)	3
Potassium hydrogen phosphate (K_2HPO_4)	2.5
Urea (CH_4N_2O)	0.24
L-Tyrosine ($C_9H_{11}NO_3$)	0.29
L-ascorbic acid ($C_6H_8O_6$)	0.5
Glutamine ($C_5H_{10}N_2O_3$)	0.39
Glutamate ($C_5H_9NO_4$	0.398
Vitamin solution 100x (1 L)	
Pyridoxamine-HCl	0.5
Nicotinic acid	0.1
Riboflavin	0.1
Calcium pantothenate	0.1
Thiamine-HCl	0.1
Pyridoxine-HCl	0.2
Para-aminobenzoic acid salt	1
Biotin	1
Folic acid	0.1
B12 vitamin	0.1
Orotic acid	0.5
Thymidine	0.5
Inosine	0.5
Lipoic acid	0.25
Metals solution 100x (1L)	
$MgCl_2.(H_2O)_6$	20
$CaCl_2.(H_2O)_2$	5

$FeCl_2.(H_2O)_4$	0.5
$ZnSO_4.(H_2O)_7$	0.5
$CuSO_4.(H_2O)_5$	0.01
$CoCl_2.(H_2O)_6$	0.25
$MnSO_4.H_2O$	2.8
DNA precursors 50x (20 mL)	
Adenine	0.01
Guanine	0.01
Xanthine	0.01
Uracil	0.01
Amino acids solution (1 L)	
L-arginine-HCl	3.5
L-cysteine	2.5
L-histidine	1.5
L-proline	6.75
L-alanine	2.4
L-asparagine	3.5
L-aspartic acid	4.55
L-glycine	1.75
Isoleucine	2.5
L-leucine	4.75
L-lysine-HCl	4.4
L-methionine	1.25
L-phenylalanine	2.75
L-serine	2.25
L-threonine	1.75
L-tryptophan	0.5
L-valine	3.25

Table 2 Composition of the solutions used to prepare 1 liter of CDM glu. Adapted from the master thesis of Dereinne Denis, "Improvement of the ComRS quorum-sensing system for competence and predation in streptococci" (2021) (85).

6 Antibiotics

Antibiotics stock solutions of ampicillin (Amp) 100 $mg.ml^{-1}$ and streptomycin (Strep) 200 $mg.ml^{-1}$were prepared in milli-Q® water (86). Chloramphenicol (Cm) stock solution of 20 $mg.L^{-1}$ was prepared in a mix of 30% ethanol and 70% milli-Q water. All antibiotics were sterilized through 0.22 µm filtration, aliquoted and kept at -20°C. Each aliquot of 500 µl was thawed once and kept after at 4°C for about 2 weeks.

7 Preparation of electrocompetent TOP10™ *E. coli* cells

Day one:

- Inoculate 5 ml LB medium with streptomycin 50 $\mu g.ml^{-1}$ in a 15 ml falcon tube with the TOP10 *E. coli* strain and incubate O/N at 37°C under shaking at 250 rpm.
- Prepare LB medium and a sterile 10% v/v solution. Sterilize 200 ml centrifuge tubes, 250 ml shaking flasks. Store the tubes and flasks at 4°C.

Day two:

- Inoculate 100 ml LB streptomycin 50 µg.ml^{-1} with 1 ml of the TOP10 preculture, in a 250 ml shaking flask.
- Incubate at 37°C under shaking at 250 rpm for 1h30-4h00 until an OD_{600} of 0.4 is reached.
- Strop growth by placing the flask in an ice bath for 50 min. Stir gently from time to time to allow even cooling.
- Centrifuge in a cooled centrifuge at 4000 rpm at 4°C for 15 min and discard the supernatant.
- Resuspend the pellet first with 5 ml of cold sterile water and then continue up to 100 ml.
- Cool for 30 min in an ice bath. Stir gently occasionally.
- Centrifuge in a cooled centrifuge at 4000 rpm at 4°C for 15 min and discard the supernatant.
- Resuspend the pellet first with 5 ml of cold sterile water and then continue up to 50 ml.
- Centrifuge in a cooled centrifuge at 4000 rpm at 4°C for 15 min and discard the supernatant.
- Resuspend the pellet in 5 ml sterile 10% v/v glycerol solution.
- Centrifuge in a cooled centrifuge at 4000 rpm at 4°C for 15 min and discard the supernatant.
- Resuspend the pellet in 2-3 ml sterile 10% v/v glycerol solution.
- Aliquot 100 µl of the resulting culture in 1.5 ml Eppendorf tubes and store at -80°C.

8 Miniprep plasmid isolation and purification

Pre-cultures of pBAD plasmids carrying *E. coli* were started the day before in 5 ml LB streptomycin 50 µg.ml^{-1} Amp 200 µg.ml^{-1} and incubated O/N at 37°C under shaking at 250 rpm, in 15 ml falcon tubes.

Miniprep is a method for plasmid isolation and purification adapted for small volume cell cultures. Plasmid DNA purified by this method can be sent directly to sequencing, used in cloning experiments or transformed by electroporation in electrocompetent cells. The miniprep protocol used in this project is available on the supplier's website, Merck (87).

9 Polymerase chain reaction

The polymerase chain reaction (PCR) is a common molecular biology technique used to exponentially amplify a specific DNA sequence. During this project, PCR played a fundamental role in allowing the amplification of sequences of interest in sufficient quantity to create genetic constructs, conduct mutagenesis and cloning experiments, and submit DNA samples to sequencing.

9.1 Standard PCR

The PCR method used in the laboratory is based on thermal cycling. Such PCR is an enzymatic reaction involving a heat-resistant DNA polymerase, a dsDNA template

containing the sequence to be copied and a pair of ssDNA primers, each complementary to a short sequence at the 3' extremity of the DNA template. All these components are mixed in an appropriate buffer containing available dNTPs.

The polymerization reaction takes place in three steps, each at a specific temperature. The first step consists in the denaturation of dsDNA into ssDNA. The temperature at which dsDNA denatures depends on the length and the sequence of the DNA molecule. This step is carried out here at 98°C to allow total denaturation of the dsDNA.

The second step, annealing, consists in reaching a lower temperature to allow the binding of the primers to the 3' extremities of each strand of the DNA template. This temperature must not be too low because it would favor non-specific binding of the primers. The optimal annealing temperature is specific for each primer. The temperature chosen for this step is therefore often the lowest of the temperatures for each primer or a compromise between the two. The annealing temperature of the primers was calculated via the New England Biolabs™ (NEB) website (88).

The third step of the PCR cycle is the extension step. The temperature of this step depends on the polymerase used. The NEB™ Phusion® High-Fidelity DNA polymerase used in this project, is a 5'-3' polymerase that possess an optimal activity temperature of 72°C. The duration of this step depends on the length of the sequence to be copied. According to the supplier, the Phusion polymerase allows the implementation of 1000 bp in 15-30 seconds, depending on the DNA template nature.

PCR master mix (50 µl)	
Component	Quantity
Phusion HF buffer 5x	10 µl
dNTPs 10 nM solution	1 µl
Forward primer 10µM solution	1 µl
Reverse primer 10µM solution	1 µl
Template DNA	10-250 ng
Phusion DNA polymerase	0.5 µl
Nuclease-free water	Up to 50 µl

Table 3 Standard PCR master mix for Phusion polymerase

The standard PCR protocol includes the preparation of a 50 µl master mix. DNA templates used were either plasmids and the results of minipreps, or genomic DNA, or even other PCR products. In the case of genomic DNA templates, 1 µl of a saturated O/N bacterial pre-culture was used without prior lysis as it was the case for the amplification of streptococcal genomic DNA.

Thermocycling conditions for standard PCR			
Step		Temperature	Time
Initial Denaturation		98°C	30 s
30 Cycles	**Denaturation**	98°C	10 s
	Annealing	Depends on primers sequences	30 s
	Extension	72°C	Depends on copied DNA sequence length
Final extension		72°C	7 min
Hold		4°C	∞

Table 4 Thermocycling conditions used for routine PCRs in this project.

Each PCR experiments included a negative control in which the template DNA was replaced by nuclease-free water, and when possible, a positive control in which the primer pair was replaced by other primers whose effectiveness had already been proven in similar conditions. Once the cycle was completed, PCR products were then stored at 4°C for up to a month or stored at -20°C for long-term conservation.

9.2 Overlap extension PCR

The overlap extension PCR (OE-PCR) is an alternative PCR method used to assemble multiple DNA fragments into a single larger molecule. Each OE-PCR consists of two steps preceded by a preliminary standard PCR.

The preliminary step is a standard PCR amplification of each fragment to be joined. For this step, each primer at the end to be joined possess a 5' overhang whose sequence is complementary to the beginning of the next fragment. Then, during the OE-PCR, all the fragments to be assembled (previously purified) are added to the PCR mix in equimolar concentration, plus a pair of primers complementary to the ends of the final desired molecule. The protocol of an OE-PCR consists of two programs of consecutive PCR cycles. The first program (1st stage) is shorter and has a lower annealing temperature. It allows the different fragments to bind together via their homologous region. They then act as primers for the first extension phase. During the second stage, the final molecule is finally amplified using the external primers.

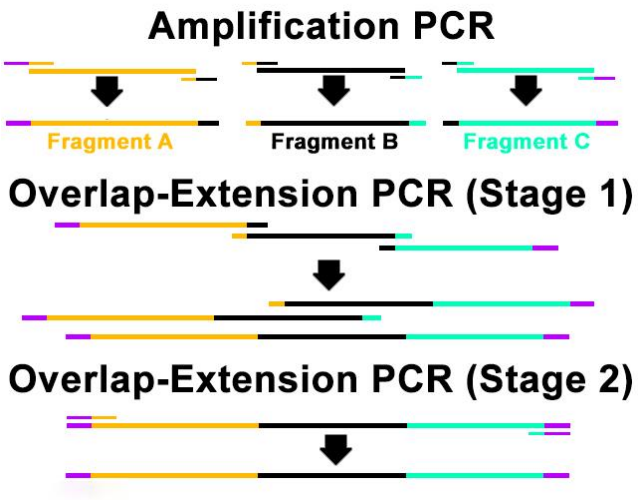


Figure 18 The three steps of overlap extension PCR. The first step consists of standard PCR amplifications of the fragments to be joined. Each pair of primers used in this step possess an homology region in their overhang with the next fragment to be joined. The second step consists in the pair-wise binding of the fragments via their homology region. They then act as primers for a first extension (stage 1). Finally, outer primers permit the amplification of the final desired molecule (stage 2). Adapted from the "Overlap extension polymerase chain reaction" Wikipedia article in English (89).

OE-PCR was a key technique in this project by permitting to flank our genetic constructs with long regions homologous to the genome of *S. thermophilus* and *S. vestibularis*. This step was essential for their insertion in the genome by homologous recombination during natural transformation experiments.

OE-PCR master mix (50 µl)	
Component	Quantity
Phusion HF buffer 5x	10 µl
dNTPs 10 nM solution	1 µl
Forward primer 10µM solution	1 µl
Reverse primer 10µM solution	1 µl
DNA fragments	10-250 ng in equimolar concentrations
Phusion DNA polymerase	0.5 µl
Nuclease-free water	Up to 50 µl

Table 5 OE-PCR master mix for Phusion polymerase

Thermocycling conditions for OE-PCR			
Step		Temperature	Time
Initial Denaturation		98°C	30 s
1st stage 7 Cycles	**Denaturation**	98°C	10 s
	Annealing	Depends on fragments homology regions sequences	30 s
	Extension	72°C	Depends on copied DNA sequence length
2nd stage 25 Cycles	**Denaturation**	98°C	10 s
	Annealing	Depends on outer primers sequences	30 s
	Extension	72°C	Depends on copied DNA sequence length
Final extension		72°C	7 min
Hold		4°C	∞

Table 6 Thermocycling conditions used for OE-PCRs in this project.

9.3 PCR products purification

PCR products were purified from remaining primers, nucleotides and enzymes, by using the QIAGEN™ QIAquick® PCR Purification Kit (90). Purified products were ready to use for further experiments or for sequencing. Purified PCR products were stored at 4°C for up to 2 weeks and stored at -20°C for long term conservation.

10 DNA quantification and purity analysis

Miniprep plasmid DNA and purified PCR products concentration and purity were analyzed by measuring their absorbance at 260 nm and their 260/280 ratio in a Nanodrop™ 2000 small volume spectrophotometer (91).

11 QIAGEN QuickChange™ site-directed mutagenesis

Point mutations of the *comR* gene were performed by site-directed PCR mutagenesis using the pBAD-*comR*$_{Sth}$-*streptag* vector as template. Primer design was based on the article of Zheng et al. (2004) (92), which proposes a protocol allowing to insert up to 17.5% of mutations in partially overlapping primers, complementary to the targeted site.

General primer design rules were the following:

- At least eight non-overlapping bases should be introduced at the 3' primer end.
- Mutations should be included into both primers.
- At least one G or C should be placed at the end of each terminus.

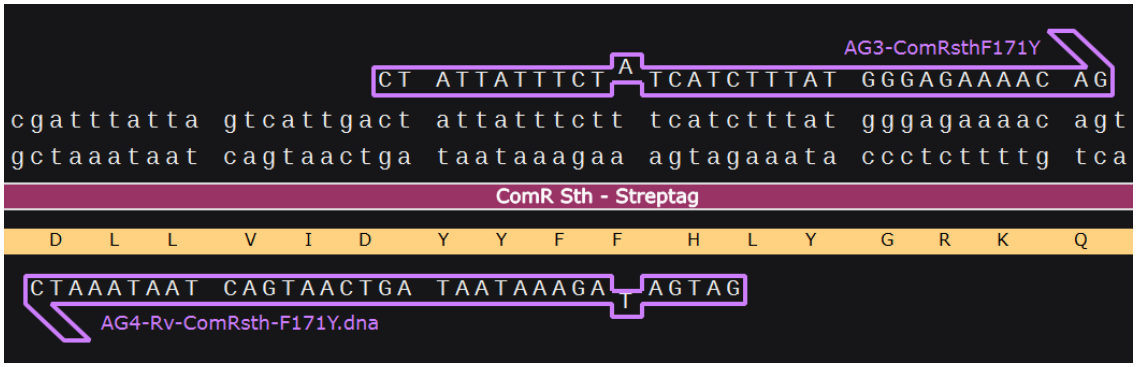


Figure 19 Example of QuickChange™ primer design used for the creation of the *comR*$_{Sth\ F171Y}$ mutant. The sequences framed in purple are those of the primers used. The mutation is present in each primer, at a distance of at least 11 bp of their 3' end. The sequences are viewed using the Snapgene Viewer© program.

12 Gibson assembly cloning

Gibson assembly is a one-step isothermal cloning method used for assembling multiple overlapping DNA molecules by combining the actions of a 5' exonuclease, a DNA polymerase and a DNA ligase (93). The action of the exonuclease will first digest the 5' ends of each DNA molecule, allowing them to bind by their homologous region. The polymerase then replaces the missing bp after annealing and the ligase finally covalently binds the sequences at their ends.

Sequences to be assembled must contain between 20 and 150 bp of homology between them according to the protocol used in the laboratory. DNA molecules must be added at equimolar concentration for an amount ranging from 10-100 ng for each fragment ≤ 6 kb. Gibson assembly cloning was used in this project to create hybrid *comR* combining certain regions of *S. vestibularis* with those of *S. thermophilus*, such as "*comR*$_{\alpha 7,\ \alpha 12\ Sve}$", inserted in the pBAD vector. The *comR* region of interest and the desired plasmid part were amplified by PCR using primers with overhangs containing a sequence homologous to that of the other fragment end. After purification, the PCR products were added to a Gibson master mix and incubated at 50°C for 60 min.

The Gibson assembly master mix is prepared in two steps. First, a buffer solution suitable for enzyme activity and conservation, and containing the required dNTPs is prepared, the "ISO 5X buffer". Then, the enzymes are added to what is referred as the ready-to-use "Gibson master mix". Their composition is listed in the tables below.

ISO Buffer 5x (6 ml)	
Component	Quantity
1 M Tris-HCl pH 7.5	3 ml
2 M $MgCl_2$	150 µl
100 mM dGTP	60 µl
100 mM dATP	60 µl
100 mM dTTP	60 µl
100 mM dCTP	60 µl
1 M DTT	300 µl
PEG-8000	1.5 g
100 mM NAD	300 µl
Nuclease-free water	Up to 6 ml

Table 7 ISO Buffer 5x composition

The ISO buffer 5x was aliquoted in 320 µl volumes and stored at -20°C.

Gibson Assembly master mix (1.2 ml)	
Component	Quantity
ISO buffer 5x	320 µl
10 U/ µl T5 exonuclease	0.64 µl
2 U/µl Phusion polymerase	20 µl
40 U/µl Taq ligase	160 µl
Nuclease-free water	Up to 1.2 ml

Table 8 Gibson assembly master mix composition

The Gibson assembly master mix was aliquoted in 15 µl volumes, stored for at least one year at -20°C, and can withstand 10 thawing cycles without enzyme activity loss.

After assembly, Gibson products were desalinated by dialysis on 13 mm diameter 0.025 µm pore size MF-Millipore™ mixed cellulose esters membrane filters (94), on nuclease-free water during 1 hour at room temperature. During dialysis, 1µl of NEB DpnI (20 U/µl) methylation-sensitive restriction enzyme (95) was also added to discard any remaining wt plasmid DNA template before electroporation in electrocompetent *E. coli* TOP10 cells.

13 Agarose gel electrophoresis

Agarose gel electrophoresis was performed on DNA samples after each PCR experiment and on restricted plasmids to the check the size of the genetic constructs. Agarose gels were made of 0.8% agarose dissolved in water. Before being loaded into the gel, 5 µl of the DNA samples are mixed with 1 µl of loading buffer. This loading buffer contains an orange G dye for visual tracking of DNA migration during the electrophoresis (96). The agarose gel is then immersed in a TAE buffer and a potential difference of 120 to 140 V is then applied to the electrodes on both sides of the gel for a period varying from 20 min to 1h. After the electrophoresis, the gel is immersed in a solution containing ethidium bromide, a fluorescent DNA intercalating agent(97), for 30 min. A UV light source is then used to reveal the DNA bands. By comparing the DNA bands to a reference, their position on the gel is used to estimate the size of the DNA molecules, while their light intensity gives information on the concentration of the loaded samples. The reference used is the molecular weight marker, "SmartLadder 200 bp – 10 kb", from Eurogentec™ (98).

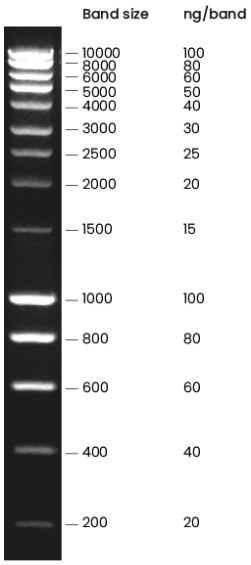


Figure 20 "SmartLadder" molecular weight marker used in agarose gel electrophoresis. DNA band positions and intensities and their corresponding sizes and concentrations are explained on the figure. Adapted from Eurogentec's website (98).

