UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Caracterización de dos activadores transcripcionales: MafR de Enterococcus faecalis y MgaP de Streptococcus pneumoniae

Characterization of two transcriptional activators: MafR of Enterococcus faecalis and MgaP of Streptococcus pneumoniae

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Ana Moreno Blanco

DIRECTORA

Alicia Bravo García

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular

TESIS DOCTORAL

Caracterización de dos activadores transcripcionales: MafR de *Enterococcus faecalis* y MgaP de *Streptococcus pneumoniae*

Characterization of two transcriptional activators: MafR of *Enterococcus faecalis* and MgaP of *Streptococcus pneumoniae*

Ana Moreno Blanco

Directora:

Alicia Bravo García

Madrid, 2020

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

CARACTERIZACIÓN DE DOS ACTIVADORES TRANSCRIPCIONALES: MafR DE *Enterococcus faecalis* Y MgaP DE *Streptococcus pneumoniae*

CHARACTERIZATION OF TWO TRANSCRIPTIONAL ACTIVATORS: MafR OF *Enterococcus faecalis* AND MgaP OF *Streptococcus pneumoniae*

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

ANA MORENO BLANCO

DIRECTORA

ALICIA BRAVO GARCÍA

CENTRO DE INVESTIGACIONES BIOLÓGICAS MARGARITA SALAS

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC)

MADRID, 2020

A mi madre, mi padre y mis hermanos.

" Que no sea inmortal puesto que es llama,

pero que sea eterno mientras dure"

Vinicius de Moraes

Poco espacio para expresar todo lo que llevo dentro. Ciertas palabras se repiten una y otra vez a lo largo de estos agradecimientos, "orgullo", "confianza", "vida", "motivar"... No es falta de léxico, es ser afortunada por haberos encontrado en mi camino.

A ti, **Alicia**, por tu infinita paciencia y buen hacer. Por darme la oportunidad y confiar en mí. Muchas gracias por tu eterna dedicación y todas las horas que has invertido en mi formación. Gracias por tu generosidad y por todas las celebraciones que hemos podido disfrutar, y por las que espero poder seguir celebrando. La vida está hecha de momentos, así que vive.

Manolo, por tu inmensa sabiduría, tus ganas de aprender y sobre todo, tus ganas de "chinchar" y meter el gusanillo de la curiosidad en los cerebritos que te vas encontrando. A los dos, “más, porque mejor es imposible”. No podría haber aprendido en un sitio mejor.

Sofi, gracias por todo lo que me has enseñado tanto como persona como profesionalmente. Qué bonito que nuestra amistad empezase con el *flow* del DuoFlow, se fortaleciese con el Ac. Acético y ^{32}P, y continuara con ¡¡Atrévete-te!! Queda pendiente el burocracy-scaperoom, y todo lo que aún no hemos descubierto. Aunque no ha sido posible, para mí siempre serás mi Co-directora, y sin duda, mi couch-animadora-motivadora. Sin ti y sin Alicia, no lo hubiera conseguido.

Itzi, a ti por tus ganas de vida. Gracias por motivar a todo aquel que se cruza en tu camino, por ofrecer todo de ti y dar hasta las entrañas, por ser como eres, y enseñar como enseñas.

A **Glori** y **Paloma**, por dejar que me acercara a vuestras mentes, que aunque son muy diferentes entre sí, son de lo más maravilloso del país. **Glori** por tu infinito apoyo, ánimo e inteligencia. **Paloma** por tu curiosidad, por acogerme como otra más del labo y por compartirme tu amor y experiencias por la ciencia.

Josecin, echaré de menos pisarte tus zapatos impolutos y patearte por debajo de la mesa. Muchas gracias por aguantarme y escucharme; y sobre todo por responder siempre con una carcajada al "¿Tenéis comidita? - me muero de hambre", cuando aparecía en vuestro laboratorio a altas horas de la tarde (también a Glori, mi oficial proveedora de pastas y quesos, y a "GSD management"). **Rafica**, gracias por tu enorme y constante sonrisa que le mejora el día a cualquiera, por tener siempre una buena palabra en la boca y estar ahí con tus “¡Qué pasa Anica!”. **Javi**, también conocido como CO de GSD-management. Por tus infinitos "Che Ana, Relájate!" y tus incesantes propuestas de planes, eso sí, siempre con la "bucita campera". **Diego**, por tener siempre palabras bonitas y ser siempre tan "excelente". A **Cathy**, la chica que más brilla de todo el CIB. Thank you for your support and be always by my side. You are just amazing. **Adriancin**, por las risas y los bailes que nos hemos echado, que algún día nos lanzarán al estrellato, “porque en España la ciencia está muy mal, y eso ha sido así, de siempre…”; Gracias por el ambiente que creas allí donde vas, siendo el "β-glucano" de la planta tres. **Gorettita**, por la alegría y energía que desprendes allí donde vas. **Norhane**, merci beaucoup pour tes sourires et ta paix intérieure.

A **Mª Luz**, **Montse**, **Kenza**, **Annel**, **Paula**, **Inés, Lore, Vero, Jingwen, Virtu** y todas las personas que han pasado por el CIB, brindándome su ayuda y dejando en mí su huella. Gracias por los "breaks" por los pasillos o en el comedor.