14 Transformation of *S. thermophilus* strains and selection

On the first day, a pre-culture of *S. thermophilus* is started in a closed 1.5 ml Eppendorf tube containing 1 ml of CDM glu and incubated at 37°C. On the second day, this strain is diluted in 1 ml CDM glu to an OD_{600} of 0.05 and incubated at 37°C. After 1h15-1h30 maximum of incubation, 10 µl of 100 µM XIP_{Sth} and 10 µl of transforming DNA at a minimum concentration of 10 $ng.\mu l^{-1}$ are added, then the culture is replaced in incubation at 37°C between 3 and 5 hours. The transforming DNA contains a chloramphenicol resistance cassette enabling transforming colonies to grow on selective medium. A negative control is also performed in which the DNA is replaced by water to ensure the efficiency of the antibiotic selection. When the time is up, 100 µl of the transforming culture and the negative control are plated on M17 glu Cm^5 Petri dishes and placed in anaerobic incubation overnight. A growth positive control is also performed by plating the transforming culture and the negative control on antibiotic-free M17 glu medium to ensure the viability of the culture medium. On the third day, isolated transformant colonies are streaked again on M17 glu Cm^5 and incubated at 37°C overnight under anaerobic conditions. Of the resulting clones, 8 cryotubes of M17 glu Cm^5 DMSO 10% are inoculated with isolated CFU. Diagnostic PCRs are then performed on the 8 candidates and their products are sent for sequencing.

15 SDS-PAGE

Gel electrophoresis separates proteins according to their size by running them through a polyacrylamide gel subjected to an electric field (99). Before being loaded, the samples' concentrations are normalized. The samples are then mixed with crack buffer containing sodium dodecyl sulfate (SDS) and β-mercaptoethanol. SDS is a surfactant that denatures protein structures and allows electrical charges to be distributed evenly over the entire protein, so that the size and not the isoelectric point of the protein is the only factor in the speed of sample migration. β-mercaptoethanol is a reducing agent used to reduce covalent bonds such as disulfide bonds.

Polyacrylamide gels used in this project consist of two layers of different composition and pH. The upper layer, also called the stacking gel, is buffered at pH 6.8 and has a lower percentage of acrylamide. Its role is to line up the proteins before they enter the lower layer (also called the resolving gel) where their separation occurs. The pH of the resolving gel is buffered at 8.8.

Polyacrylamide gels were made in the laboratory with the reagents listed in the following table. A polyacrylamide concentration of 12.5% was chosen to enhance the observation of 30 kDa proteins such as ComR. Ammonium persulfate (APS) is the initiator radical of the acrylamide polymerization reaction and tetramethylethylenediamine (TEMED) is the catalyst. SDS-PAGE were realized using the Bio-Rad Mini-PROTEAN® III protein electrophoresis chamber (100). The molecular weight marker used in SDS-PAGE experiments is the "Page Ruler" from Thermo Fisher (101) which indicates a protein size range from 10 kDa to 180 kDa. Protein electrophoresis was performed at a potential difference of 160 V for 1h-1h30. Protein gel staining was realized using InstantBlue™ Coomassie protein stain (102). This quick protein gel staining technique does not require destaining nor any additional steps. Polyacrylamide gels are submerged for 15-20 minutes in InstantBlue and then briefly rinsed in Milli-Q water.

SDS-PAGE buffers	
Buffer	Composition
Resolving gel buffer 4x	1.5 M Tris buffer pH 8.8 SDS 4 $g.l^{-1}$
Stacking gel buffer 4x	0.5 M Tris buffer pH 6.8 SDS 4 $g.l^{-1}$
Running buffer 10x	0.25 M Tris-Base 1.9 M Glycine SDS 10 $g.l^{-1}$

Table 9 Polyacrylamide gel buffers for SDS-PAGE

SDS-PAGE gel layer		
Layer	Composition	Volume (ml)
Stacking gel (5 ml) **5% polyacrylamide**	Stacking gel buffer 4x Acrylamide:bisacrylamide 40% (37.5 : 1) TEMED APS Milli-Q water	1.25 0.625 0.02 0.02 Up to 5 ml
Resolving gel (10 ml) **12.5% polyacrylamide**	Resolving gel buffer 4x Acrylamide:bisacrylamide 40% (37.5 : 1) TEMED APS Milli-Q water	2.5 3.125 0.02 0.063 Up to 10 ml

Table 10 Polyacrylamide gel layers composition for SDS-PAGE

16 Recombinant ComR protein expression

The expression of recombinant ComR protein is performed in *E. coli* TOP10 cells and is regulated by the *pBAD* promoter of the *ara* operon carried by the pBAD vector (103). The goal of overexpressing *comR* in this system is to produce enough protein to perform *in vitro* assays after purification using the C-ter streptag. The wt *comR* genes of *S. thermophilus* and *S. vestibularis* as well as the full TPR domain exchanges mutants of *comR* were cloned into the pBAD-HisA by restriction ligation. Hybrid TPR *comR* with alpha-

helices exchanges were created from the wt templates by overlap PCR on the entire plasmid and recirculation by Gibson assembly.

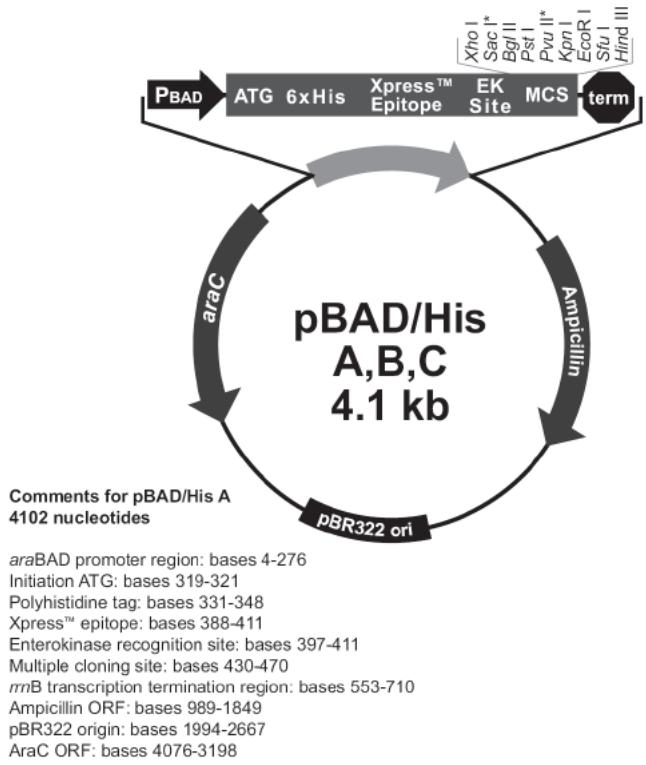


Figure 21 Features summary of the pBAD/His vector. Adapted from the invitrogen™ *pBAD/His A, B and C User Guide* (80).

16.1 Expression induction

The promoter of the *ara* operon of the pBAD vector allows tight regulation of the expression of a gene over multiple ranges. The AraC regulator inhibits the expression of the P_{BAD} promoter. Low levels of P_{BAD} activation can even be further repressed by the addition of glucose. On the contrary, the addition of L-arabinose to the growth medium shifts the action of AraC which becomes an activator allowing an overexpression of the gene of interest (103).

In the laboratory, the following protocol was used to produce purified ComR-streptag proteins.

Day one:

- Inoculate 10 ml LB medium $Strp^{50}$ Amp^{200} in a 50 ml falcon tube with a filter cap, with the desired TOP10 *E. coli* strain and incubate O/N at 37°C under shaking at 250 rpm. A negative control of TOP10 *E. coli* without plasmid is also prepared in the same conditions.

Day two:

- Inoculate 50 ml LB medium $Strp^{50}$ Amp^{200} at OD_{600} 0.05 in a 250 ml sterile shaking flask with the preculture.
- Incubate at 42°C under shaking at 250 rpm for about 1h30 until OD_{600} 0.5.
- Transfer the flask containing the culture in an ice bath for 10 min to stop the growth.

- Sample 2 ml of the culture to represent the conditions before induction (T_0). Measure the OD_{600} absorbance using 1 ml. Centrifuge 1 ml of this sample. Discard the supernatant and store the pellet at -20°C.
- Add sterile L-arabinose diluted previously in a 10% w/v glycerol solution to the medium to a final concentration of 0.05%.
- Incubate at 30°C under shaking for 4 hours.
- Sample 2 ml of the culture to represent the conditions after induction (T_f). Measure the OD_{600} absorbance using 1 ml. Centrifuge 1 ml of this sample. Discard the supernatant and store the pellet at -20°C.
- Centrifuge the culture at 4,000 *g* in a refrigerated centrifuge at 4°C for 10 min. Discard the supernatant and resuspend the pellet in 20 ml of W buffer.
- Centrifuge again at 4,000 *g* in a refrigerated centrifuge at 4°C for 10 min. Discard the supernatant and store the pellet at -80°C.

The T_0 and the T_f fractions are then prepared to check for ComR overexpression on SDS-PAGE.

Overexpression test sample preparation:

- T_0 and T_f pellets are dissolved in Milli-Q water and mixed with 5x crack buffer. The dilutions are normalized based on the previous OD_{600} measurements of the cultures following this formula: for an OD_{600} of 1 A, the pellet is dissolved in 25 µl of Milli-Q water and is then mixed with10 µl of crack buffer 5x.
- The fractions are then heated at 100°C for 10 min.
- Centrifuge the fractions at maximum speed on a benchtop centrifuge during 5 min.
- The supernatants are ready to be loaded on a polyacrylamide gel.

Crack buffer 5x (10 ml)	
Component	Quantity
1M Tris.HCl pH 6.8	0.6 ml
50% Glycerol	5 ml
10% Sodium dodecyl sulfate (SDS)	2 ml
1% Bromophenol Blue	1 ml
0.5M Mercaptoethanol	0.5 ml
Milli-Q water	Up to 10 ml

Table 11 Crack buffer 5x composition

Buffer W 1x (50 ml)	
Component	Concentration
Tris pH 8	100 mM
NaCl	150 mM
EDTA	1 mM
Milli-Q water	Up to 50 ml

Table 12 W buffer 1x composition

16.2 Cell lysis by sonication

If the ComR overexpression test on SDS-PAGE is successful, the cells are then lysed by sonication.

- The pellets stored at -80°C are thawed at room temperature and dissolved in 3 ml of W buffer. Lysozyme is then added at a final concentration of 0.5 $mg.ml^{-1}$.

- Shake gently during 30 min to homogenize the solution then transfer it in a 15 ml falcon tube. Keep the tubes immerged on ice from this moment on.
- Sonicate 2 times the solution during 10 min each time using the following parameters: 6 cycles of 25 s sonication followed by 75 s of rest.
- Centrifuge the tubes at 4,000 *g* at 4°C in a refrigerated centrifuge during 1 hour.
- Recover the supernatant and add roughly the same volume of W buffer to the pellet to extract as much soluble protein as possible without redissolving the pellet.
- Retrieve 50 µl of the supernatant to represent the unfiltered fraction of the soluble protein (NF) and store at -20°C.
- Filter the rest of the supernatant using a 0.45 µm filter. This fraction is now referred to as the filtered fraction (F). Store at -20°C.
- Dissolve the pellets containing the insoluble protein fractions (P) with 3 ml of W buffer. Store at -20°C.

16.3 Solubility test of protein fractions

All fractions that may contain protein, sampled from expression induction to cell lysis, are migrated on SDS-PAGE to analyze whether the ComR protein is soluble or not.

As in the overexpression assay, the non-filtered (NF) and filtered (F) soluble fractions, as well as the insoluble fraction (P) are mixed with 5x Crack Buffer. In this case, 16 µl of these 3 fractions are mixed with 4 µl of 5x Crack Buffer. These samples are then heated for 10 min at 100°C, centrifuged on a benchtop centrifuge at maximum speed during 5 min and then cooled in an ice bath for 10 min. After this preparation, these fractions are ready to be analyzed on SDS-PAGE.

Using a molecular weight scale, fractions from the negative control and optionally a purified ComR protein as a positive control, the band (s) corresponding to ComR synthesis are identified and protein solubility is deduced according to its concentration in the corresponding fraction.

17 Total protein extraction from *S. thermophilus* cultures

In parallel to the production of recombinant ComR proteins in *E. coli* for *in vitro* assays, total protein extractions from *S. thermophilus* cultures were performed to analyze the production of mutant ComR under native conditions by western blot.

Cell lysis was performed using the FastPrep-24™ classic bead beating grinder and lysis system. The total protein extraction protocol is described below.

Buffer A 1x (100 ml)	
Component	Concentration
Tris pH 8	50 mM
NaCl	50 mM
EDTA	1 mM
Milli-Q water	Up to 100 ml

Table 13 Buffer A 1x composition

Day one:

- Inoculate 500 µl CDM glu Cm[5] with the desired *S. thermophilus* strains.
- Prepare 100 ml of buffer A and store it at 4°C.

Day two:

- Inoculate 2 ml CDM glu Cm[5] at OD_{600} 0.05 A for each preculture.
- Incubate at 37°C for 3 hours until OD_{600} reach between 0.3 and 0.5.
- After measuring OD_{600}. normalize samples by taking the volume equivalent to 1 ml at OD 0.5 and transfer it to Eppendorf tubes.
- Centrifuge at 5,000 rpm in a benchtop centrifuge for 3 min. Discard the supernatant and resuspend the pellet with 1 ml of buffer A at 4°C.
- Centrifuge again at 5,000 rpm in a benchtop centrifuge for 3 min. Discard the supernatant. Resuspend the pellet with 100 µl of buffer A at 4°C and transfer it to a screw cap microcentrifuge tube. Add 100 µl of Fast-Prep lysis matrix beads and homogenize each tube.
- Place each tube in the Fast-Prep-24 test tube holder in an evenly spaced pattern.
- Start 4 cycles of 40 seconds each at 4.5 $m.s^{-1}$ speed. Cool the tubes in an ice bath 1 min between each cycle.
- Centrifuge at 13,000 rpm in a benchtop centrifuge for 3 min.
- Transfer 100 µl of the supernatant of each tube in an Eppendorf tube and add 25 µl of crack buffer.
- Heat the tubes at 100°C for 10 min.
- Centrifuge at maximum speed in a benchtop centrifuge for 5 min.
- Cool the samples in an ice bath for 10 min.
- The supernatants are ready to be loaded on a polyacrylamide gel and can be stored at -20°C.

Two SDS-PAGEs are prepared by adding 10 µl of each sample to each well. One will be Coomassie blue stained to control the lysis step and a western blot will be performed on the other.

18 Streptavidin biotin protein purification

The expression vector pBAD-*comR-streptag* was designed with a strep-tag II affinity tag C-terminally fused to *comR*. in order to purify the recombinant ComR-streptag protein by affinity chromatography. Streptag II is an 8-aa peptide sequence (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) with high affinity for streptavidin (104). Strep-tag II offers the advantage of being stable, biologically inert and does not interfere with protein folding or functionality.

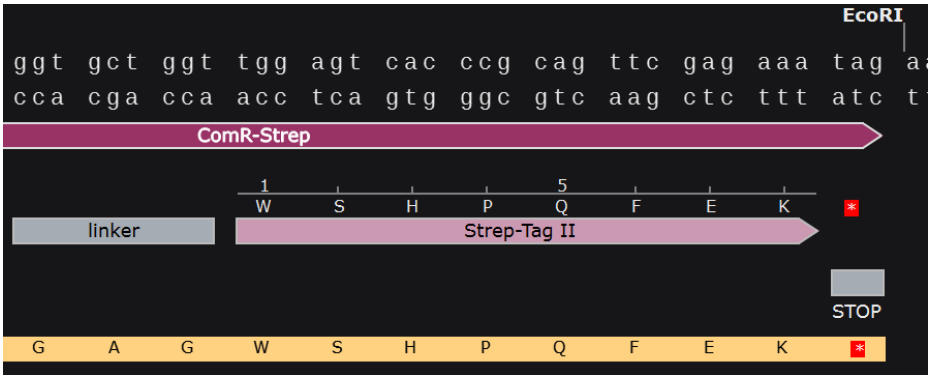


Figure 22 End of the *comR*$_{Sth}$ sequence C-terminally fused with the Strep-Tag II in the pBAD vector. The Strep-Tag II sequence is shown in pink. A 9 bp linker sequence joins the affinity tag to the C-ter end of *comR*$_{Sth}$.

In the laboratory, recombinant ComR proteins were purified by affinity chromatography on a 1 ml Strep-Tactin (an engineered streptavidin derivative) Sepharose column available from IBA™ (105), according to the following protocol:

All operations are carried out in a cold room at 4°C, with gloves on.

- Run 2 x 1 ml buffer W through the column.
- Dispense the filtered lysis supernatant (fraction F from §16.2) onto the column and save the last 100-200 µl.
- Wash the column with 5 x 2 ml buffer W and keep 500 µl of the last wash (fraction W).
- Elute the sample with 6 x 500 µl of buffer W + D-desthiobiotin 2.5 mM (0.0214 mg/40ml) and recover each fraction in a different 1.5 ml microtube containing 50 µl glycerol (fractions E1 to E6).
- Homogenize and store samples at -20°C.

The next day, wash the column for further use.

- Run 2 x 2 ml buffer R.
- Add 0.5 ml of buffer R onto the column and leave it overnight.

Finish cleaning the next day or even after a weekend.

- Run 5 x 2 ml buffer R.
- Then 3 x 2 ml buffer W at pH 10.
- Finally, 2 x 2 ml buffer W at pH 8.
- Apply 1 ml to the column and store it at 4°C.

All samples from ComR production (overexpression and cell lysis) to its purification are then visualized on SDS-PAGE to capture an overall view of the process. The two purest and most concentrated elution fractions are combined and used for further experiments. The concentration of purified ComR-streptag protein is then measured using its absorbance at 280 nm.

19 Western blot

Western blot is an analytical technique used to specifically detect a protein in a sample through the use of antibodies and to estimate its concentration. First, a primary antibody targets the protein of interest, then a secondary antibody binds to the primary antibody in turn. In this project, the secondary antibody is coupled to an horseradish peroxidase enzyme (HRP) which is involved in a chemiluminescence reaction allowing protein visualization.

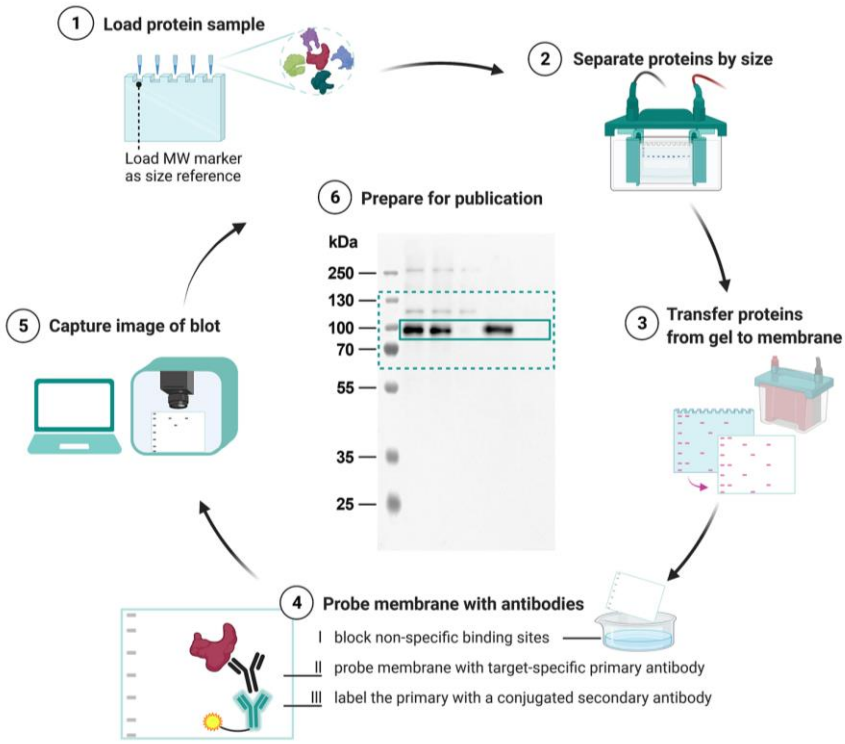


Figure 23 Overview of the steps to identify a protein by western blot. The several steps allowing the identification of a protein by western blot are presented here, from protein separation by SDS-PAGE (1), to the identification of the protein of interest (6). Adapted from Kroon C. et al. (2022) (106).

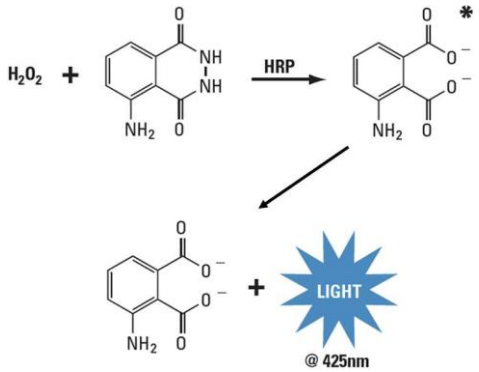


Figure 24 The chemiluminescent reaction behind the revelation of the targeted protein by western blot. Schematic describing the chemiluminescent reaction of oxidation of luminol by hydrogen peroxide, catalyzed by the horseradish peroxidase (HRP) coupled to the secondary antibody used in the western blot. Adapted from the Thermo Fisher Scientific Belgium website (107).

In this work, a primary anti-ComR antibody (108) was used to perform western blots on purified ComR-streptag proteins and on total protein extractions from *S. thermophilus* cultures. In the second case, this technique is also useful to assess the production of ComR mutants under native *in vivo* conditions and their potential toxicity in their host.

19.1 Protein transfer from polyacrylamide gels to nitrocellulose membranes

After protein size separation by SDS-PAGE, it is necessary to transfer them to a membrane before immunodetection. In the laboratory, the transfer of the proteins contained in the polyacrylamide gel to a nitrocellulose membrane is performed by electrophoresis using a Bio-Rad Mini Trans-Blot® electrophoretic transfer cell (109) according to the following protocol.

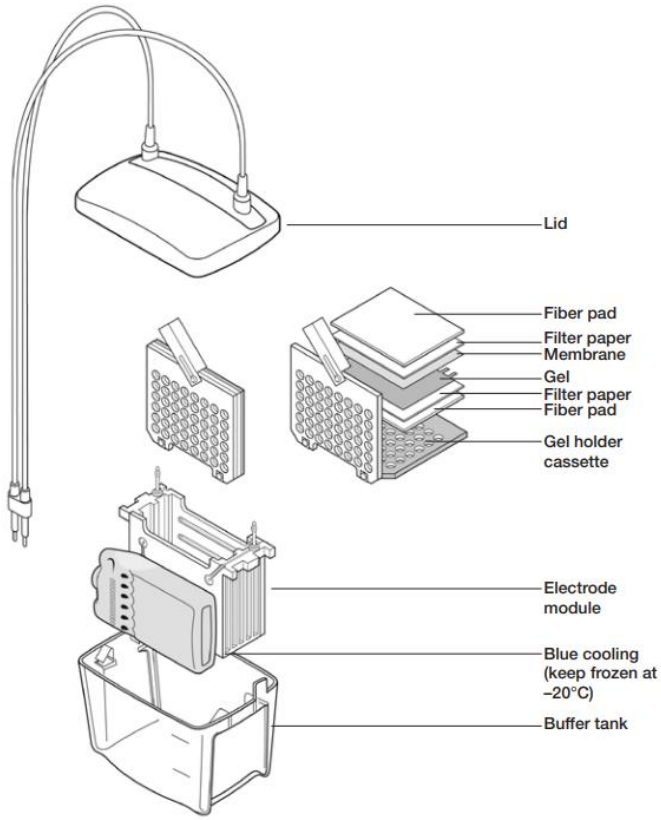


Figure 25 Components of the Mini Trans-Blot electrophoretic transfer cell. Adapted from the Bio-Rad Mini Trans-Blot electrophoretic transfer cell instruction manual(109).

- First, prepare 700 ml of transfer buffer 1x 10% methanol by adding 70 ml TBS 10x and 70 ml of methanol to 560 ml of distilled water.
- Fill the buffer tank with 700 ml 10% methanol transfer buffer.
- A fiber pad, three pieces of Whatman filter paper the size of the gel and the polyacrylamide gel are placed inside the gel holder cassette, in the order described in Fig. 25.
- The nitrocellulose membrane is then soaked in transfer buffer and placed in contact with the polyacrylamide gel. Check the adhesion of the membrane to the gel and remove any air bubble.
- Place three Whatman filter papers and a fiber pad on top and close the gel holder cassette.
- Place the gel holder cassette in the electrode module and submerge the whole assembly in the buffer tank filled with transfer buffer.
- Place a cooling pack at 4°C next to the electrode module and close the lid.
- The transfer is performed in a cold room at 4°C at 90 V during 1 hour.

Western blot buffers	
Buffer	Composition
Transfer buffer 10x	0.25 M Tris buffer 1.9 M Glycine
TBS 10x	NaCl 80 $g.l^{-1}$ 1 M Tris buffer pH 7.8
Milk TBS 1x Tween solution	TBS 1x Milk powder 50 $g.l^{-1}$ Tween 20% 2$ml.l^{-1}$

Table 14 Western blot buffers composition

19.2 Antibody membrane probing and X-ray film development

After transfer, the protein-free sites of the membrane are blocked with a milk solution to prevent non-specific binding of antibodies to the membrane. The membrane is then probed with anti-ComR primary antibody. A solution containing secondary antibodies conjugated to HRP then labels the primary antibodies. The membrane is then covered with a solution containing luminol and hydrogen peroxide, producing a chemiluminescence reaction at the positions occupied by ComR proteins. The blot is eventually developed on x-ray film in a dark room. The film is placed on the membrane and bleached by light emission. Different exposure times are tried to find optimal image contrast and reduced background noise. The film is then developed in a film developer.

The protocol in details is the following:

- Wash the membrane in distilled water and soak it in milk TBS 1x solution during 1 hour in low agitation at room temperature.
- Remove and store the milk solution and add the primary antibody solution.
- Remove and store at -20°C the primary antibody solution. Wash 4 times the membrane with TBS 1x during 3 min.
- Add the secondary antibody in the milk solution used for blocking and add it to the membrane.
- Remove the secondary antibody solution and discard it. Wash 4 times the membrane with TBS 1x during 3 min.
- Gently dry the membrane and place it in the transparent plastic sleeve of a developing cassette.
- In an Eppendorf tube, mix 300 µl of hydrogen peroxide with 300 µl of luminol solution to prepare the revelation solution.
- Spread the revelation solution evenly over the membrane and close the cassette, leaving no air bubbles and wait 3 minutes.
- Place an X-ray film on the plastic sleeve at the position of the membrane and close the cassette. At this point starts the exposure time which can vary to optimize image quality.
- Develop the X-ray film in the film developer, “Optimax X-Ray film processor” (110).

20 Electrophoretic mobility shift assay

The electrophoretic mobility shift assay (EMSA) is an affinity electrophoresis technique used to study protein-DNA interaction. One or more proteins are loaded onto a polyacrylamide gel under non-denaturing conditions with a given labeled DNA sequence (111). If there is binding between the protein (s) and DNA, then the complex formed will migrate more slowly than the free nucleic acid during electrophoresis and its corresponding band will be higher on the gel. On the contrary, if there is no binding, the band will be at the same level as that of the negative control, where only DNA is loaded. EMSA can also be used to estimate the stoichiometry of DNA-protein complex formation.

During this project, EMSAs were performed with the recombinant ComR-streptag produced in *E. coli*, a DNA probe corresponding to the P_{comS} region from *S.th.* coupled to the fluorophore Cy3, and the synthetic XIPs from *S.th.* and *S.ve*. $ComR_{Ssa}$ in the presence of its XIP is used as a positive control, while $ComR_{Sth}$ in the absence of XIP is the negative

control. Cy3 is a fluorophore that absorbs at 550 nm and emits at 570 nm in the green spectrum.

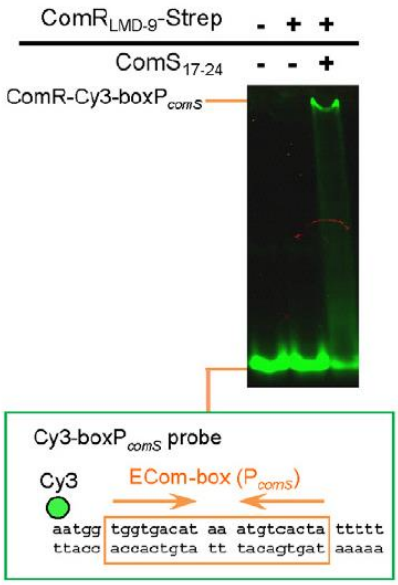


Figure 26 EMSA demonstrating the formation of a ComR$_{Sth}$-XIP$_{Sth}$-DNA complex. Formation of a ComR-XIP-DNA complex between the ComR of *S.th.* LMD-9, its XIP (ComS$_{17-24}$) and the ComR-Cy3-boxP$_{comS}$ DNA probe corresponding to the 30 bp of the P_{comS} promoter region where the ComR-box is located as shown in the green inset. This probe is conjugated to the Cy3 fluorophore. As shown by this *in vitro* experiment, ComR binding to DNA is ComS-dependent. Adapted from Fontaine et al. (2013) (57).

21 Luciferase reporter assay

Luciferase reporter assay is an *in vivo* experiment that indirectly measures the expression level of a gene at transcriptional level. The *luxAB* gene codes for a luciferase, an enzyme that catalyzes a luminous reaction with n-decyl aldehyde as substrate.

In this work, *S. thermophilus* LMD-9 strains carrying mutant *comR* were created with the Δ*comS*::P_{32}*cat blp*::P_{comS} *luxAB* genotype. Part of the bacteriocin *blp* locus (non-essential, late competence genes) is replaced by the reporter gene preceded by the *comS* promoter, P_{comS}. The expression of *luxAB* is thereby regulated by the ComRS system, as P_{comS} has a ComR-box, like P_{comX}, the promoter of *comX*. As explained in §4.1.3, by binding to the ComR-box, the ComR-XIP complex activates the expression of *comX*, which is the central regulator of competence. The expression of *luxAB* thus enables us to indirectly measure the activation of competence in our mutants over time.

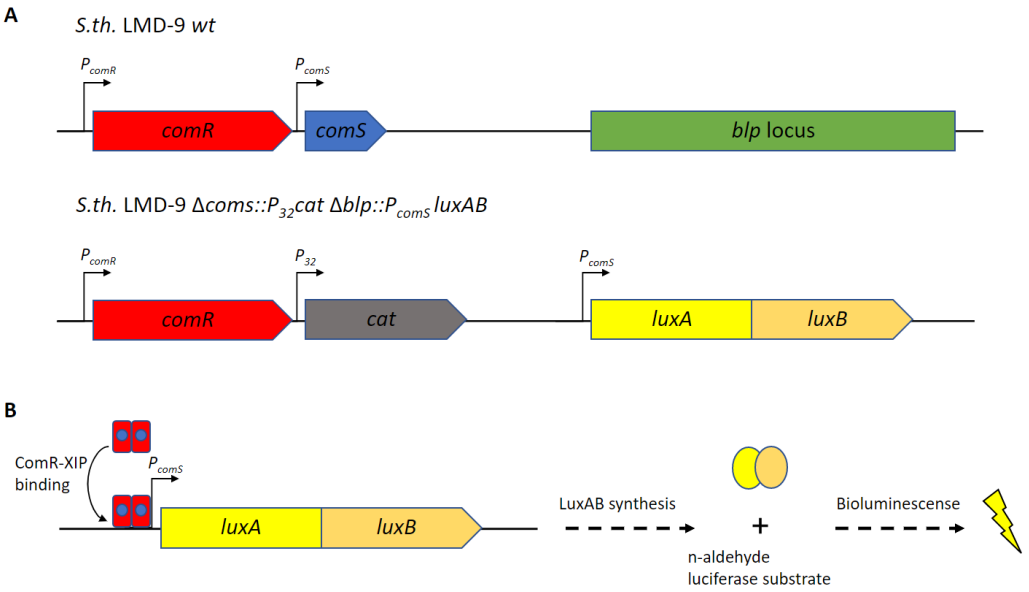


Figure 27 Overview of the reporter lux AB system in *S.th.* strains. (A) Comparison of the genetic organization between *S.th. wt* and the reporter strains *S.th. ΔcomS::P*$_{32}$*cat Δblp::P*$_{comS}$ *luxAB.* (B) Functional overview of the *luxAB* luciferase system in *S.th.* reporter strains.

In our experimental protocol, competence induction is controlled by the addition of synthetic XIP$_{Sth}$ (ComS$_{17-24}$). As *S. thermophilus* strains are Δ*comS*, there is no positive feedback loop. The addition of XIP takes place between 1h15 and 1h30 of incubation at 37°C of an initial dilution at OD$_{600}$ 0.05, as it has been shown that it is at this cell density that *S. thermophilus* LMD-9 is the most permissive to natural competence. Under these conditions, ComR concentration and its interaction with XIP are limiting parameters for competence induction.

The detailed experimental protocol is as follows:

- Pre-cultures of *S.th.* are prepared the previous day in 1 ml CDM glu Cm[5] in 1.5 ml Eppendorf.
- Dilute each strain to OD$_{600}$ of 0.05 in 1.5 ml CDM glu and distribute 300 µl per well for each replicate in a 96-well Greiner CELLSTAR® plate (112) compatible with the Varioskan® Flax spectral scanning multimode reader (113). A blank is made in a well containing 300 µl of CDM, and the empty wells are filled with 300 µl of distilled water.
- Incubate the plate with the lid closed at 37°C for precisely 1h15.
- Add XIP at the desired concentration and distilled water instead for the negative controls.
- Prepare luciferase substrate solution by diluting 125 µl nonyl aldehyde in 10 ml mineral oil. Add 50 µl of nonyl aldehyde solution to each gap between each well on the plate using a multi-channel pipette.
- Insert the plate in the Varioskan. The instrument is set up to measure luminescence and OD$_{600}$ for each well every 15 min for at least 4 hours.

The light emission results obtained are then normalized with respect to the cell densities of the cultures (OD$_{600}$), from which the absorbance of the negative control is deduced using this formula:

$$\text{Specific luminescence} = \frac{luminescence_{\ culture} - luminescence_{\ control}}{OD_{600}\ culture - OD_{\ 600}CDM}$$

Results

1 Experimental strategy

Streptococcus vestibularis is closely-related to *Streptococcus thermophilus*. Both species belong to the salivarius group of streptococci and possess XIPs which feature type I peptides, characterized by the presence of a hydrophobic (V/L)P (F/Y)F motif. Their ComR sequences are highly conserved, with an identity of over 80%. Comparing their domains one by one, the identity of their DNA-binding HTH domain is over 90%, compared with 79% for their pheromone-binding TPR domain. This differential evolution of their domains suggests a co-evolution between the pheromone-binding domains and their XIP pheromones, which is typical of ComRS systems (34). Considering these factors, we hypothesize that the ComRs of these two species share similar activation and selectivity mechanisms.

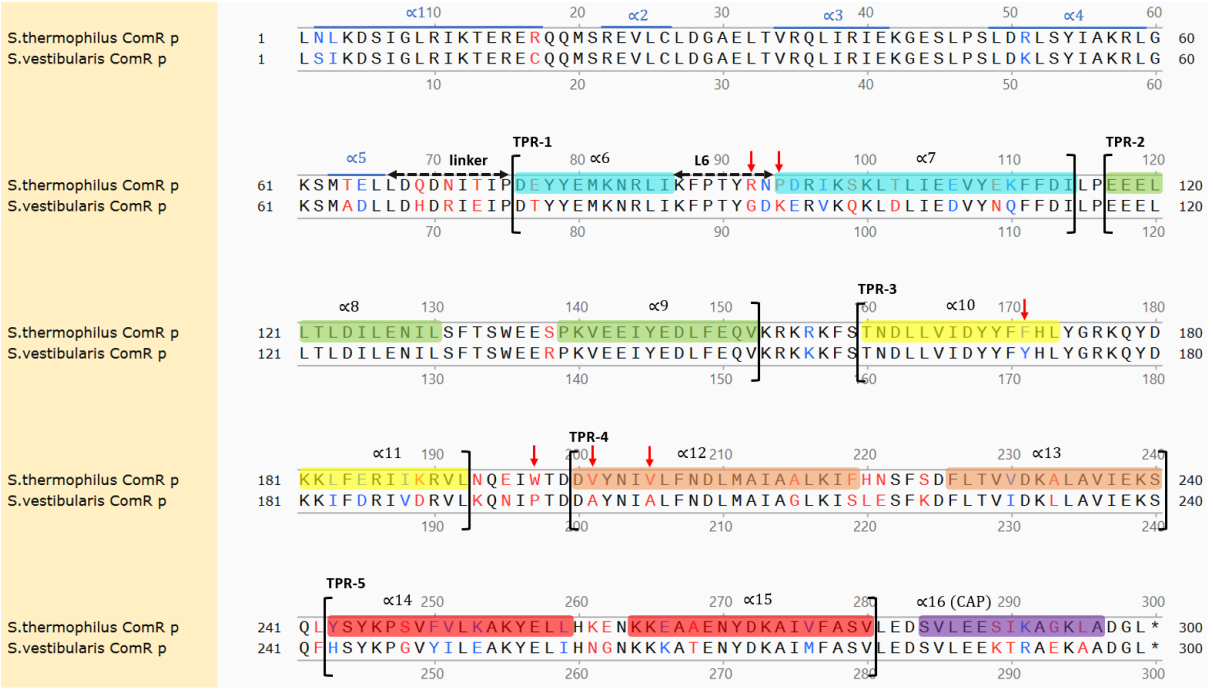


Figure 28 Sequence alignment of ComR from *S. thermophilus* LMD-9, and *S. vestibularis* F0396. Protein sequence alignment made with SnapGene® software, using the BLOSUM62 matrix. Conserved residues are shown in black, similar residues in blue and not similar residues in red. The α-helices are marked in blue for the HTH domain and highlighted in color for the TPR domain. They are colored in pairs, each pair representing a TPR motif, except for the last solitary helix, called CAP, in purple. Each TPR motif is displayed between brackets in black. The linker region and the loop between the 6th and 7th helices (L6) are indicated by dashed double arrows. Mutated residues during this project are marked by red vertical arrows. Adapted from Ledesma-Garcia et al. (2020) (68).

Region	Number of residues	Identity	Similarity
ComR	**299**	**81.2%**	**89.0%**
HTH-domain	**66**	**90.9%**	**97.0%**
Linker	**9**	**66.7%**	**66.7%**
TPR-domain	**224**	**79.0%**	**87.5%**

Table 15 Sequence identity and similarity between $ComR_{Sth}$ and $ComR_{Sve}$ and their domains. Sequence conservation is based on the alignment shown in Fig. 28.

On the basis of the $ComR_{Sth}$ crystal structure in its active form, complexed with DNA, and sequence alignments with $ComR_{Sve}$ and other orthologs (52), potentially important regions in the mechanism of activation and specificity of ComR for XIP have been identified for a mutational analysis.