A **Rosa** y su gente, sois los que habéis estado en mi peor momento de la Tesis. Gracias por confiar en mí y darme la oportunidad de unirme a vuestro mundo en una situación tan delicada. Nunca tendré forma de agradeceros la ayuda, facilidad y alegría que me habéis dado.

A mis **XXLKS**, simplemente por estar siempre, por apoyarme y secuestrarme con un "tómate una cerveza, que eso te quita todo". Gracias por todos los viajes y retiros: por Cadalso, Peguerinos, Xeraco, Cork, Interrail… por todo en general, porque ¡Bastante hemos fregao' ya! Y obviamente, gracias por el soporte, sicológico, técnico, de diseño y administrativo, entre otros mil. Aquí, yo ya cierro mi cuaderno de "aborto". Os quiero churris.

A mis **Biologogos**, si de algo puedo estar más orgullosa, aparte de haber acabado la carrera, es de haberos conocido y tener este grupo. Lo mejor de la universidad, sin duda, habéis sido vosotros. Desde los "visu-culo" en el metro, pasando por los vivac en la Pedriza hasta las noches por Madrid. Especial mención a **Chen**, sin ti a saber qué sería de mí. A ti por ser tú, porque tú entero alegras el alma, mi alma. A mi **rubito**, **Javi**, por estar siempre en mis sueños y por haberme dejado un huequito en ese corazoncito. **Blanditas**, gracias por escuchar mis dramas continuos y estar ahí. Quedan pendientes esos viajes que nos cultivan el espíritu. **Marina** y **Ali**, con vosotras, un mundo. **Marina**, por esa mirada que inclina la cabeza y extiende los brazos, para acabar dándome ese abrazo que me levanta y saca del pozo. Siempre con tu sonrisa y ese corazón que tienes que retumba de lo fuerte que late. **Ali**, por todos los ratos de "cosquis" y por tus constantes ganas de ayudar, gracias. Que tu inquietud y arte por nuevas manualidades no desaparezca nunca, artista. **Hajni**, the strongest woman in all Hungary. Without you, **Lori** and **Daniela**, we had not been able to survive. Thank you for all your love and kind.

A **Julia** que siempre me muestra su admiración, haciendo que confíe más en mi misma. Contigo todo es fácil y simple, y pese a las adversidades siempre sales a flote: "la palmera que se dobla, pero que resiste el huracán". No puedo estar más orgullosa de ti. Gracias por estar en mi vida.

Águeda, porque la distancia nunca fue problema para nosotras. Por ese mes que ha forjado nuestra amistad año tras año. Gracias por incluirme en tu enorme mundo. "Anem a la platja bussejar" y de cenar "algo ligerito".

A **Paqui** y **Manolo**, gracias por toda vuestra ayuda y generosidad. Gracias por acogerme como otra hija más.

Me siento orgullosa de muchas cosas en mi vida, pero sin duda la que está en el top es mi **Familia**. Porque aunque nunca acabasteis de entender que es esto, nunca dejasteis de apoyarme y confiar en que lo lograría. A mis tetes por abrirme el camino e ir curtiéndome en la

vida. **Iván**, a ti por tu corazón y por darlo todo. Porque lo que te pedimos se convierte en deseos cumplidos. También por tu cariño y afecto, "Salta pequeño canguro". **Cris**, gracias por llegar a estar familia y por todo lo que estás haciendo por nosotros.

Fran, gracias por tu esfuerzo por querer entender lo que hago y el motivo por el cual lo hago, significa mucho para mí. No se me olvidarán tus palabras recién entrada en la carrera, "sé la mejor en lo que haces, y nunca te faltará trabajo". Quizás nunca llegue a ser la mejor, pero si me seguiré esforzando en conseguirlo, por vosotros y por mí. Gracias por actuar como un muro de carga cuando todo tambalea. A ti **Almu**, por acogerme como una hermana pequeña y ayudarme siempre que te lo pido. **Cuñis**, porque realmente pienso que habéis mejorado esta familia.

A ti **papá**, por todas tardes que hemos pasado juntos en la calle, jugando, y no entre cuatro paredes. Por enseñarme la vida, desde habilidades sociales y hasta el amor por los seres vivos. Gracias por tu infinita paciencia y tus "Bueno, no pasa nada" cuando la "liaba". Gracias por todo lo que habéis trabajado tu y mamá para que nunca nos faltase nada a los tetes y a mí. Gracias por enseñarme bondad y respeto. **Mamá**, por ti, la vida. No conozco a otra persona que esté más orgullosa de sus hijos, pero tampoco conozco a unos hijos que estén más orgullosos de su madre, de ti. Tu forma de querernos, nos protege de todo. Al final, daba igual cual fuera el resultado, tu siempre has estado orgullosa de nosotros por llegar, motivándonos a seguir luchando. Tu también, **papá**. Gracias por todo lo que has/habéis luchado y trabajado para formar esta familia. Gracias por quitaros el pan de la boca para dárnoslo a nosotros. Gracias por darme la oportunidad de estudiar en la universidad y darme la libertad de formarme como persona, explorando otros ambientes. Por confiar en mí. Gracias por sembrar la semilla que me hace ser un "culo inquieto". Gracias por enseñarme que si te gusta lo que haces, el trabajo es otro juego. Gracias por existir, mamá.