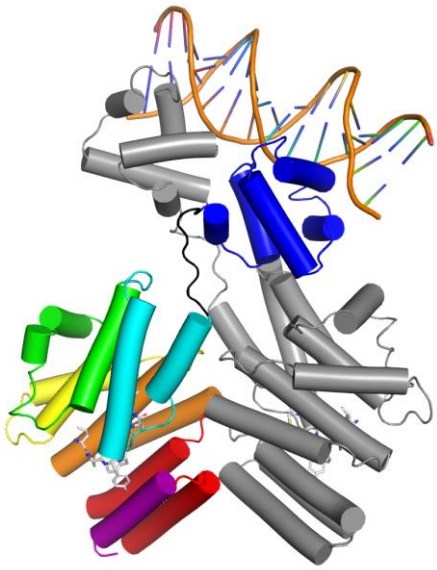


Figure 29 3D structure of ComR from *S. thermophilus* in complex with DNA and its signalling peptide $ComS_{17-24}$. Representation of the crystal structure of the dimeric complex $ComR_{Sth}$·XIP·DNA on PyMol software. The α-helices and the linker region are shown in cartoon, while the two XIP are shown in sticks. DNA is represented in cartoon, in orange. The linker region is in black. Only one ComR monomer is colored while the other appears in dark grey. The helix pairs in each TPR motif are colored in the same way as in the alignment in Fig. 28. PDB ID: 5JUB (114).

Our mutagenesis involves replacing residues, regions or entire domains of $ComR_{Sth}$ with those of $ComR_{Sve}$. These substitutions were made chromosomally in a *S. thermophilus* LMD-9 strain containing a luciferase reporter gene under the control of the ComRS system (see page 58) and some selected constructs were transferred to *E. coli* in a pBAD plasmid for recombinant expression of ComR*-streptag. Our approach is designed to verify the activation and specificity of ComR* by XIP_{Sth} and XIP_{Sve}, and the functionality of the ComR·XIP complex, by combining *in vivo* luciferase assays with *in vitro* EMSA. The luciferase assays were used to observe the gene regulatory function of the ComRS system at transcriptional level in *S. thermophilus*, through the bioluminescence generated by LuxAB production. In addition, these experiments will enable us to monitor the activation kinetics of the ComRS system and the impact of activation of this physiological process on growth via OD_{600} measurements. Moreover, western blot analyses were carried out on mutant strains to ensure ComR* production in *S. thermophilus*. On the other hand, *in vitro* EMSA with purified ComR*-streptag should help us to confirm the specific DNA-binding capacity of ComR·XIP complexes for the ComR-box DNA. These two approaches are therefore complementary. In addition, luciferase assays were also carried out with hybrid XIP peptides in order to further dissect the effects of ComR mutations and their interactions with specific XIP residues. The sequence of these hybrid XIPs is closer to that of XIP_{Sth}, but some residues have been replaced by those of *S. vestibularis*. Finally, considering the knowledge acquired in the meantime, mutations of the most interesting mutants were also combined with the aim of cumulating their effects. Last but not least, a final experiment conducted alongside was to find out if *S. vestibularis* F0396 is competent in CDM conditions by adding XIP_{Sve}.

S.thermophilus	L	P	Y	F	A	G	C	L
S.vestibularis	V	P	F	F	M	I	Y	Y
Hybrid1	L	P	F	F	A	G	Y	Y
Hybrid2	L	P	Y	F	A	G	Y	Y

Figure 30 Sequence alignment of XIP from *S. thermophilus* LMD-9, *S. vestibularis* F0396 and the two hybrid XIP used in this project. Sequence alignment was performed with the Uniprot protein sequence alignment tool (115) which uses the Clustal Omega program. Residues appear in blue according to the level of conservation between each sequence.

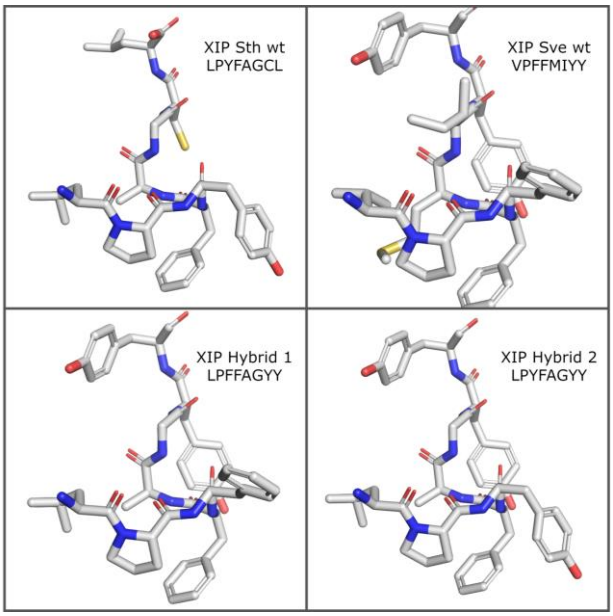


Figure 31 Representation of the structures of the XIPs used in this project. The structure of XIP_{Sth} is derived from the 3D structure of the $ComR_{Sth}$·XIP·DNA complex solved by Talagas et al. (2016) (72). The orientation of each residue in this structure has therefore functional significance, as it is positioned in the pocket of the activated ComR. The structures of the other 3 XIPs are derived from this structure, but from simple residues substitution, without taking into account interactions with ComR. The orientation of each mutated residue is therefore of no experimental significance. This representation is intended to help visualize the effects of these substitutions in terms of bulkiness and new interaction possibilities. Representations made with the PyMol program.

2 $ComR_{Sve}$ exchange and domain swap mutants

First, *$comR_{Sth}$* was completely replaced by *$comR_{Sve}$* in the genome of a *S. thermophilus* strain to create strain LF118 - *$comR_{Sth}$::$comR_{Sve}$*. The aim of this construct is to check whether $ComR_{Sve}$ is functional in *S. thermophilus* and, if so, to study the kinetics of $ComR_{Sve}$ activation in comparison with $ComR_{Sth}$ and the possibility of cross-talk with XIP_{Sth}. At the same time, domain-swap mutants were obtained by replacing the XIP-binding-TPR domain of $ComR_{Sth}$ with that of *S. vestibularis*, while keeping the DNA-binding-HTH domain of *S. thermophilus*. In one strain, the linker region of *S. thermophilus* was preserved (TPR Sve mutant) and in another, it was also substituted by the linker of *S. vestibularis* (TPR + Linker Sve mutant). The aim of the domain-swap mutants is to verify whether the substitutions present in the HTH domain and/or the linker have an important role in the mechanism of recognition and activation by XIP.

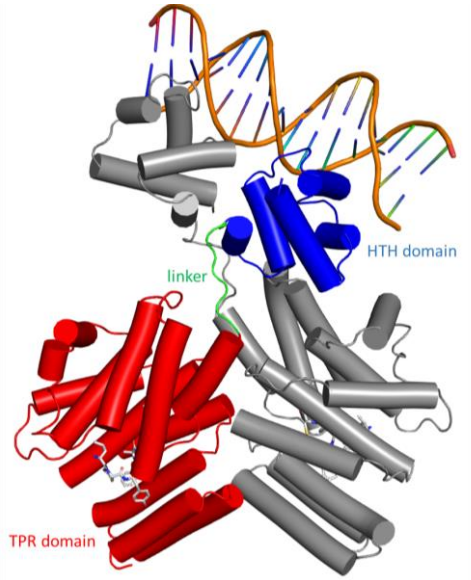


Figure 32 Overview of the HTH and TPR domains of ComR$_{Sth}$ connected by the linker region. This is a variation of Fig. 29 where the HTH DNA-binding domain appears in blue, the TPR pheromone-binding domain in red, and the linker region in green.

Fig. 33 shows the kinetics of specific Lux activity from *S. thermophilus* luciferase reporter strains with ComR$_{Sth}$ and ComR$_{Sve}$, as well as their OD$_{600}$ over time, with and without the addition of XIP$_{Sth}$ and XIP$_{Sve}$. As explained on page 58, the definition of specific luminescence is the following:

$$\text{Specific luminescence} = \frac{luminescence_{\ culture} - luminescence_{\ control}}{OD_{600}\, culture - OD_{\ 600} CDM}$$

It is therefore a measure of the luminescence of a reporter strain caused solely by the addition of XIP to the medium, normalized by the OD$_{600}$ of the culture. In the following column charts, only the maximum luminescence of the different mutants is shown, corresponding to the peak luminescence observed during the luciferase experiments.

The activation profile of the ComRS system of *S. thermophilus* by its cognate XIP is the same as that described in the literature. Competence being an energy-intensive process, there is a delay in growth when a strain becomes competent following the addition of XIP. This delay is visible at the start of the Lux activity peak. At this point, there is a delay in the onset of the exponential phase of growth of the ComR$_{Sth}$ strain with XIP$_{Sth}$ compared with the same strain without XIP. The cell density plateaus reached (stationary phases) are not always at the same level between experiments. They depend, among other things, on how the culture develops in the well and are therefore not really representative of a XIP effect. The same ComRS activation profile is visible for the strain carrying ComR$_{Sve}$ when XIP$_{Sve}$ is added. Peak intensity is slightly higher here, although compilation of our results suggests that ComR$_{Sth}$ is slightly more active with its cognate XIP than ComR$_{Sve}$ with its own XIP in *S. thermophilus*. Furthermore, the same growth delay is observed in the ComR$_{Sve}$ strain when its cognate XIP is added. Competence therefore appears to be functional in *S. thermophilus* in CDM conditions when ComR$_{Sth}$ and XIP$_{Sth}$ are replaced by those of *S. vestibularis*.

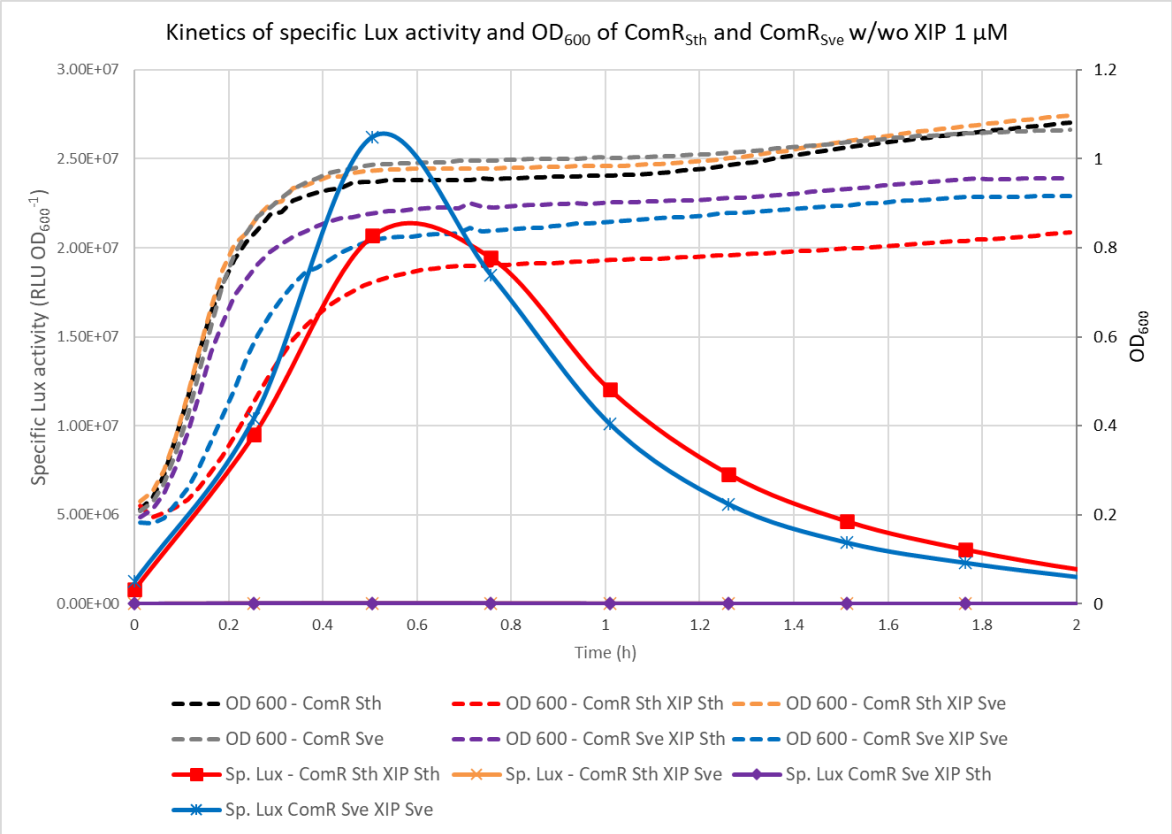


Figure 33 Kinetics of specific Lux activity and OD_{600} of $ComR_{Sth}$ and $ComR_{Sve}$ reporter strains, with and without XIP at 1 µM. These results are from a representative experiment from 9 independent repetitions.

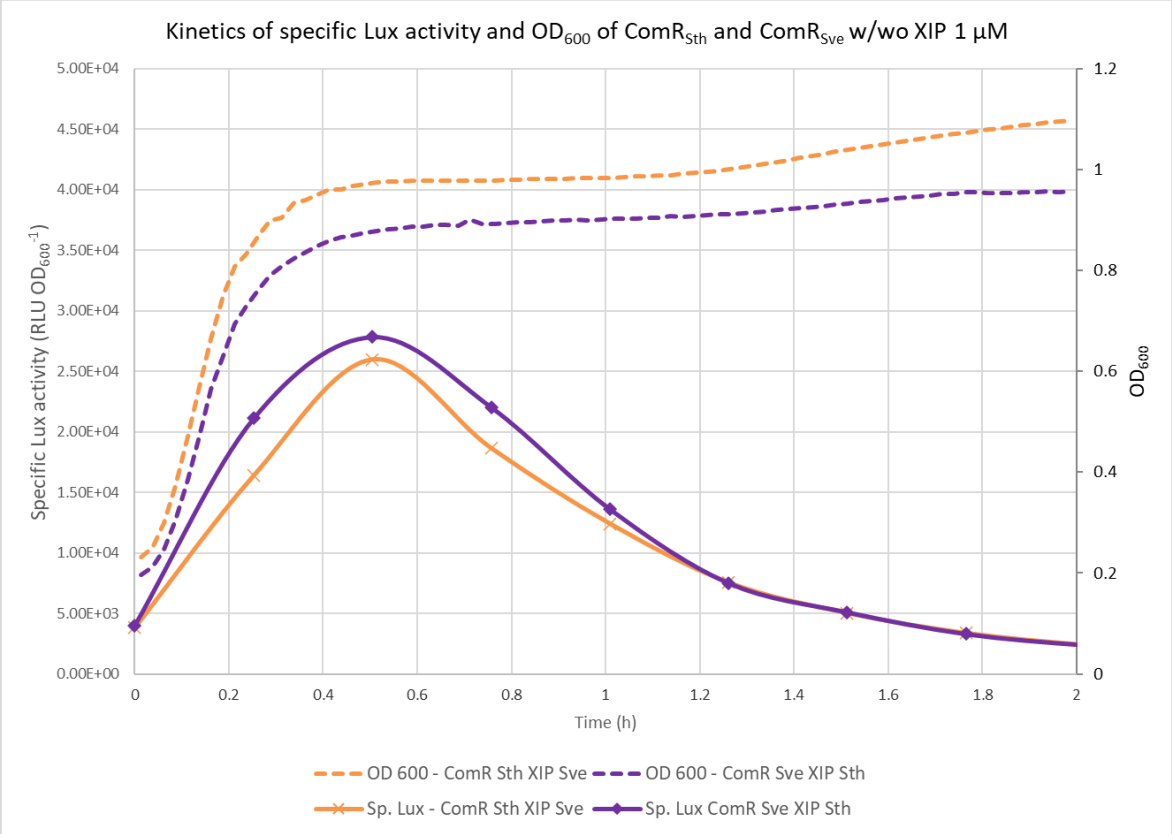
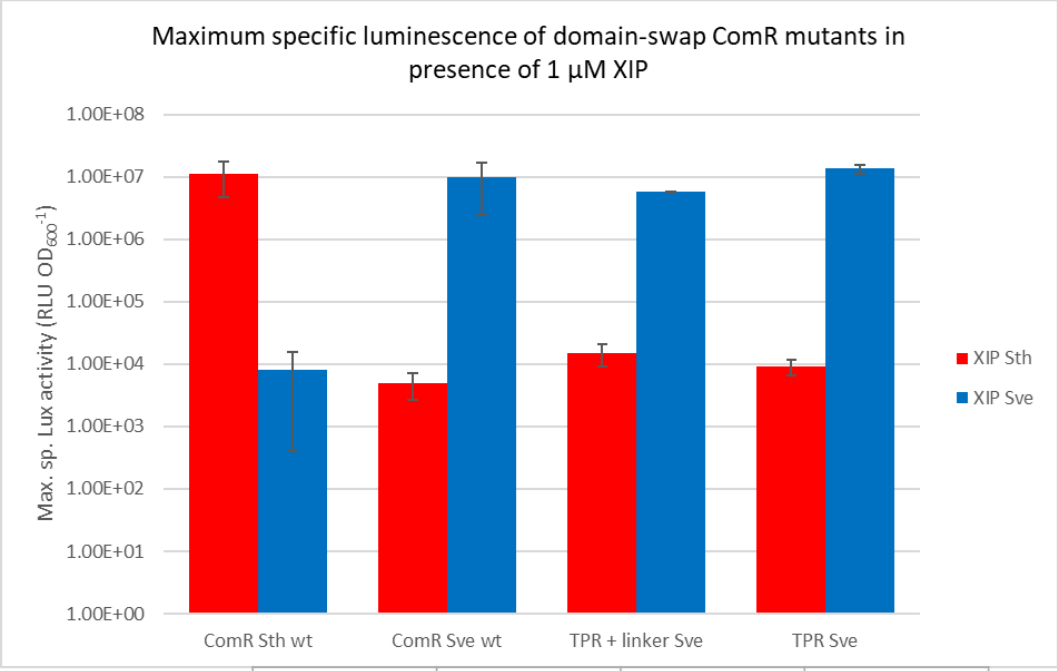


Figure 34 Kinetics of specific Lux activity and OD_{600} of $ComR_{Sth}$ and $ComR_{Sve}$ reporter strains, in response to XIP of the other species. These results are from the same representative experiment as in Fig. 33 based on 9 independent repetitions.

Fig. 34 shows that the peak intensity of light emission is 1,000 times lower when ComRs are treated with XIPs from the other species. It seems that there is still a minimal effect on transcriptional activity, but we cannot consider this to be cross-talk. Moreover, there is no growth delay at the start of the exponential phase, indicating that competence development was not initiated under these conditions.



	ComR Sth wt	ComR Sve wt	TPR + linker Sve	TPR Sve
XIP Sth	1.12E+07	4.98E+03	1.48E+04	9.02E+03
Fold change cf. $ComR_{Sth}$	+1.0	-2252.7	-756.1	-1244.4
Fold change cf. $ComR_{Sve}$	+2252.7	+1.0	+3.0	+1.8
XIP Sve	8.14E+03	9.74E+06	5.68E+06	1.35E+07
Fold change cf. $ComR_{Sth}$	+1.0	+1196.4	+697.6	+1662.5
Fold change cf. $ComR_{Sve}$	-1196.4	+1.0	-1.7	+1.4

Figure 35 Maximum specific luminescence of domain-swap ComR mutants in presence of 1 µM XIP. Each value is the average result of at least 3 different experiments. The standard deviation is represented by capped lines in both directions. Data labels and fold changes (compared to $ComR_{Sth}$ and $ComR_{Sve}$ with each XIP) are presented in the table below.

The average maximum specific Lux activity of domain-swap mutants (see Fig. 35) shows that, as previously stated, there is no cross-activation between the ComRS system in *S. thermophilus* and that in *S. vestibularis*. Furthermore, domain-swap mutants are activated in *S. thermophilus* by their cognate XIP while retaining their specificity, which logically shows that the residues essential for XIP recognition and the initiation of the activation mechanism are located in the TPR domain. Interestingly, the TPR Sve mutant, which retains the linker and HTH domain of Sth, appears to be slightly more active (around 2-fold, statistically significant, *t*-test, $P < 0.05$) with XIP_{Sve} than the TPR + linker Sve mutant (whose Sth linker has been replaced by that of *S. vestibularis*). Nevertheless, this indicates that the linker has a minimal role in ComR activation between $ComR_{Sth}$ and $ComR_{Sve}$.

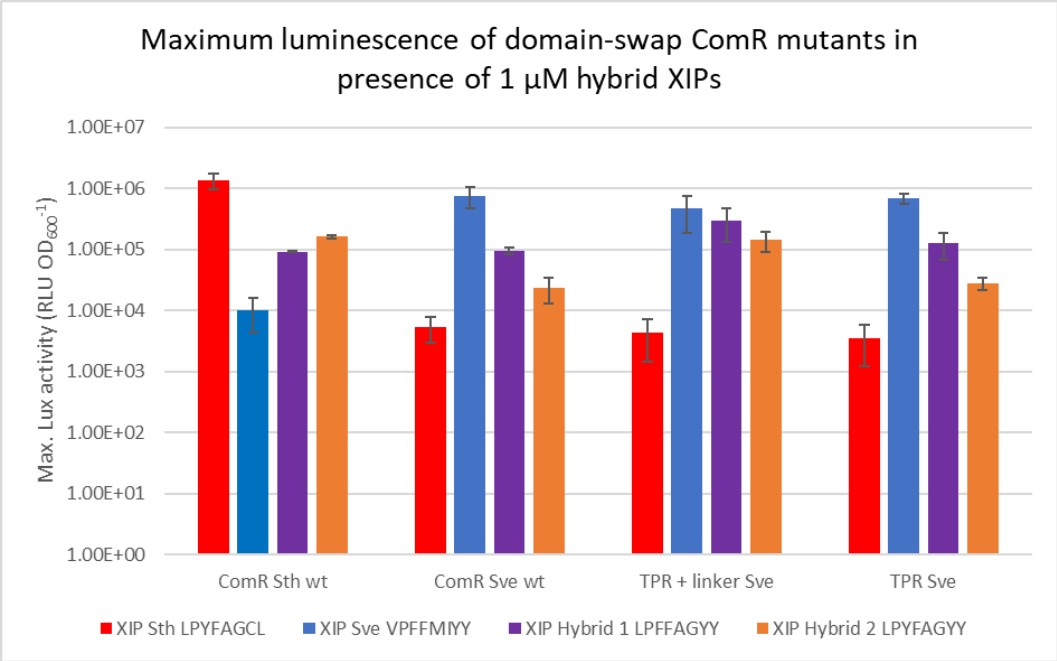


Figure 36 Maximum luminescence of domain-swap ComR mutants in presence of 1 µM hybrid XIPs. Each value displaying error bars is the average result of at least 2 different experiments. Values without error bars come from a single experiment. When present, standard deviation is represented by capped lines in both directions.

The luciferase assays with hybrid XIPs followed a slightly different methodology to that of the luciferase assays presenting results only for XIP_{Sth} and XIP_{Sve}. The Lux activity observed is not measured with the same method. In these experiments, the measured luminescence is no longer specific according to the following formula:

$$\text{Luminescence} = \frac{luminescence_{culture}}{OD_{600}\,culture - OD_{600}CDM}$$

This change in methodology modifies the absolute values of the observed measurements, but there is no risk of constitutive luminescence (and therefore activity) because the luminescence of all these mutants has already been observed without the addition of XIP in other experiments.

For this reason, the results of the two sets of experiments (shown in Fig. 35 and 36) are not mixed, and different values may exist for the same ComR* with the same XIP in the other dataset.

If we look at the results of hybrid peptide luciferase experiments, we see that the $ComR_{Sve}$ and TPR Sve strains behave in the same way with hybrid XIPs, while the TPR + linker Sve mutant appears to be more permissive, which was already the case with XIP_{Sth} in previous data (see Fig. 35). We also note that Y19F, C23Y and L24Y substitutions in XIP_{Sth} (hybrid 1) allow partial activation recovery with $ComR_{Sve}$ and the domain swap mutants. Only few mutations in XIP are needed to allow recognition by the TPR of *S. vestibularis* and to lay the foundations for cross-talk with $ComR_{Sve}$ and domain swap mutants.

3 Point mutants

Four $ComR_{Sth}$ point mutants were designed at first. The first substitution identified was R92G. In Talagas et al 2016 (72), R92 is described as a critical residue in the ComR activation mechanism. This residue which forms hydrophobic interactions with XIP is

part of the loop between α6 and α7 of the TPR-1 motif (called L6). Loop L6 contains the highly conserved P89-T90-Y91 motif, which plays a crucial role in the transmission of the conformational change from TPR-1 to TPR-2, after XIP binding, which results in the break of α9 allowing the release of the HTH domain. The second substitution is P94K. P94 also borders L6, but above all is a radical modification that could lead to a change in the position of α7 and the creation of new interactions with the positively charged lysine. These two substitutions led to the creation of the R92G-P94K double mutant.

F171 belongs to α10 of TPR-3. Still according to Talagas et al 2016 (72), in $ComR_{Sth}$, the conformational change of TPR-2 reorients α10 and F171 would then have a role in the formation of the hydrophobic pocket near the N-ter XIP_{Sth} L17 residue. Its tyrosine substitution in the F171Y mutant would then help to accommodate the XIP_{Sve} N-ter valine. Overall, the role of F171 is not very clear and we hope to gain a better understanding of it in this project.

Another targeted position is the W197P substitution located between α11 and α12. α12 forms the bottom of the peptide-binding pocket and is described as a key factor in TPR selectivity (52,72). By substituting W197P, we hope to modify the position of α12 so that it can fulfill its role by being located at a suitable distance from XIP_{Sve}.

Finally, positions V201 and V205 of α12 will be mutated in the V201A-V205A double mutant, as the substitutions of these two residues close to the C-ter L24 of XIP_{Sth} would help to accommodate the bulkier double C-ter Y residues of XIP_{Sve}. The C-ter part of XIP is indeed seen as the most important part during XIP binding in the TPR pocket, which initiates the ComR activation mechanism.

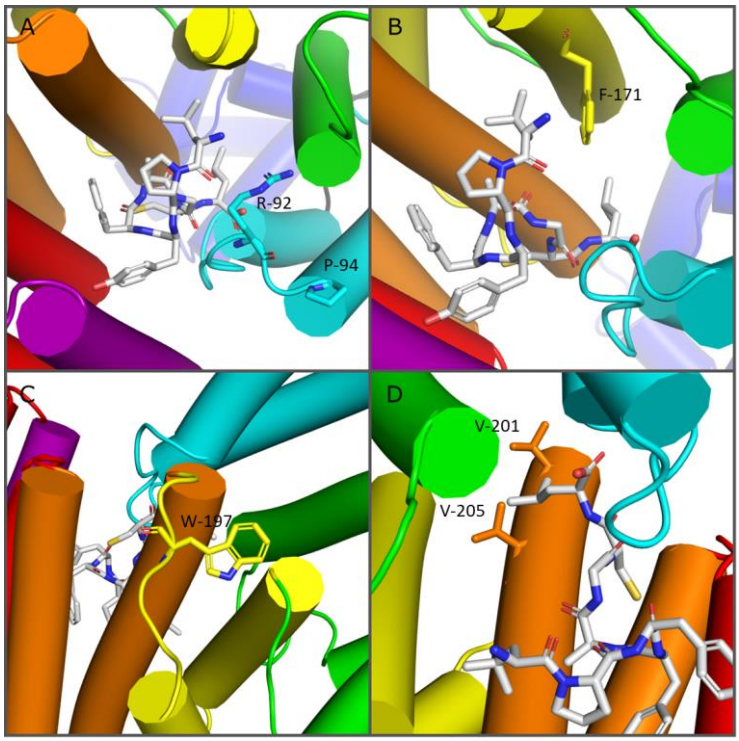
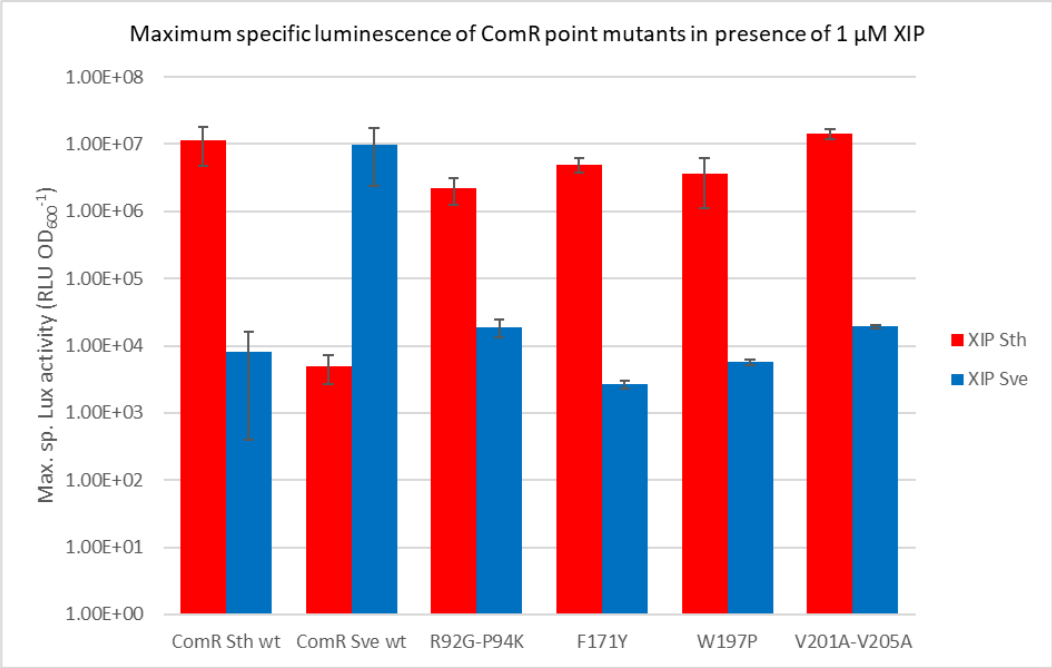


Figure 37 Visualization of $ComR_{Sth}$ residues targeted for mutagenesis in this project. Version of the 3D structure of Fig. 29 zoomed in on the residues targeted for mutagenesis during this project. Each insert shows a residue of $ComR_{Sth}$ targeted for the creation of points mutants. (A) R92G-P94K. (B) F171Y. (C) W197P. (D) V201A-V205A.



	ComR Sth wt	ComR Sve wt	R92G-P94K	F171Y	W197P	V201A-V205A
XIP Sth	1.12E+07	4.98E+03	2.19E+06	4.98E+06	3.61E+06	1.42E+07
Fold change cf. $ComR_{Sth}$	+1.0	-2252.7	-5.1	-2.3	-3.1	+1.3
XIP Sve	8.14E+03	9.74E+06	1.86E+04	2.64E+03	5.71E+03	1.92E+04
Fold change cf. $ComR_{Sth}$	+1.0	+1196.4	+2.3	-3.1	-1.4	+2.4

Figure 38 Maximum specific luminescence of ComR point mutants in presence of 1 µM XIP. Each value is the average result of at least 2 different experiments. The standard deviation is represented by capped lines in both direction. Data labels and fold changes (compared to $ComR_{Sth}$ with each XIP) are presented in the table below.

In Fig. 38, no strong activation or change in selectivity for XIP_{Sve} is observed in the four point mutants. However, the R92G-P94K and V201A-V205A mutants show a slight increase (around 2-fold, non-statistically significant for R92G-P94K and statistically significant for V201A-V205A, *t*-test, $P < 0.05$) in activity compared with $ComR_{Sth}$ wt in the presence of XIP_{Sve}. This could reflect a first step in the recognition of XIP_{Sve}. Interestingly, the V201A-V205A mutant fully maintains its activity in the presence of XIP_{Sth}, while the slight increase in activity in the R92G-P94K mutant for XIP_{Sve} is accompanied by a decrease (around 5-fold, statistically significant, *t*-test, $P < 0.05$) in activity for its cognate peptide. Based on these results, it was decided to combine the point mutations present on helices α7 and α12 to create the quintuple mutant R92G-P94K-W197P-V201A-V205A.

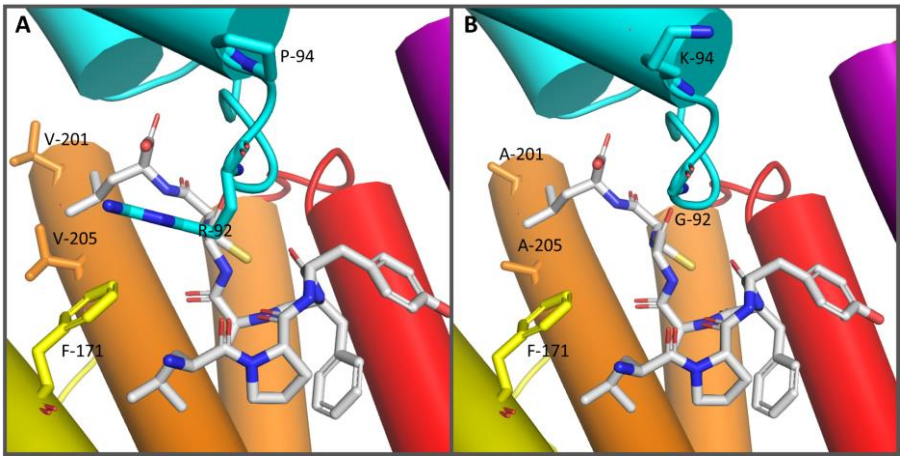


Figure 39 Visualization of mutated residues in the creation of the ComR$_{Sth}$ R92G-P94K-W197P-V201A-V205A point mutant. Version of the 3D structure of Fig. 29 zoomed in on the residues targeted for the creation of the. aforementioned point mutant. The position of W197 does not allow it to be seen at the same time as the other positions. Cited residues are shown in sticks. (A) Residues in ComR$_{Sth}$ wt. (B) Residues after mutation. The orientation of the mutated residues was suggested by the software and does not correspond to experimental data.

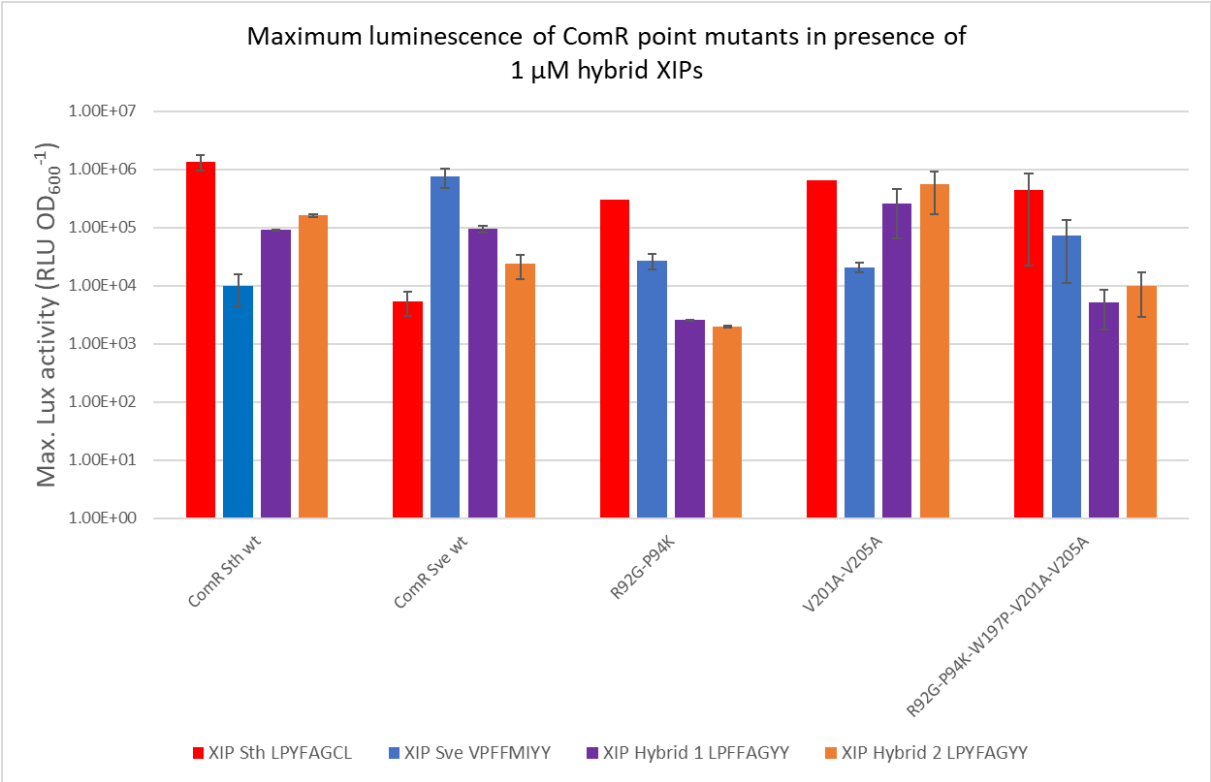


Figure 40 Maximum luminescence of ComR point mutants in presence of 1 µM hybrid XIPs. Each value displaying error bars is the average result of at least 2 different experiments. Values without error bars come from a single experiment. When present, standard deviation is represented by capped lines in both directions.

In the experimental set shown in Fig. 40, the V201A-V205A mutant is strongly activated by hybrid XIPs 1 and 2, which contain two C-ter Y residues instead of C23 and L24 in XIP$_{Sth}$. This result confirms the importance of these positions in their interaction with the C-ter part of XIP and shows that the two alanine residues compensate for the bulkiness of the two tyrosine residues. Finally, the results for the quintuple mutant show high activation in the presence of XIP$_{Sth}$ and moderate activation in the presence of XIP$_{Sve}$ (around 8-fold, statistically significant, *t*-test, $P < 0.05$) compared to ComR$_{Sth}$ wt. The five-mutant would therefore be more permissive while maintaining a preference for XIP$_{Sth}$. Surprisingly, it is not as well activated by XIP hybrids 1 and 2. This latter observation is

contradictory to the activity profiles of double point mutants R92G-P94K and V201A-V205A, and we find it difficult to interpret this result. It would therefore be useful to consolidate such results by repeating experiments with these hybrid XIPs before issuing an initial interpretation.

4 Alpha-helix exchange chimeric mutants

Chimeric $ComR_{Sth}$ mutants in which alpha helices and/or entire TPR motifs were replaced by those of *S. vestibularis* were also tested. The alpha-helices of the TPR-1 and TPR-4 motifs, i.e. alpha-helices α6-α7 and α12-α13 respectively, were targeted first. In a second time, mutants combining certain helix exchanges in both TPRs were additionally assayed. As previously explained, TPR-1 and, in particular, loop L6 play a crucial role in the activation mechanism triggered by peptide binding, while TPR-4 and α12 in particular are thought to play a role in XIP selectivity.