Porque no me imagino una familia mejor, os quiero. Sin vosotros, nunca hubiera levantado el vuelo.

Héctor, a ti te lo llevo diciendo una vida. Gracias por ser el soporte y pilar de mi vida. Gracias por aceptar y querer todo lo que llevo dentro, por estar a mi lado, por abrazarme y darme el amor que recarga mi pila. Por tu paciencia y tu habilidad tranquilizándome. Gracias por todo lo que has hecho por mi estos años, pero sobre todo, por lo que has hecho por mi estos meses. Gracias por marcarme el ritmo, cuando me aceleraba quemando los frenos, y por tus "tranquila, que todo va a salir bien". Sin duda sin ti, tampoco lo hubiera conseguido. Lo que te costó llegar al fondo de mi corazón, y lo adherido que estás ahora; porque…“Lo justo, lo necesario, lo suficiente”. Te quiero, pequeño.

Me dejo muchísima gente en el tintero, pero no puedo alargar más, por lo tanto, a todas las personas que os habéis cruzado en mi camino y que de alguna manera me habéis liberado, GRACIAS. Nos vemos por las calles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ABC transporter	**A**TP-**B**inding **C**assette **transporter**
APS	**A**mmonium **P**er**s**ulfate
BHI	**B**rain **H**eart **I**nfusion
bp	**b**ase **p**air
BSA	**B**ovine **S**erum **A**lbumin
cDNA	**c**omplementary **DNA**
cpm	**c**ounts **p**er **m**inute
CPS	**C**apsular **p**oly**s**accharide
CSP-1	**C**ompetence **S**timulating **P**eptide-1
C_T	**T**hreshold **C**ycles
DMSO	**Dim**ethyl **S**ulf**o**xide
dNTPs	**D**eoxyribo**n**ucleotide **T**ri**p**hosphates
DOC	**D**e**o**xy**c**holate
dsDNA	**d**ouble-**s**tranded **DNA**
DTT	**Di**thio**t**hreitol
ECF transporter	**E**nergy-**C**oupling **F**actor **transporter**
EDTA	**E**thylene**d**iamine **t**etra-acetic **a**cid
EMSA	**E**lectrophoretic **M**obility **S**hift **A**ssay
FC	**F**old **C**hange
FPLC	**F**ast **P**ressure **L**iquid **C**hromatography
Fus	**Fus**idic **a**cid
GFP	**G**reen **F**luorescent **P**rotein

HPLC	**H**igh-**P**ressure **L**iquid **C**hromatography
IPTG	**I**so**p**ropyl-β-D-**t**hio**g**alactopyranoside
Km	**K**ana**m**ycin
MSCRAMM	**M**icrobial **S**urface **C**omponents **R**ecognizing **A**dhesive **M**atrix **M**olecules
NCBI	**N**ational **C**enter for **B**iotechnology **I**nformation
nt	**n**ucleo**t**ide
OD	**O**ptical **D**ensity
PA	**P**hosphatase **A**lkaline
PCR	**P**olymerase **C**hain **R**eaction
PEI	**P**oly**e**thylene**i**mine
PNK	**P**oly**n**ucleotide **K**inase
PRD	**P**TS **R**egulation **D**omain
PTS	**P**hosphoenolpyruvate:carbohydrate phospho**t**ransferase **S**ystem
P-type ATPase	**P**hosophorylated intermediate-**type ATPase**
qRT-PCR	**Q**uantitative **R**everse **T**ranscription **P**olymerase **C**hain **R**eaction
Rif	**Rif**ampicin
RNAP	**RNA P**olymerase
rpm	**r**evolutions **p**er **m**inute
RT-PCR	**R**everse **T**ranscription **P**olymerase **C**hain **R**eaction
SDS	**S**odium **D**odecyl **S**ulfate
Tc	**T**etra**c**ycline
TCS	**T**wo-**C**omponent **S**ystem
TEMED	**Te**tra**m**ethyl**e**thylene**d**iamine
Tris	**Tris** (Hydroxymethyl) Aminomethane

TY	**T**ryptone **Y**east Extract
UV	**U**ltra**v**iolet

RESUMEN

Las bacterias Gram-positivas *Enterococcus faecalis* y *Streptococcus pneumoniae* son habitantes naturales del ser humano. *E. faecalis* es un comensal del tracto gastrointestinal, mientras que *S. pneumoniae* puede encontrarse en la nasofaringe de individuos sanos. Sin embargo, su capacidad de adaptación a condiciones adversas les permite colonizar otros nichos y causar infecciones graves. En general, estos procesos de adaptación requieren la acción de proteínas que actúan como reguladores globales de la transcripción. Estas proteínas activan y/o reprimen la transcripción de numerosos genes en respuesta a señales ambientales específicas. La familia de reguladores globales conocida como Mga/AtxA incluye varias proteínas de bacterias patógenas Gram-positivas: AtxA (*Bacillus anthracis*), Mga (*S. pyogenes*), Mga*Spn* (*S. pneumoniae*) y MafR (*E. faecalis*). Varios estudios indican que estos reguladores transcripcionales están involucrados en la virulencia de estas bacterias (Fouet, 2010; Hemsley *et al.*, 2003; McIver, 2009; Ruiz-Cruz *et al.*, 2016).