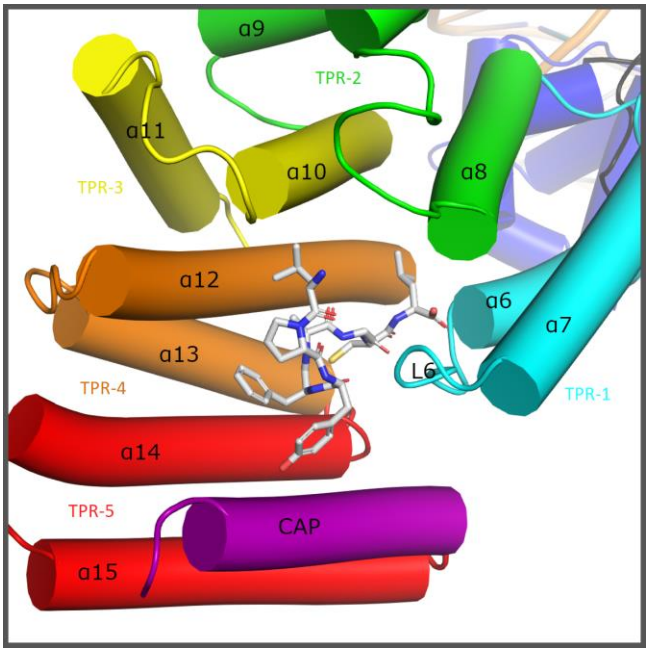
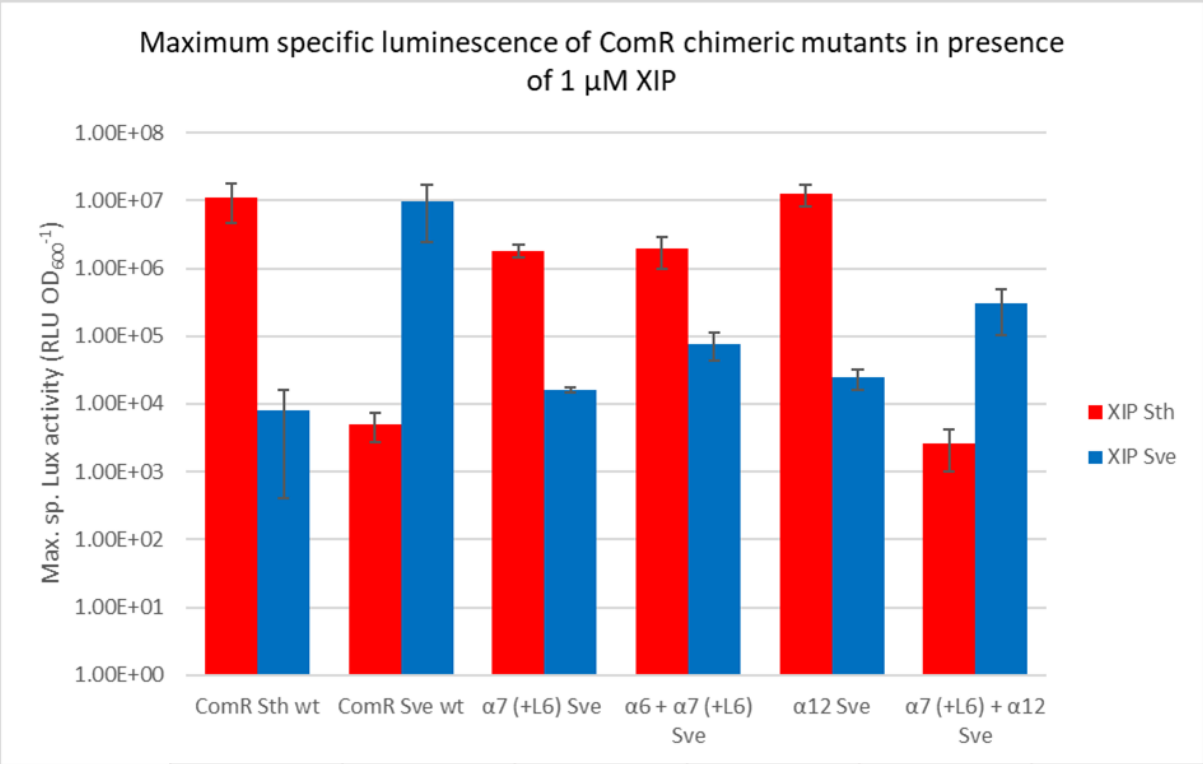


Figure 41 Overview of the α-helices of the TPR domain of $ComR_{Sth}$. Version of the 3D structure of Fig. 29 zoomed in on the TPR domain and centered around XIP to highlight their position relative to the peptide. Each alpha-helix is numbered and the TPR motifs formed by the pairs of helices are written in color.



	ComR Sth wt	ComR Sve wt	α7 (+L6) Sve	α6 + α7 (+L6) Sve	α12 Sve	α7 (+L6) + α12 Sve
XIP Sth	1.12E+07	4.98E+03	1.82E+06	1.92E+06	1.29E+07	2.58E+03
Fold change cf. $ComR_{Sth}$	+1.0	-2252.7	-6.2	-5.8	+1.1	-4342.3
XIP Sve	8.14E+03	9.74E+06	1.60E+04	7.75E+04	2.42E+04	2.98E+05
Fold change cf. $ComR_{Sth}$	+1.0	+1196.4	+2.0	+9.5	+3.0	+36.6

Figure 42 Maximum specific luminescence of ComR chimeric mutants in presence of 1 µM XIP. Each value is the average result of at least 2 different experiments. The standard deviation is represented by capped lines in both directions. Data labels and fold changes (compared to $ComR_{Sth}$ with each XIP) are presented in the table below.

On Fig. 42, α7 and L6 exchange results in the same activation profile as the R92G-P94K double mutant (see Fig. 38 and 40, slight decrease and increase in ComR activation regarding XIP_{Sth} and XIP_{Sve}, respectively), confirming the relative importance of these residues in this part of the TPR. The exchange of the complete TPR-1 motif (α6 + α7 (+L6)) further enhances XIP_{Sve} recognition compared with the exchange of α7 (+L6), while maintaining high activation by XIP_{Sth}. Similarly, the activation profile of α12 confirms the initial effects of the V201A-V205A substitutions (see Fig. 38 and 40, slight increase in ComR activation by XIP_{Sve}). Finally, the combined exchange of helices α7 (+L6) and α12 is the first mutant to show an inversion of selectivity in favor of XIP_{Sve}, to the detriment of its activation by XIP_{Sth}. This chimeric mutant demonstrates that mutations in helices α7 (+L6) and α12 are additive to XIP_{Sve} recognition. Comparison of the activation profile of this mutant with the quintuple point mutant shows that the chimeric mutant is slightly more active in the presence of XIP_{Sve}, while the point mutant retains its activation capacity for XIP_{Sth} (see Fig. 40). One interpretation of this result could be that the swap of α7 (+L6) and α12 adds additional mutations that contribute to the strictness of activation by XIP_{Sve}. It is conceivable that these substitutions could create specific contact points with XIP_{Sve} or, alternatively, modulate the shape of the peptide-binding pocket to correspond more to XIP_{Sve} structure.

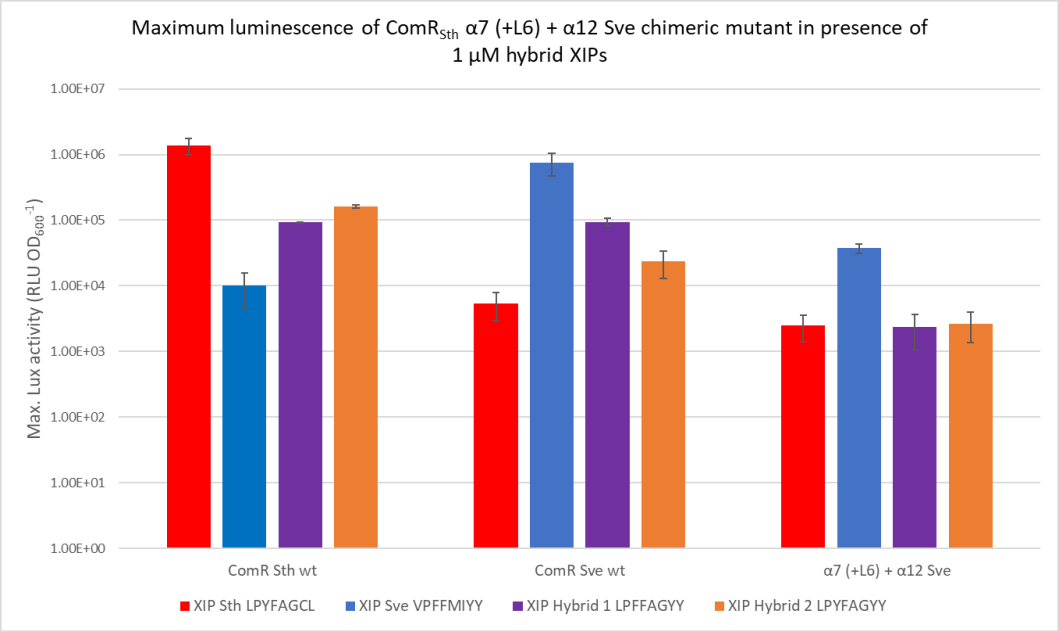


Figure 43 Maximum luminescence of ComR$_{Sth}$ α7 (+L6) + α12 Sve chimeric mutant in presence of 1 µM hybrid XIPs. Each value displaying error bars is the average result of at least 2 different experiments. Standard deviation is represented by capped lines in both directions.

Finally, the luciferase results with hybrid XIPs for α7 (+L6) + α12 Sve (Fig. 43) show specific activation for XIP$_{Sve}$ compared with hybrid XIPs. This result further demonstrates that the residues included in these helices provide the basis for a strict activation profile for the entire XIP$_{Sve}$ sequence.

5 Luciferase assays at low XIP concentration

Complementary to the luciferase assays at 1 µM in XIP, experiments were also performed at a lower concentration of 50 nM (Fig. 44). The aim of these experiments was to observe differences in activation potentially caused by lower affinity of ComR* for XIP, and to study ComR* activation at a concentration closer to values of biological significance.

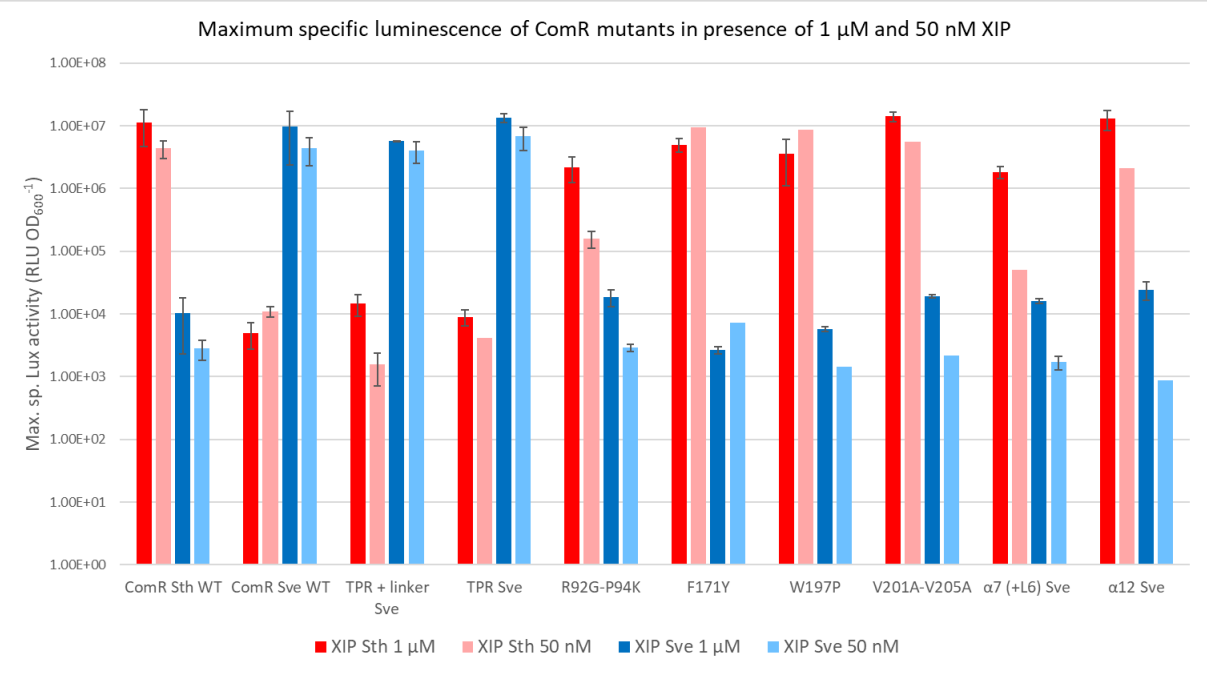


Figure 44 Comparison of maximum specific luminescence of ComR mutants in presence of 1 µM and 50 nM XIP. Each value displaying error bars is the average result of at least 2 different experiments. Values without error bars come from a single experiment. When present, standard deviation is represented by capped lines in both direction.

Two trends can be observed when comparing results obtained at a concentration of 50 nM XIP with those at 1 µM. First, ComR* with a selective activation profile for one XIP or the other do not show significantly different results for the two concentrations. Examples in this category are the two wt ComR for their cognate XIP or the W197P mutant. On the other hand, mutants with a more permissive activation profile show significantly lower activation results at 50 nM concentration. In this other category, we might include the R92G-P94K or the α12 Sve mutants, for example. One possible interpretation of this result is that the more permissive mutants form weaker associations with XIP. Their affinity for XIP would therefore be weaker. As permissiveness increases, affinity for a specific XIP decreases. *In vitro* affinity experiments would confirm this assertion.

6 Western blots

Western blots were performed on *S. thermophilus* LMD-9 luciferase reporter strains to verify and quantify ComR* production.

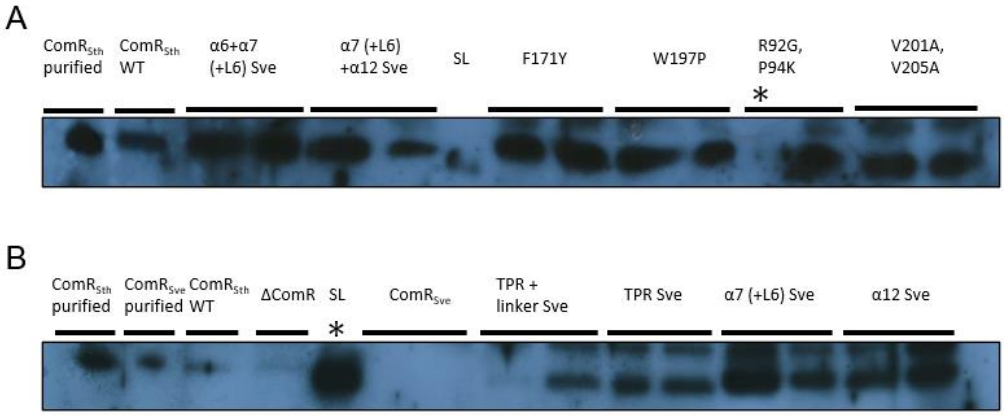


Figure 45 Western blots of mutant ComR *S. thermophilus* LMD-9 reporter strains used for luciferase experiments. Each well contains a purified recombinant ComR-streptag protein or the total protein extract of an *S. thermophilus* LMD-9 strain carrying a mutant ComR, probed with an anti-ComR antibody. Total protein extracts of 2 candidates are shown for each mutant. The name of each ComR mutant is labelled above each well. (SL) refers to a molecular weight marker. ΔComR refers to an isogenic strain in which *comR* has been deleted. A and B correspond to two independent western blots following the same protocol. (A) The asterisk indicates an R92G-P94K candidate that was subsequently excluded from our experiments following this result. (B) The asterisk indicates a problem with the molecular weight marker in this experiment.

ComR* is present in all our mutants, except for the $ComR_{Sve}$ exchange mutant. There is therefore a potential problem with the toxicity or the production of $ComR_{Sve}$ in *S. thermophilus*, or a technical problem for this specific extract in the western blot. This problem remains to be confirmed by repeating the experiment. Furthermore, given the results of the luciferase assays for this mutant, a total absence of ComR seems unlikely, although a lower expression could explain the slightly lower activity of $ComR_{Sve}$ exchange compared with $ComR_{Sth}$ in the presence of their respective XIPs. Except for this discrepancy, the western blot results confirm the reliability of the luciferase data.

7 ComR*-streptag purification

In order to perform *in vitro* assays, in particular EMSA, certain ComR* were C-terminally fused to the streptag II sequence, heterologously expressed in *E. coli* and purified on an affinity column. Overexpression tests, solubility tests and a complete overview of the purification of the TPR + linker Sve mutant are shown in the following figures. In this project, we report the successful purification of soluble and pure $ComR_{Sth}$, $ComR_{Sve}$, TPR + linker Sve and TPR Sve ComR-streptag proteins. However, due to a lack of time, these purified proteins were not tested for their ability to bind DNA in presence of XIP *in vitro*.

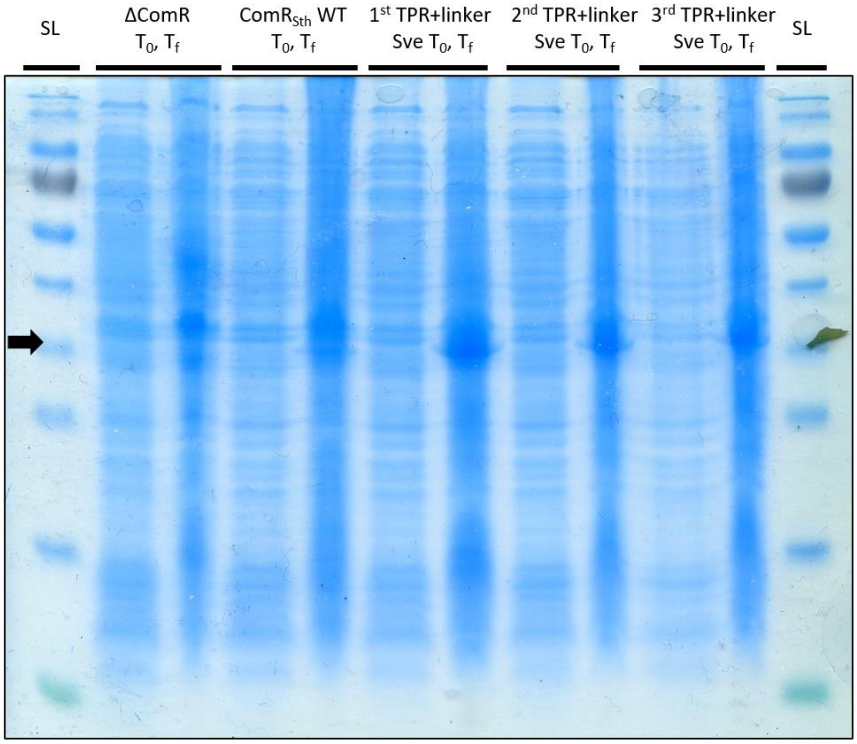


Figure 46 Overexpression test of a recombinant ComR mutant in *E. coli* TOP10 pBAD-*comR*-streptag*. SDS-PAGE gel showing total protein extracts of *E. coli* TOP10 pBAD-*comR*-streptag* lysates before induction of *comR* expression (T_0) and after induction (T_f). Three isogenic candidates each carrying the *$comR_{Sth}$* gene in which the TPR domain and linker region have been replaced by those of *S. vestibularis* are shown. ΔComR refers to an isogenic strain in which *comR-streptag* is not present in the pBAD plasmid. (SL) refers to a molecular weight marker. The black arrow indicates the expected position for the ComR protein.

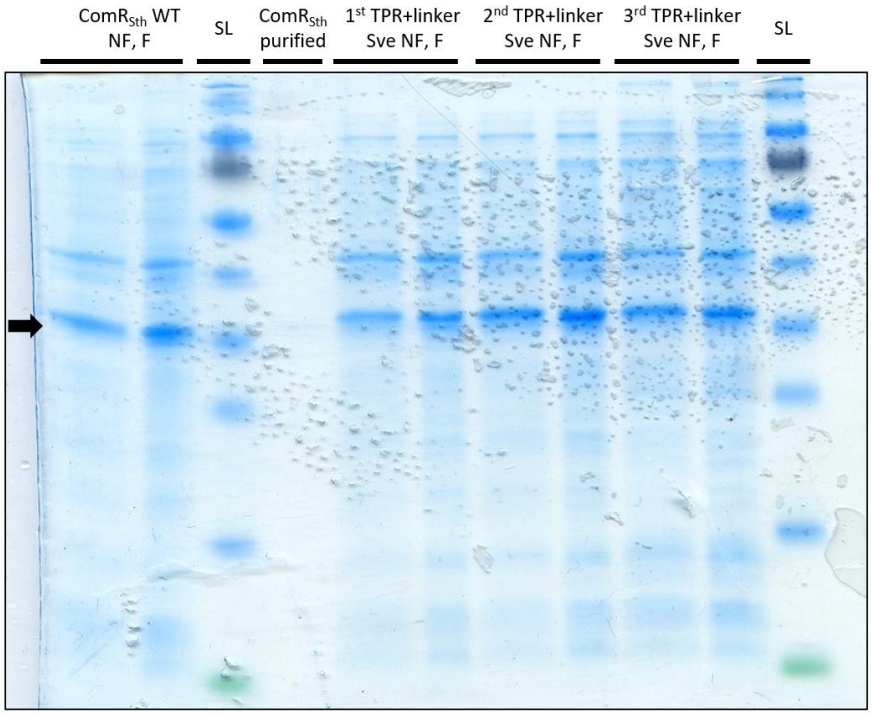


Figure 47 Solubility test of a recombinant ComR mutant in *E. coli* TOP10 pBAD-*comR*-streptag*. SDS-PAGE gel showing soluble fractions of *E. coli* TOP10 pBAD-*comR*-streptag* lysates, after the centrifugation steps, unfiltered (NF), and filtered through a 0.45 µm filter. Three isogenic candidates each carrying the *$comR_{Sth}$* gene in which the TPR domain and linker region have been replaced by those of *S. vestibularis* are shown. (SL) refers to a molecular weight marker. The black arrow indicates the expected position for the ComR protein.

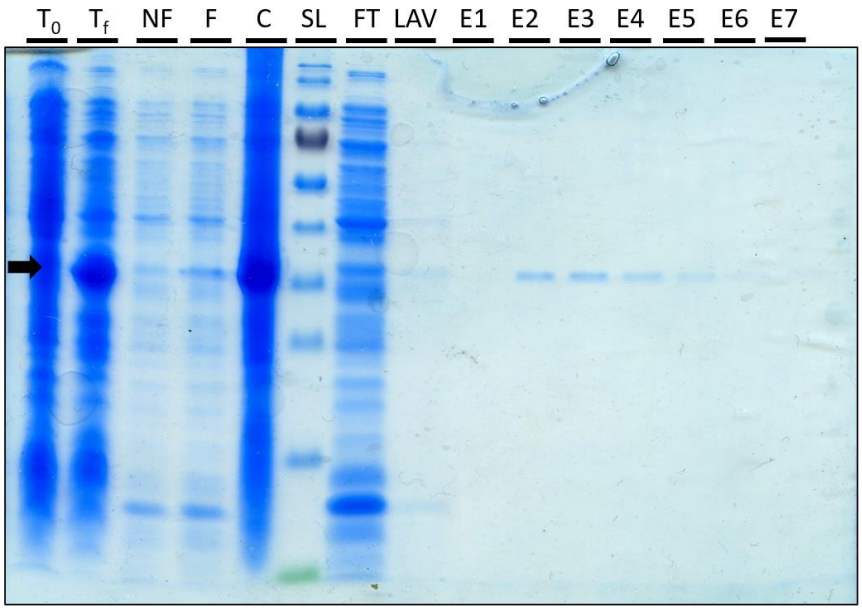


Figure 48 Summary of the purification of a recombinant ComR mutant in *E. coli* TOP10 pBAD-*comR*-streptag*. SDS-PAGE gel showing the different steps and fractions of the purification of the recombinant $ComR_{sth}$ TPR+linker Sve streptagged protein. (T_0) and (T_f) refer to total protein extracts of the culture lysates before and after induction respectively. (NF) refers to the total protein fractions after several centrifugation steps, unfiltered (NF) and filtered (F) through a 0.45 µm filter. (C) refers to the insoluble protein fraction recovered from the pellet. (FT) refers to the fraction recovered during the fixation step. (LAV) refers to the fraction recovered from the washing of the column after elution. (E1-E7) represent the elution fractions. (SL) refers to a molecular weight marker. The black arrow indicates the expected position for the ComR protein.

8 Natural transformation of *S. vestibularis* F0396

The natural transformation of *S. vestibularis* had not yet been described in the literature at the time of this project. It was therefore interesting to see whether competence could be induced in this species under laboratory conditions. The transformability of *S. vestibularis* was thus tested on CDM glu with and without the addition of its cognate XIP_{Sve} at concentrations of 1 µM and 5 µM, following the same protocol as for *S. thermophilus* (cfr. page 48). The experiment consists in adding a DNA fragment containing a chloramphenicol resistance cassette (insertion into the upstream region of $ComR_{Sve}$) to a liquid culture of *S. vestibularis* F0396 wt, and then to spread this culture on selective plates containing chloramphenicol.

XIP_{Sve} concentration	Transformed CFU count per mL
0 µM	0
1 µM	235
5 µM	1695

Table 16 Number of transformed *S. vestibularis* F0396 wt colonies per ml after addition of XIP_{Sve} at concentrations of 1 and 5 µM.

We report in this project that *S. vestibularis* F0396 is transformable in CDM after addition of its cognate XIP_{Sve}, but that its spontaneous transformation (in the absence of external addition of XIP_{Sve}) was not observed under these conditions. As this was a diagnostic one-shot experiment alongside this project, it was not repeated, and the values reported are much lower than those published in Ledesma-Garcia et al (2020) (68). Moreover, it was not possible to calculate transformation rates for this test.

Discussion

During this project, we made some progress in understanding the structural determinants which are dictating the XIP selectivity of ComR activation in species belonging to the salivarius group of streptococci, i.e. *Streptococcus thermophilus* and *Streptococcus vestibularis*. Our first result is that $ComR_{Sve}$ is fully functional in *S. thermophilus* and that it can even replace $ComR_{Sth}$. Indeed, its *in vivo* activation profile in this host is highly similar, as revealed by the kinetics of the luciferase experiments (Fig. 33). We also report concurrently that *S. vestibularis* F0396 is competent under laboratory conditions and that its competence system is functional when XIP_{Sve} is provided in the culture medium (Table 16).

Although highly similar, our results show that the two ComRS type I systems do not cross-talk. Our first aim was to explore the role played by the linker of ComR, the less conserved region between the N-ter HTH domain and the C-ter TPR domain, in activation and specificity (Fig. 35 and Fig. 36). The domain swap mutants highlighted that the linker potentially plays a minor role. Given that the most active mutant in this category is the TPR Sve mutant, which retains the linker Sth and contains the TPR Sve domain, it is possible that the role of the linker is related to interactions with the HTH domain. This hypothesis will need further investigation. It is worth noting that Y19F, C23Y and L24Y substitutions in hybrid XIPs partially restore the activation of $ComR_{Sve}$ and the domain swap mutants (Fig. 36). This shows that only few mutations in XIP are sufficient to enable recognition of a pheromone by the sensing system of another species. This is also an indirect evidence that the ComRS system of one species is capable of rapid adaptation to isolate itself and/or communicate more widely with other species.

Through the study of point and chimeric mutants (Fig. 38, 40, 42 and 43) we confirmed the importance of the R92/G92 substitution, and more globally of the L6 loop and the TPR-1 motif, in the peptide-induced activation mechanism of ComR as described in the article by Talagas et al. (2016) (72). However, the activation results of the R92G-P94K mutant and the corresponding chimeric mutant α7 (+L6) Sve do not allow us to identify the exact role of G92 in the $ComR_{Sve}$ activation mechanism. Experiments with hybrid XIPs demonstrated the importance of substitutions in the V201A-V205A double mutant, which compensate for the presence of the two C-ter tyrosine residues in XIP_{Sve}. On the other hand, this project failed to identify a significant effect of the F171Y and W197P mutations in the mechanism of ComR activation and selectivity. In the article by Talagas et al. (2016) (72), the XIP_{Sth}-activated F171A/Y174A double mutant is no longer able to bind DNA *in vitro*, even though it retains a high affinity for XIP_{Sth}. Given its position in $ComR_{Sth}$, it was proposed that Y171 interacts with the N-ter part of XIP_{Sve}. To test this hypothesis, it would be interesting to carry out further experiments with hybrid XIPs retaining the valine at position 1 of XIP_{Sve} and the F171Y mutant.

Finally, by combining the mutations present in α7 (+L6) and α12, we report the creation of two ComR* which are moderately activated by XIP_{Sve}, i.e. α7 (+L6) + α12 (Fig. 42) and R92G-P94K-W197P-V201A-V205A (Fig. 40). The chimeric mutant displays a complete reversion of selectivity for XIP_{Sve}, while the quintuple point mutant retains the ability to be activated by XIP_{Sth}. These results show that the effects of these mutations are additive,

and demonstrate the importance of these positions in the mechanism of activation and selectivity for the XIP_{Sve} peptide. Interestingly, the chimeric mutant α7 (+L6) + α12 Sve shows a stricter activation profile and is not activated by hybrid XIPs (Fig. 43), unlike the five-point mutant (Fig. 40). The additional mutations present in the exchanged alpha-helices would therefore play a more significant role in the strictness of activation by XIP_{Sve}. However, the activation levels of these mutants remain lower than those of $ComR_{Sth}$ and $ComR_{Sve}$ wild-types, showing that in our case, gains in recognition for one XIP in particular are coupled with a loss of activation for the other.

The results of this master's thesis project contributed to the publication of a PNAS article, "Molecular dissection of pheromone selectivity in the competence signaling system ComRS of streptococci" by Ledesma-Garcia et al. (2020) (68). The results in common follow the same trends, but the values of the published luciferase experiments are much more accurate and allow a clearer and more rigorous assessment of the effects of the different mutations. This difference is mainly due to the method applied to acquire these luciferase results. In the case of the master's thesis project, each ComR* reporter strain occupied a different position in the 96-well plate between each experiment, since new mutants were added to each run and the plate organization was modified accordingly. Perhaps too many ComR* strains were also loaded at the same time, increasing the risk of luminescence cross-contamination between neighboring wells.

Besides these technical details, the PNAS publication presents the resolution of the 3D structure of the $ComR_{Sve}$ monomer in apo form, as well as that of the $ComR_{Sve}$-XIP_{Sve} complex. It also extends the *in vivo* luciferase experiments by pursuing the mutational analysis of different regions of ComR (such as α14 and the CAP helix) and testing the effects of different hybrid XIPs, some of those sequences are closer to XIP_{Sve}. Finally, the paper complements these data with a series of *in vitro* experiments. Namely, measurements of the *in vitro* affinity of ComR* for XIP in fluorescence polarization assays; a study of ComR·XIP complex formation of $ComR_{Sth}$ and $ComR_{Sve}$ for their cognate XIPs using calorimetry experiments, and the verification of the ability of ComR*-XIP complexes to bind DNA in EMSAs.

The results of this paper confirmed the importance of the R92G and V205A substitutions in XIP_{Sve} recognition and identified their precise roles by using hybrid XIPs. R92/G92 interacts with G6/I6 of XIP_{Sth} and XIP_{Sve} respectively (Fig. 49). The role of Y171 in $ComR_{Sve}$ has also been elucidated. Residue Y171 interacts with the methionine at position 5 of XIP_{Sve} (Fig. 49). However, substitutions P94K, W197P and V201A are no longer considered relevant. Finally, this paper presents a quintuple point mutant R92G-V205A-S248G-S289K-I290T that shows complete reversion of selectivity for XIP_{Sve} at a comparable level of activation as $ComR_{Sve}$ wt. This proves that few mutations in ComR are also capable of adapting receptor recognition for different pheromone sequences.

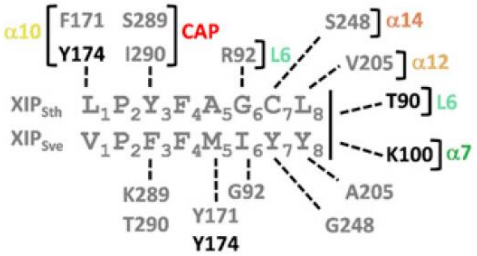


Figure 49 Summary of ComR-XIP interactions. Identical and different ComR residues interacting with peptides are in black and grey, respectively. Adapted from Ledesma-Garcia et al. (2020) (68).

Perspectives

In order to extend our knowledge of the activation and selectivity mechanisms of type II and type III ComRS systems, it would be interesting to apply the same experimental strategy to a species belonging to these categories, in an attempt to create a cross-talk with a type I species. The lower levels of sequence conservation of the ComRs of these species with that of *Streptococcus thermophilus*, compared with *Streptococcus vestibularis*, and the presence of basic and/or acidic residues in most of the types II and III XIPs suggest, however, a higher challenge than the current project.

To understand the ecological significance of the strict vs. permissive characteristics of ComRS systems, competition experiments by predation between species with modified ComRs (e.g. more permissive to pathogen XIPs) could be designed. In parallel, these tests could be compared with competitive transformation experiments taking place in a medium containing sub-lethal doses of antibiotics as well as extracellular DNA containing a resistance gene to compare these two roles associated with competence development.

Finally, the knowledge obtained with this project could prove useful to design new biomedical applications, such as high affinity XIP prototypes capable to block ComR activation (Quorum Quenching), and in turn competence induction and other associated processes such as biofilm formation or virulence increase. Another example of medical application would be the design of *Streptococcus salivarius* (beneficial commensal of the human digestive tract) strains capable of recognizing XIP pheromones from pathogenic streptococci such as *Streptococcus pyogenes* to establish a predatory response.

References

1. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proc Natl Acad Sci U S A [Internet]. 2018 Jun 19 [cited 2022 Jul 31];115(25):6506–11. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1711842115
2. Ochman H, Lawrence[2] JG, Groisman[3] EA. Lateral gene transfer and the nature of bacterial innovation [Internet]. Vol. 405, Nature. 2000. Available from: www.nature.com
3. Sela I, Wolf YI, Koonin E v. Selection and Genome Plasticity as the Key Factors in the Evolution of Bacteria. Phys Rev X. 2019 Aug 5;9(3).
4. Choi IG, Kim SH. Global extent of horizontal gene transfer [Internet]. 2007. Available from: www.pnas.org/cgi/content/full/
5. García-Aljaro C, Ballesté E, Muniesa M. Beyond the canonical strategies of horizontal gene transfer in prokaryotes. Curr Opin Microbiol [Internet]. 2017 Aug 1 [cited 2022 Jul 31];38:95–105. Available from: https://pubmed.ncbi.nlm.nih.gov/28600959/
6. Domingues S, Nielsen KM. Membrane vesicles and horizontal gene transfer in prokaryotes. Curr Opin Microbiol [Internet]. 2017 Aug 1 [cited 2022 Jul 31];38:16–21. Available from: https://pubmed.ncbi.nlm.nih.gov/28441577/
7. Dubnau D, Blokesch M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. Vol. 53, Annual Review of Genetics. Annual Reviews Inc.; 2019. p. 217–37.
8. Johnsborg O, Eldholm V, Håvarstein LS. Natural genetic transformation: prevalence, mechanisms and function. Res Microbiol. 2007 Dec;158(10):767–78.
9. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Vol. 3, Nature Reviews Microbiology. 2005. p. 711–21.
10. Jain R, Rivera MC, Moore JE, Lake JA. Horizontal gene transfer accelerates genome innovation and evolution. Mol Biol Evol. 2003;20(10):1598–602.
11. Jeltsch A. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene [Internet]. 2003 Oct 23 [cited 2022 Jul 31];317(1–2):13–6. Available from: https://pubmed.ncbi.nlm.nih.gov/14604787/
12. Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: Distribution, shared mechanisms and divergent control. Vol. 12, Nature Reviews Microbiology. 2014. p. 181–96.
13. Griffith F. The Significance of Pneumococcal Types. J Hyg (Lond) [Internet]. 1928 [cited 2022 Jul 31];27(2):113–59. Available from: https://pubmed.ncbi.nlm.nih.gov/20474956/
14. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med [Internet]. 1944 Feb 1 [cited 2022 Jul 31];79(2):137–58. Available from: https://pubmed.ncbi.nlm.nih.gov/19871359/
15. Blokesch M. Natural competence for transformation. Vol. 26, Current Biology. Cell Press; 2016. p. R1126–30.