MafR de *Enterococcus faecalis*

Estudios previos de nuestro laboratorio mostraron que la proteína MafR de *E. faecalis* V583 ($MafR_{V583}$) se une a fragmentos de ADN lineal generando complejos multiméricos y que esta unión tiene poca o ninguna especificidad de secuencia (Ruiz-Cruz, 2015). Además, mediante *microarrays* diseñados para la estirpe OG1RF, se observó que MafR activa, directa o indirectamente, la transcripción de al menos 87 genes. Muchos de estos genes están organizados en operones y codifican proteínas implicadas en la utilización de fuentes de carbono (Ruiz-Cruz *et al.*, 2016). En esta Tesis, hemos trabajado en los siguientes **objetivos: (a)** Análisis de las propiedades de unión a ADN de la proteína MafR de la estirpe OG1RF ($MafR_{OG1RF}$) y **(b)** Identificación de genes regulados directamente por $MafR_{OG1RF}$.

Resultados y Conclusiones

$MafR_{OG1RF}$ difiere de $MafR_{V583}$ en cinco aminoácidos, tres de ellos (Ala37Thr, Gln131Leu, Met145Thr) localizados en el supuesto dominio de unión a ADN (residuos 11 a 164). Hemos purificado una versión de $MafR_{OG1RF}$ con cola de histidinas ($MafR_{OG1RF}$-His), así como una variante de $MafR_{OG1RF}$-His que carece de los tres primeros aminoácidos ($MafR_{OG1RF}$Δ3N-His). Mediante experimentos de retraso en gel, hemos observado que $MafR_{OG1RF}$-His, al contrario que $MafR_{OG1RF}$Δ3N-His, genera complejos multiméricos en diferentes fragmentos de ADN lineal. Por consiguiente, ni la presencia de la cola de histidinas en el extremo C-terminal, ni las sustituciones aminoacídicas afectan la unión de MafR al ADN. Sin embargo, la deleción de los tres primeros aminoácidos resulta en una variante incapaz de interaccionar con el ADN, lo cual

sugiere que esos residuos son cruciales para la estructura y/o función de la proteína. La formación de complejos multiméricos en fragmentos de ADN lineal ha sido también descrita para el regulador transcripcional Mga*Spn* de neumococo (Solano-Collado *et al.*, 2013).

Mediante técnicas de qRT-PCR, *primer extension*, fusiones transcripcionales y *footprinting* con DNasa I, hemos demostrado que MafR activa directamente la transcripción de los genes *OG1RF_12294*, *OG1RF_11486* y *OG1RF_10478*. Según la predicción de programas bioinformáticos, estos genes estarían involucrados en el transporte de calcio, en la incorporación de un precursor de la queuosina y en el transporte de celobiosa, respectivamente, lo que sugiere que MafR tiene un papel regulador en estos procesos. También, hemos demostrado que MafR incrementa la actividad de sus promotores diana uniéndose a un sitio específico que solapa con el elemento -35. Los tres sitios de unión de MafR identificados en este trabajo muestran una identidad de secuencia baja y están localizados en, o flanqueados por, regiones de posible *bendability*. Proponemos que MafR reconoce sitios específicos del ADN por un mecanismo que implica el reconocimiento de características estructurales del ADN (*shape readout mechanism*).

MgaP de *Streptococcus pneumoniae*

La estirpe R6 de *S. pneumoniae* posee varios clústeres genéticos que no se encuentran en la estirpe TIGR4 (Brückner *et al.*, 2004). Uno de ellos contiene dos genes divergentes, *spr1403* (*pclA*) y *spr1404* (aquí llamado *mgaP*). El análisis de la distribución de este clúster entre numerosos aislados clínicos reveló que era un clúster específico de estirpe (Imai *et al.*, 2011; Paterson *et al.*, 2008). Buscando homologías de secuencias aminoacídicas, encontramos que MgaP tiene similitud (60.3%) con el regulador Mga*Spn*, el cual es un miembro de la familia Mga/AtxA que está presente en TIGR4 y R6 (Hemsley *et al.*, 2003; Solano-Collado *et al.*, 2012). En esta Tesis, hemos trabajado en los siguientes **objetivos: (a)** Analizar la expresión del gen *mgaP* en *S. pneumoniae* R6, **(b)** Estudiar el efecto de MgaP sobre la transcripción del gen *pclA* (proteína de tipo colágeno), y **(c)** Analizar si las proteínas MgaP y Mga*Spn* tienen características comunes en su interacción con el ADN.

Resultados y Conclusiones

Mediante qRT-PCR, hemos observado que el gen *mgaP* se transcribe en condiciones de laboratorio. Comparado con la fase estacionaria, la transcripción de *mgaP* es mayor en la fase logarítmica del crecimiento bacteriano. Además, por *primer extension* y fusiones

transcripcionales, hemos mostrado que la transcripción de *mgaP* comienza en el promotor *PmgaP*, el cual no parece estar auto-regulado.

Hemos demostrado que MgaP es un activador transcripcional. Utilizando estirpes de neumococo que sintetizan diferentes niveles de MgaP, hemos observado que MgaP incrementa los niveles de transcripción del gen *pclA* (qRT-PCR), así como la actividad del promotor *PpclA* (fusiones transcripcionales). Esta activación requiere la presencia de una región localizada aguas arriba del promotor *PpclA* (posiciones -76 a -106). Además, mediante *footprinting* con DNasa I y usando una versión de MgaP con cola de histidinas (MgaP-His), hemos demostrado que MgaP reconoce un sitio específico aguas arriba del promotor *PpclA* (posiciones -68 a -169).