16. Rutherford ST, Bassler BL. Bacterial quorum sensing: Its role in virulence and possibilities for its control. Vol. 2, Cold Spring Harbor Perspectives in Medicine. Cold Spring Harbor Laboratory Press; 2012.
17. Waters CM, Bassler BL. Quorum sensing: Cell-to-cell communication in bacteria. Vol. 21, Annual Review of Cell and Developmental Biology. 2005. p. 319–46.
18. Federle MJ, Bassler BL. Interspecies communication in bacteria. Vol. 112, Journal of Clinical Investigation. The American Society for Clinical Investigation; 2003. p. 1291–9.
19. West SA, Griffin AS, Gardner A, Diggle SP. Social evolution theory for microorganisms. Nature Reviews Microbiology 2006 4:8 [Internet]. 2006 Aug [cited 2022 Jul 31];4(8):597–607. Available from: https://www.nature.com/articles/nrmicro1461
20. Diggle SP, Gardner A, West SA, Griffin AS. Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? Philosophical Transactions of the Royal Society B: Biological Sciences [Internet]. 2007 Jul 7 [cited 2022 Jul 31];362(1483):1241. Available from: /pmc/articles/PMC2435587/
21. Baltrus DA, Guillemin K. Multiple phases of competence occur during the Helicobacter pylori growth cycle. FEMS Microbiol Lett [Internet]. 2006 Feb [cited 2022 Jul 31];255(1):148–55. Available from: https://pubmed.ncbi.nlm.nih.gov/16436074/
22. Mirouze N, Desai Y, Raj A, Dubnau D. Spo0A∼P Imposes a Temporal Gate for the Bimodal Expression of Competence in Bacillus subtilis. PLoS Genet [Internet]. 2012 Mar [cited 2022 Jul 31];8(3):1002586. Available from: /pmc/articles/PMC3297582/
23. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. Vol. 38, FEMS Microbiology Reviews. Blackwell Publishing Ltd; 2014. p. 473–92.
24. Berka RM, Hahn J, Albano M, Draskovic I, Persuh M, Cui X, et al. Microarray analysis of the Bacillus subtilis K-state: genome-wide expression changes dependent on ComK. Mol Microbiol [Internet]. 2002 [cited 2022 Jul 31];43(5):1331–45. Available from: https://pubmed.ncbi.nlm.nih.gov/11918817/
25. Sinha S, Mell J, Redfield R. The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. Mol Microbiol [Internet]. 2013 Jun [cited 2022 Jul 31];88(6):1106. Available from: /pmc/articles/PMC3739930/
26. Charpentier X, Polard P, Claverys JP. Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS? Curr Opin Microbiol. 2012 Oct 1;15(5):570–6.
27. Charpentier X, Polard P, Claverys JP. Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS? Curr Opin Microbiol. 2012 Oct 1;15(5):570–6.
28. Moradigaravand D, Engelstädter J. The evolution of natural competence: Disentangling costs and benefits of sex in bacteria. American Naturalist. 2013 Oct;182(4).
29. Ibáñez de Aldecoa AL, Zafra O, González-Pastor JE. Mechanisms and regulation of extracellular DNA release and its biological roles in microbial communities. Vol. 8, Frontiers in Microbiology. Frontiers Media S.A.; 2017.
30. Dillard JP, Seifert HS. A variable genetic island specific for Neisseria gonorrhoeae is involved in providing DNA for natural transformation and is found more often

in disseminated infection isolates. Mol Microbiol [Internet]. 2001 [cited 2022 Aug 1];41(1):263–77. Available from: https://pubmed.ncbi.nlm.nih.gov/11454218/
31. Hamilton HL, Domínguez NM, Schwartz KJ, Hackett KT, Dillard JP. Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system. Mol Microbiol [Internet]. 2005 Mar [cited 2022 Aug 1];55(6):1704–21. Available from: https://pubmed.ncbi.nlm.nih.gov/15752195/
32. Steinmoen H, Knutsen E, Håvarstein LS. Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. Proc Natl Acad Sci U S A [Internet]. 2002 May 28 [cited 2022 Aug 1];99(11):7681–6. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.112464599
33. Borgeaud S, Metzger LC, Scrignari T, Blokesch M. The type VI secretion system of Vibrio cholerae fosters horizontal gene transfer. Science [Internet]. 2015 Jan 2 [cited 2022 Aug 1];347(6217):63–7. Available from: https://pubmed.ncbi.nlm.nih.gov/25554784/
34. Fontaine L, Wahl A, Fléchard M, Mignolet J, Hols P. Regulation of competence for natural transformation in streptococci. Infection, Genetics and Evolution. 2015 Jul 1;33:343–60.
35. Hofreuter D, Odenbreit S, Haas R. Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. Mol Microbiol [Internet]. 2001 [cited 2022 Aug 1];41(2):379–91. Available from: https://pubmed.ncbi.nlm.nih.gov/11489125/
36. Mejeans V, Claverysq JP. The Journal of Biological Chemistry DNA Processing during Entry in Transformation of Streptococcus pneumoniae*. 1993;268(8):5594–9.
37. Méjean V, Claverys JP. Polarity of DNA entry in transformation of Streptococcus pneumoniae. MGG Molecular & General Genetics. 1988 Aug;213(2–3):444–8.
38. Sisco KL, Smith HO. Sequence-specific DNA uptake in Haemophilus transformation. Proc Natl Acad Sci U S A [Internet]. 1979 [cited 2022 Aug 1];76(2):972–6. Available from: https://pubmed.ncbi.nlm.nih.gov/311478/
39. Goodman SD, Scocca JJ. Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae. Proc Natl Acad Sci U S A [Internet]. 1988 [cited 2022 Aug 1];85(18):6982–6. Available from: https://pubmed.ncbi.nlm.nih.gov/3137581/
40. Chen I, Dubnau D. DNA uptake during bacterial transformation. Vol. 2, Nature Reviews Microbiology. 2004. p. 241–9.
41. Ray CGeorge, Ryan KJ. Sherris Medical Microbiology : an Introduction to Infectious Diseases. 2004;992.
42. Henry George Liddell, Robert Scott, A Greek-English Lexicon, στρεπτ-ός [Internet]. [cited 2022 Aug 2]. Available from: https://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=strepto/s
43. Facklam R. What happened to the streptococci: Overview of taxonomic and nomenclature changes. Vol. 15, Clinical Microbiology Reviews. 2002. p. 613–30.
44. Delorme C. Safety assessment of dairy microorganisms: Streptococcus thermophilus. Int J Food Microbiol. 2008 Sep 1;126(3):274–7.
45. Sherman JM. The Streptococci. Bacteriol Rev [Internet]. 1937 Dec [cited 2022 Aug 2];1(1):3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC440821/

46. Lancefield RC. A serological differentiation of human and other groups of hemolytic streptococci. J Exp Med [Internet]. 1933 Mar 3 [cited 2022 Aug 2];57(4):571. Available from: /pmc/articles/PMC2132252/?report=abstract
47. Raina V, Nayak T, Ray L, Kumari K, Suar M. A Polyphasic Taxonomic Approach for Designation and Description of Novel Microbial Species. Microbial Diversity in the Genomic Era. 2018 Oct 16;137–52.
48. Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol Rev [Internet]. 1994 Oct 1 [cited 2022 Aug 2];15(2–3):155–73. Available from: https://academic.oup.com/femsre/article/15/2-3/155/612145
49. Delorme C, Abraham AL, Renault P, Guédon E. Genomics of Streptococcus salivarius, a major human commensal. Infection, Genetics and Evolution. 2015 Jul 1;33:381–92.
50. Chi F, Nolte O, Bergmann C, Ip M, Hakenbeck R. Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in Streptococcus pneumoniae, S. mitis and S. oralis. Int J Med Microbiol [Internet]. 2007 Nov 12 [cited 2022 Aug 2];297(7–8):503–12. Available from: https://pubmed.ncbi.nlm.nih.gov/17459765/
51. Johnsborg O, Håvarstein LS. Regulation of natural genetic transformation and acquisition of transforming DNA in Streptococcus pneumoniae. In: FEMS Microbiology Reviews. 2009. p. 627–42.
52. Shanker E, Morrison DA, Talagas A, Nessler S, Federle MJ, Prehna G. Pheromone Recognition and Selectivity by ComR Proteins among Streptococcus Species. PLoS Pathog. 2016 Dec 1;12(12).
53. Fontaine L, Boutry C, De Frahan MH, Delplace B, Fremaux C, Horvath P, et al. A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. J Bacteriol. 2010 Mar;192(5):1444–54.
54. Mashburn-Warren L, Morrison DA, Federle MJ. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. Mol Microbiol. 2010 Nov;78(3):589–606.
55. Lingeswaran A, Metton C, Henry C, Monnet V, Juillard V, Gardan R. Export of RGG quorum sensing peptides is mediated by the PPTAB ABC transporter in streptococcus thermophilus strain LMD-9. Genes (Basel). 2020 Sep 1;11(9):1–22.
56. Gardan R, Besset C, Gitton C, Guillot A, Fontaine L, Hols P, et al. Extracellular life cycle of ComS, the competence-stimulating peptide of Streptococcus thermophilus. J Bacteriol. 2013 Apr;195(8):1845–55.
57. Fontaine L, Goffin P, Dubout H, Delplace B, Baulard A, Lecat-Guillet N, et al. Mechanism of competence activation by the ComRS signalling system in streptococci. Mol Microbiol. 2013 Mar;87(6):1113–32.
58. Knoops A, Capelle F Vande, Fontaine L, Verhaegen M, Mignolet J, Goffin P, et al. The CovRS Environmental Sensor Directly Controls the ComRS Signaling System To Orchestrate Competence Bimodality in Salivarius Streptococci [Internet]. 2022. Available from: https://journals.asm.org/journal/mbio
59. Knoops A, Ledesma-García L, Waegemans A, Lamontagne M, Decat B, Degand H, et al. Competence shut-off by intracellular pheromone degradation in salivarius streptococci. PLoS Genet. 2022 May 25;18(5).
60. Wahl A, Servais F, Drucbert AS, Foulon C, Fontaine L, Hols P. Control of natural transformation in salivarius streptococci through specific degradation of σX by the MecA-ClpCP protease complex. J Bacteriol. 2014;196(15):2807–16.

61. Schleifer KH, Ehrmann M, Krusch U, Neve H. Revival of the Species Streptococcus thermophilus (ex Orla-Jensen, 1919) nom. rev. Syst Appl Microbiol. 1991 Oct 1;14(4):386–8.
62. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R. Specialized adaptation of a lactic acid bacterium to the milk environment: The comparative genomics of Streptococcus thermophilus LMD-9. Microb Cell Fact. 2011 Aug 30;10(SUPPL. 1).
63. Iyer R, Tomar SK, Uma Maheswari T, Singh R. Streptococcus thermophilus strains: Multifunctional lactic acid bacteria. Vol. 20, International Dairy Journal. 2010. p. 133–41.
64. Gardan R, Besset C, Guillot A, Gitton C, Monnet V. The oligopeptide transport system is essential for the development of natural competence in Streptococcus thermophilus strain LMD-9. J Bacteriol. 2009 Jul;191(14):4647–55.
65. Whiley RA, Hardie JM. Streptococcus vestibularis sp. nov. from the Human Oral Cavity. Vol. 38, International Journal of Systematic Bacteriology. 1988.
66. Doyuk E, Ormerod OJ, Bowler ICJW. Native value endocarditis due to Streptococcus vestibularis and Streptococcus oralis. Journal of Infection. 2002;45(1):39–41.
67. Simsek AD, Sezer S, Ozdemir NF, Mehmet H. Streptococcus vestibularis bacteremia following dental extraction in a patient on long-term hemodialysis: a case report. NDT Plus [Internet]. 2008 Aug [cited 2022 Aug 3];1(4):276. Available from: /pmc/articles/PMC4421209/
68. Ledesma-Garcia L, Thuillier J, Guzman-Espinola A, Ensinck I, de la Sierra-Gallay IL, Lazar N, et al. Molecular dissection of pheromone selectivity in the competence signaling system ComRS of streptococci. Proc Natl Acad Sci U S A [Internet]. 2020 Apr 7 [cited 2023 Aug 8];117(14):7745–54. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1916085117
69. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg proteins associated with internalized small hydrophobic peptides: A new quorum-sensing mechanism in streptococci. Mol Microbiol. 2011 May;80(4):1102–19.
70. Zaccaria E, van Baarlen P, de Greeff A, Morrison DA, Smith H, Wells JM. Control of Competence for DNA Transformation in Streptococcus suis by Genetically Transferable Pherotypes. PLoS One [Internet]. 2014 Jun 26 [cited 2022 Aug 28];9(6):e99394. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0099394
71. Blatch GL, Lässle M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays [Internet]. 1999 [cited 2022 Jul 30];21(11):932–9. Available from: https://pubmed.ncbi.nlm.nih.gov/10517866/
72. Talagas A, Fontaine L, Ledesma-Garca L, Mignolet J, Li de la Sierra-Gallay I, Lazar N, et al. Structural Insights into Streptococcal Competence Regulation by the Cell-to-Cell Communication System ComRS. PLoS Pathog. 2016 Dec 1;12(12).
73. Neiditch MB, Capodagli GC, Prehna G, Federle MJ. Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria Introduction: a family of cytoplasmic proteins that serve as pheromone receptors. 2017; Available from: https://doi.org/10.1146/annurev-genet-120116-
74. Wang CY, Dawid S. Mobilization of Bacteriocins during Competence in Streptococci. Trends Microbiol. 2018 May 1;26(5):389–91.
75. Li YH, Tang N, Aspiras MB, Lau PCY, Lee JH, Ellen RP, et al. A quorum-sensing signaling system essential for genetic competence in Streptococcus mutans is involved in biofilm formation. J Bacteriol. 2002;184(10):2699–708.

76. Streptococcus thermophilus Orla-Jensen - BAA-491 | ATCC [Internet]. [cited 2023 Apr 3]. Available from: https://www.atcc.org/products/baa-491
77. ASM18007v1 - Genome - Assembly - NCBI [Internet]. [cited 2023 Apr 3]. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000180075.1
78. One Shot™ TOP10 Electrocomp™ *E. coli* [Internet]. [cited 2023 Apr 3]. Available from: https://www.thermofisher.com/order/catalog/product/C404050
79. TOP10 - EcoliWiki [Internet]. [cited 2023 Apr 3]. Available from: https://ecoliwiki.org/colipedia/index.php/TOP10
80. pBAD/His Kit [Internet]. [cited 2023 Apr 4]. Available from: https://www.thermofisher.com/order/catalog/product/V43001
81. BD BBL™ GasPak™ jars [Internet]. [cited 2023 Apr 27]. Available from: https://www.bd.com/en-us/products-and-solutions/products/product-families/bd-bbl-gaspak-jars
82. Bertani G. Studies on Lysogenesis I. : The Mode of Phage Liberation by Lysogenic Escherichia coli1. J Bacteriol [Internet]. 1951 Sep [cited 2023 Apr 27];62(3):293. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC386127/
83. BD Difco™ Dehydrated Culture Media: M 17 Broth M17 Broth; 500g BD Difco™ Dehydrated Culture Media: M 17 Broth | Fisher Scientific [Internet]. [cited 2023 Apr 27]. Available from: https://www.fishersci.be/shop/products/bd-difco-dehydrated-culture-media-m-17-broth/11708293
84. 6.3B: Complex and Synthetic Media - Biology LibreTexts [Internet]. [cited 2023 Apr 27]. Available from: https://bio.libretexts.org/Bookshelves/Microbiology/Microbiology_(Boundless)/06%3A_Culturing_Microorganisms/6.03%3A_Culturing_Bacteria/6.3B%3A_Complex_and_Synthetic_Media
85. Improvement of the ComRS quorum-sensing system for competence and predation in streptococci | Mémoire UCL [Internet]. [cited 2023 Apr 27]. Available from: https://dial.uclouvain.be/memoire/ucl/en/object/thesis%3A28301
86. Ultrapure water purification systems, Milli-Q® EQ 7000 | VWR [Internet]. [cited 2023 Apr 27]. Available from: https://be.vwr.com/store/product/en/35476154/ultrapure-water-purification-systems-milli-q-eq-7000
87. GenElute Plasmid DNA Miniprep Kit - 70 Purifications Gen Elute [Internet]. [cited 2023 Apr 28]. Available from: https://www.sigmaaldrich.com/BE/en/product/sigma/pln70
88. NEB Tm Calculator [Internet]. [cited 2023 Apr 28]. Available from: https://tmcalculator.neb.com/#!/main
89. Overlap extension polymerase chain reaction - Wikipedia [Internet]. [cited 2023 Apr 29]. Available from: https://en.wikipedia.org/wiki/Overlap_extension_polymerase_chain_reaction
90. QIAquick PCR Purification Kit [Internet]. [cited 2023 Apr 30]. Available from: https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-pcr-purification-kit?catno=28104
91. NanoDrop™ 2000/2000c Spectrophotometers [Internet]. [cited 2023 Apr 28]. Available from: https://www.thermofisher.com/order/catalog/product/ND-2000
92. Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 2004;32(14).

93. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343–5.
94. Membrane MF-Millipore, esters de cellulose, hydrophile, 0,025 µm, 13 mm, blanche, unie | VSWP01300 [Internet]. [cited 2023 May 1]. Available from: https://www.merckmillipore.com/BE/fr/product/MF-Millipore-Membrane-Filter-0.025m-pore-size,MM_NF-VSWP01300#relations
95. DpnI | NEB [Internet]. [cited 2023 May 1]. Available from: https://international.neb.com/products/r0176-dpni#Product%20Information
96. Orange DNA Loading Dye (6X) [Internet]. [cited 2023 May 7]. Available from: https://www.thermofisher.com/order/catalog/product/R0631
97. Lepecq JB, Paoletti C. A fluorescent complex between ethidium bromide and nucleic acids. Physical-Chemical characterization. J Mol Biol. 1967 Jul 14;27(1):87–106.
98. DNA molecular weight - Smart Ladder | Eurogentec [Internet]. [cited 2023 May 4]. Available from: https://www.eurogentec.com/en/smartladder
99. Smith BJ. SDS Polyacrylamide Gel Electrophoresis of Proteins. In: Walker J.M. (eds) Proteins. Methods in Molecular Biology™. Walker JM, editor. J Chromatogr A. 1984;1:41–55.
100. Mini-PROTEAN ® 3 Cell Instruction Manual Catalog Numbers 165-3301 165-3302.
101. PageRuler™ Prestained Protein Ladder, 10 to 180 kDa [Internet]. [cited 2023 May 21]. Available from: https://www.thermofisher.com/order/catalog/product/26616
102. InstantBlue™ protein stain | VWR [Internet]. [cited 2023 May 20]. Available from: https://fr.vwr.com/store/product/23392979/protein-stain-instantblue
103. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol [Internet]. 1995 [cited 2023 May 14];177(14):4121. Available from: /pmc/articles/PMC177145/?report=abstract
104. Schmidt TGM, Skerra A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nature Protocols 2007 2:6 [Internet]. 2007 Jun 14 [cited 2023 May 29];2(6):1528–35. Available from: https://www.nature.com/articles/nprot.2007.209
105. Strep-Tactin® Sepharose® column | 1 ml | 2-1202-001 [Internet]. [cited 2023 May 29]. Available from: https://www.iba-lifesciences.com/strep-tactin-sepharose-column/2-1202-001
106. Kroon C, Breuer L, Jones L, An J, Akan A, Ali EAM, et al. Blind spots on western blots: Assessment of common problems in western blot figures and methods reporting with recommendations to improve them. PLoS Biol [Internet]. 2022 Sep 1 [cited 2023 May 21];20(9):e3001783. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001783
107. Chemiluminescent Western Blotting | Thermo Fisher Scientific - BE [Internet]. [cited 2023 May 21]. Available from: https://www.thermofisher.com/be/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/chemiluminescent-western-blotting.html

108. Mignolet J, Fontaine L, Sass A, Nannan C, Mahillon J, Coenye T, et al. Circuitry Rewiring Directly Couples Competence to Predation in the Gut Dweller Streptococcus salivarius. Cell Rep. 2018 Feb 13;22(7):1627–38.
109. Mini Trans-Blot ® Electrophoretic Transfer Cell Instruction Manual.
110. Protec OPTIMAX X-Ray Film Processor [Internet]. [cited 2023 May 27]. Available from: https://www.raytechdiagnostics.com/Product/Details/Protec-OPTIMAX-X-Ray-Film-Processor/Default
111. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nature Protocols 2007 2:8 [Internet]. 2007 Jul 26 [cited 2023 May 29];2(8):1849–61. Available from: https://www.nature.com/articles/nprot.2007.249
112. CELLSTAR® plaque, 96w, F, µClear®, TC, - 655098 [Internet]. [cited 2023 May 30]. Available from: https://shop.gbo.com/fr/belgium/products/bioscience/culture-cellulaire/microplaques-cellstar/microplaques-96-puits-pour-culture-cellulaire-transparente/655098.html
113. Lecteur de plaques multi-modes, Varioskan® Flash | VWR [Internet]. [cited 2023 May 30]. Available from: https://fr.vwr.com/store/product/591798/lecteur-de-plaques-multi-modes-varioskan-flash
114. RCSB PDB - 5JUB: Crystal structure of ComR from S.thermophilus in complex with DNA and its signalling peptide ComS. [Internet]. [cited 2023 Jul 23]. Available from: https://www.rcsb.org/structure/5jub
115. Align | UniProt [Internet]. [cited 2023 Jul 23]. Available from: https://www.uniprot.org/align

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/sc**