Estudios previos de nuestro laboratorio mostraron que Mga*Spn* activa al promotor *P1623B* uniéndose a un sitio específico aguas arriba del promotor (posiciones -60 a -99) (Solano-Collado *et al.*, 2013). En este trabajo, por análisis *in silico*, hemos encontrado que MgaP y Mga*Spn* presentan una idéntica organización de dominios funcionales. También, mediante ensayos de retraso en gel, hemos visto que MgaP, al igual que Mga*Spn*, puede interaccionar con fragmentos de ADN lineal sin reconocer una secuencia nucleotídica específica y generar complejos multiméricos. No obstante, mediante experimentos de *footprinting* con DNasa I, hemos demostrado que estas proteínas presentan diferentes especificidades de unión a ADN: MgaP no reconoce el sitio de unión de Mga*Spn* y viceversa.

SUMMARY

The Gram-positive bacteria *Enterococcus faecalis* and *Streptococcus pneumoniae* are inhabitants of the human body. Whereas *E. faecalis* is a usual commensal of the gastrointestinal tract, *S. pneumoniae* is normally found in the nasopharynx of healthy individuals. Nevertheless, their ability to adapt to diverse stress conditions allows them to invade other niches of the human host causing a variety of life-threatening infections. In general, bacterial adaptation to new niches requires the action of proteins that act as global transcriptional regulators. These proteins activate and/or repress the transcription of multiple genes in response to specific signals. The Mga/AtxA family of global transcriptional regulators includes various proteins of Gram-positive pathogenic bacteria: AtxA (*Bacillus anthracis*), Mga (*S. pyogenes*), Mga*Spn* (*S. pneumoniae*), and MafR (*E. faecalis*). Several studies have shown that these transcriptional regulators play a role in bacterial virulence (Fouet, 2010; Hemsley *et al.*, 2003; McIver, 2009; Ruiz-Cruz *et al.*, 2016).

MafR from *Enterococcus faecalis*

Previous studies in our laboratory showed that the MafR protein of *E. faecalis* strain V583 ($MafR_{V583}$) binds to linear DNA fragments generating multimeric complexes, and this binding has little or no sequence specificity (Ruiz-Cruz, 2015). Furthermore, using microarrays designed for strain OG1RF, it was shown that MafR activates, directly or indirectly, the transcription of at least 87 genes. Many of them are organized in operons and encode proteins involved in the utilization of carbon sources (Ruiz-Cruz *et al.*, 2016). In this Thesis, we have worked on the following **objectives: (a)** Analysis of the DNA-binding properties of MafR from strain OG1RF ($MafR_{OG1RF}$) and **(b)** Identification of genes regulated directly by $MafR_{OG1RF}$.

Results and Conclusions

Compared to $MafR_{V583,}$ $MafR_{OG1RF}$ has five amino acid changes, and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the predicted DNA-binding domain (residues 11 to 164). We have purified a His-tagged version of $MafR_{OG1RF}$ ($MafR_{OG1RF}$-His), as well as a variant of $MafR_{OG1RF}$-His that lacks the first three amino acid residues ($MafR_{OG1RF}$Δ3N-His). By gel retardation experiments, we have shown that $MafR_{OG1RF}$-His, but not $MafR_{OG1RF}$Δ3N-His, generates multimeric complexes on different linear DNA fragments. Thus, neither the presence of the His-tag at the C-terminus nor the amino acid substitutions affect the DNA-binding behaviour of MafR. However, removal of the first three residues results in a variant unable to interact with DNA, which suggests that such residues are crucial for its structure and/or

function. Formation of multimeric complexes on linear DNA fragments has also been reported for the pneumococcal Mga*Spn* transcriptional regulator (Solano-Collado *et al.*, 2013).

By qRT-PCR, primer extension, transcriptional fusions and DNase I footprinting assays, we have demonstrated that MafR activates directly the transcription of *OG1RF_12294*, *OG1RF_11486* and *OG1RF_10478* genes. These genes are predicted to be involved in calcium transport, uptake of a queuosine precursor and cellobiose transport, respectively, which suggests a regulatory role of MafR in such processes. Furthermore, we have shown that MafR increases the efficiency of its target promoters by binding to a specific site that overlaps the -35 element. The three MafR-binding sites identified in this work exhibit a low sequence identity. They are located at or flanked by regions of potential bendability. We propose that MafR could use a shape readout mechanism to achieve DNA-binding specificity.

MgaP from *Streptococcus pneumoniae*

S. pneumoniae strain R6 has several gene clusters that are not present in strain TIGR4 (Brückner *et al.*, 2004). One of them contains two genes that are divergently transcribed, *spr1403* (*pclA*) and *spr1404* (here named *mgaP*). Analyses of the distribution of this gene cluster among numerous clinical isolates revealed that it is a strain-specific gene cluster (Imai *et al.*, 2011; Paterson *et al.*, 2008). Searching for homologies, we found that MgaP has sequence similarity (60.3%) with the pneumococcal Mga*Spn* regulator, a member of the Mga/AtxA family that is present in both strains TIGR4 and R6 (Hemsley *et al.*, 2003; Solano-Collado *et al.*, 2012). In this Thesis, we have worked on the following **objectives: (a)** Expression analysis of the *mgaP* gene in *S. pneumoniae* R6, **(b)** Study of the effect of MgaP on the transcription of the *pclA* gene (pneumococcal collagen-like protein A), and **(c)** To analyse whether MgaP and Mga*Spn* have common features in their interaction with DNA.

Results and Conclusions

By qRT-PCR, we have shown that *mgaP* is transcribed under laboratory conditions. Compared to the stationary phase, transcription of the *mgaP* gene is higher at logarithmic phase. In addition, by primer extension and transcriptional fusions, we have shown that transcription of *mgaP* is initiated at the *PmgaP* promoter, which does not seem to be self-regulated.

We have demonstrated that MgaP is a transcriptional activator. First, using pneumococcal strains that synthesize different levels of MgaP, we have shown that MgaP increases the transcription level of the *pclA* gene (qRT-PCR), as well as the activity of the *PpclA* promoter

(transcriptional fusions). This activation requires a region located upstream of the *PpclA* promoter (positions -76 to -106). Furthermore, by DNase I footprinting assays and using a His-tagged version of MgaP (MgaP-His), we have established that MgaP recognizes a specific site upstream of the *PpclA* promoter (positions -68 to -169).

Previous studies in our laboratory showed that Mga*Spn* enhances the activity of the *P1623B* promoter by binding to a specific site upstream of the core promoter (positions -60 to -99) (Solano-Collado *et al.*, 2013). Now, by *in silico* analyses, we have found that MgaP and Mga*Spn* exhibit a similar domain organization. In addition, by gel retardation assays, we have shown that, like Mga*Spn*, MgaP binds to linear DNA fragments in a non-sequence-specific manner and generates multimeric complexes. Nonetheless, by DNase I footprinting experiments, we have demonstrated that theses proteins have different DNA-binding specificities. MgaP does not recognize the Mga*Spn* binding site and vice versa.

INTRODUCTION

1. THE OPPORTUNISTIC BACTERIA *Enterococcus faecalis* AND *Streptococcus pneumoniae*

Bacteria can inhabit a wide range of niches. In the last decade, the scientific interest in microorganisms living in the human body has increased considerably. One of the reasons is that several studies on the gut microbiota have shown a connection between the composition of commensal bacterial populations and the development of multiple metabolic and inflammatory human diseases, such as inflammatory bowel disease, obesity, type 2 diabetes, atherosclerosis, allergy, and colon cancer. Moreover, particular commensal bacteria of the gastrointestinal tract are known to be essential for the maintenance of human health, as they may regulate the host immune system and promote host resistance against infectious disease through competition with pathogenic microorganisms (Brestoff and Artis, 2013; Khosravi and Mazmanian, 2013; Rinninella *et al.*, 2019). Compared to the gut microbiota, the nasopharyngeal microbiota has not been extensively studied. Nevertheless, several studies have suggested that the composition of commensal bacteria may contribute to both the occurrence and the severity of viral respiratory infections (Dubourg *et al.*, 2019). During the last years, many research efforts have also been focused on human commensal bacteria able to become pathogenic under certain circumstances (opportunistic bacteria). The adaptive capacity of these bacteria is closely linked to their ability to modulate gene expression in response to specific environmental signals. This is the case of *Enterococcus faecalis,* a commensal of the gastrointestinal tract, and *Streptococcus pneumoniae*, a commensal of the nasopharynx. Both bacteria are able (i) to evade the host immune system, (ii) to adapt to diverse stress conditions, and (iii) to invade different niches of the human host causing a variety of life-threatening infections (Figure 1). Furthermore, their ability to adapt to new niches is strongly associated with the action of global transcriptional regulators.

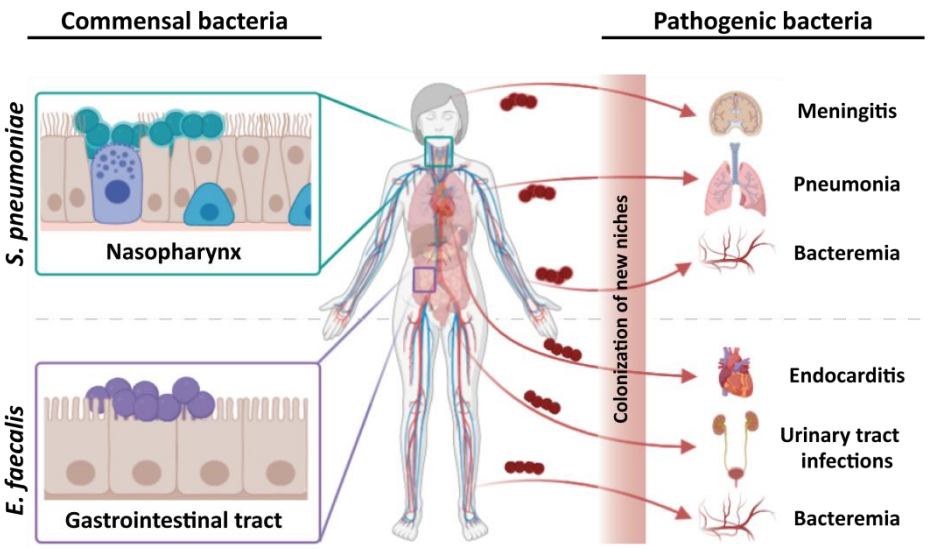


Figure 1. The opportunistic bacteria *S. pneumoniae* and *E. faecalis*. *S. pneumoniae* is a commensal of the nasopharynx, whereas *E. faecalis* is part of the gastrointestinal microbiota. As opportunistic bacteria, they are able to evade the host immune system, to adapt to diverse stress conditions, and, in consequence, to invade different niches of the human host causing a variety of life-threatening infections (*S. pneumoniae*: meningitis, pneumoniae, bacteremia, among others; *E. faecalis*: endocarditis, urinary tract infections, and bacteremia).

Enterococcus faecalis

Among the trillions of mammalian gastrointestinal tract commensal bacteria, we can find *E. faecalis* despite the enterococci typically constitute a minority of the total gut population. *E. faecalis* is a low-GC content Gram-positive, catalase-negative, non-spore forming, facultatively anaerobic bacterium, which is defined by its ovoid-shape and appears in pairs or chains (Fisher and Phillips, 2009; Murray, 1990). Apart from the intestinal tract of mammals, fish, birds and insects, *E. faecalis* also inhabits in plants, soil, marine sediments and water (Martin and Mundt, 1972; Mundt, 1961; Mundt, 1963; Van Tyne and Gilmore, 2014). Additionally, *E. faecalis* has been regularly isolated from cheese, sausages, minced beef and pork. The presence of *Enterococcus* species in animal source foods is associated with contamination due to the capacity of this genus to survive the heating process (it can survive at 60°C for 30 min) (Fisher and Phillips, 2009; Lebreton *et al.*, 2014). Genus *Enterococcus* was initially grouped within the genus *Streptococcus*, however, Sherman (1937) proposed a division based on the ability of enterococci to grow at a wide range of temperatures (from 10°C to 50°C) and pH (from 4.6 to 9.9), and also on its tolerance to high concentrations of NaCl (6.5%) (Sherman, 1937). In 1984, *Enterococcus* was proposed as a unique genus, separated from *Streptococcus*, based on DNA-DNA and DNA-rRNA hybridization experiments (Schleifer and Kilpper-Bälz, 1984). As some lactic acid bacteria, *Enterococcus* produces bacteriocins (enterocins), whose production is favoured in stressful growth conditions; most likely due to

lower growth rates that result in better utilization of energy and greater availability of metabolites for the synthesis of bacteriocins (Aasen *et al.*, 2000; De Vuyst *et al.*, 1996; Parente and Ricciardi, 1994). Besides, the metabolic potential of *E. faecalis* is huge because it can use several energy sources, such as carbohydrates, glycerol, citrate, maltose and lactate, among others (Ramsey *et al.*, 2014). The enterococcal tolerance to desiccation, starvation, extremes of pH, ionizing radiation, osmotic and oxidative stress, high heavy metal concentrations, and antibiotics, makes *E. faecalis* a bacterium rapidly adaptable to dramatic changes in many physiochemical conditions (Lebreton *et al.*, 2014; Ramsey *et al.*, 2014; Van Tyne and Gilmore, 2014). At present, *E. faecalis* is (together with *E. faecium*) the most studied enterococcal species due to its long-standing association with human disease and its emergence as a leading cause of nosocomial infections, such as urinary tract infections, bacteremia, endocarditis and wound infections (Gaca and Lemos, 2019).

E. faecalis has emerged as a multidrug-resistant (MDR) bacterium due to its intrinsically resistance to several antibiotics and its capacity to acquire new antibiotic resistances through sporadic mutations in specific genes or by horizontal gene transfer. Moreover, *E. faecalis* is considered one of the most common bacterial species associated with human disease, likely due to its ability to persist in the hospital settings (Gaca and Lemos, 2019; Hollenbeck and Rice, 2012; Ruiz-Garbajosa *et al.*, 2006). The ability of *E. faecalis* to exchange genetic material not only results in the propagation of resistance genes to other pathogens but also in the acquisition of new adaptive traits, such as bacteriocins, metabolic genes and virulence determinants (Arias and Murray, 2012; Clewell *et al.*, 2014; Hollenbeck and Rice, 2012). Although *E. faecalis* has been used for decades as probiotics in human and farm animals owing to its usually low-virulence level, nowadays its use is controversial due to its potential virulence determinants and its facility to adapt into new environments, which contribute to causing diseases (Arias and Murray, 2012; Franz *et al.*, 2011; Hanchi *et al.*, 2018).

Despite most enterococci do not produce a set of potent pro-inflammatory toxins, several virulence factors have been identified in *E. faecalis*, including secreted factors (Cyl, GelE and SprE), collagen adhesion proteins that may mediate adherence to host tissues (Ace), and some cell surface proteins involved in biofilm production (Esp, AS proteins) and pili formation (Ebp), among others (Arias and Murray, 2012; DebRoy *et al.*, 2014; Gaca and Lemos, 2019).

The secreted factor cytolysin-haemolysin (Cyl) is an exotoxin produced by some strains of *E. faecalis*, which can lyse Gram-positive bacteria and eukaryotic cells, such as human red blood cells and some human white blood cells (Ike *et al.*, 1984). Cyl is encoded on pheromone-responsive plasmids or pathogenicity islands, and it is regulated by a two-component *quorum-sensing* mechanism (Coburn *et al.*, 2004; DebRoy *et al.*, 2014; Dunny and Clewell, 1975; Jacob

et al., 1975; Segarra *et al.*, 1991). Strains expressing *cyl* are more virulent in various animal models than strains lacking *cyl* (Arias and Murray, 2012; Chow *et al.*, 1993; Garsin *et al.*, 2001; Ike *et al.*, 1984). Other secreted factors of *E. faecalis* include hydrolytic enzymes such as gelatinase (GelE) and extracellular serine proteinase (SprE). Both enzymes are able to degrade host tissues to provide nutrients, but GelE also appears to modulate the host immune response (Arias and Murray, 2012; Fisher and Phillips, 2009; Park *et al.*, 2008). GelE is specifically able to hydrolyse gelatin, casein, haemoglobin and other bioactive peptides (Fisher and Phillips, 2009). Furthermore, this zinc-metalloendopeptidase is associated with clearing misfolded proteins and participates in the activation of autolysin (peptide-glycan-degrading enzyme), which leads to the release of extracellular DNA and to the formation of biofilms (Arias and Murray, 2012; Mohamed *et al.*, 2004; Thomas *et al.*, 2009; Waters *et al.*, 2003). The *gelE* gene is cotranscribed with its downstream gene, s*prE*. The *quorum-sensing* system encoded by the *fsr* locus regulates the transcription of *gelE* and *sprE*, and it has been shown a reduction in biofilm formation when particular genes of the *fsr* locus are mutated (DebRoy *et al.*, 2014; Mylonakis *et al.*, 2002; Qin *et al.*, 2000; Singh *et al.*, 2005). Moreover, GelE directly cleaves the cell-surface protein Ace (collagen- and laminin-adhesin from *E. faecalis*), affecting the ability of *E. faecalis* to adhere to collagen (Pinkston *et al.*, 2011).

The attachment to proteins of the host extracellular matrix is an important step during *E. faecalis* infection. The Ace protein belongs to the family of microbial surface components recognizing adhesive matrix molecules (MSCRAMM), and it has been implicated in the adherence to host tissues and in the pathogenesis of endocarditis (Rich *et al.*, 1999; Singh *et al.*, 2010). Another protein related with the promotion of adhesion, colonization and evasion of the immune system is Esp (extracellular surface protein), which is a cell-wall-associated protein (Shankar *et al.*, 1999; Waar *et al.*, 2002). This protein contributes to enterococcal biofilm formation, as well as to adhesion to eukaryotic cells (Borgmann *et al.*, 2004; Tendolkar *et al.*, 2004). When the *esp* gene is disrupted, the ability of *E. faecalis* to form biofilms is impaired (Heikens *et al.*, 2007; Toledo-Arana *et al.*, 2001). The aggregation substance proteins (AS proteins) are surface-localized proteins encoded by pheromone-responsive conjugative plasmids. These proteins have been associated with *E. faecalis* binding to cultured renal epithelial cells, internalization by intestinal cells, and evasion of killing by phagocytic cells. Besides, they were shown to increase the virulence of *E. faecalis* in rabbit models of infective endocarditis (Chow *et al.*, 1993; Kreft *et al.*, 1992; Olmsted *et al.*, 1994).

There is a connection between the infection and colonization process and the enterococcal ability to produce biofilms. Moreover, the formation of pili is known to be necessary for biofilm development. One of the gene clusters implicated in the formation of pili is the *ebpABC*

operon (endocarditis- and biofilm-associated pili), which consists of three genes. This operon and an adjacent sortase-encoding gene (*bps*, biofilm and pilus-associated sortase) are essential for the assembly of pillion on the surface of *E. faecalis*. Protein Bps is also known as sortase C (SrtC) (Nallapareddy *et al.*, 2006). Several studies have shown that the production of biofilms is fundamental in endodontic and urinary tract infections, as well as endocarditis; likewise, experiments with a non-piliated mutant of *E. faecalis* have demonstrated its inability to produce biofilms (Budzik and Schneewind, 2006; Nallapareddy *et al.*, 2006; Singh *et al.*, 2007).

Other molecules that have an important role in pathogenesis are the polysaccharides of the Gram-positive cell surface, which are able to mediate evasion of the host innate immune system. Specifically, some *E. faecalis* strains have a capsular polysaccharide locus (*cps*), containing 8-9 genes (Hancock and Gilmore, 2002). Capsule-producing strains of the serotypes C and D have the *cps* genes and it has been shown that they are more resistant to complement-mediated opsonophagocytosis than unencapsulated strains (Thurlow *et al.*, 2009). Another important polysaccharide associated with the cell wall is the enterococcal polysaccharide antigen (Epa), which is a rhamnose-containing polysaccharide widespread among *E. faecalis* strains. The *epa* locus (16 genes) appears to be involved in the biosynthesis of Epa (Teng *et al.*, 2009). Different experiments with Epa mutants have shown they have an attenuated virulence phenotype, a reduced-capacity to evade the immune system, and a deficiency in biofilm formation (Singh *et al.*, 2009; Teng *et al.*, 2002; Xu *et al.*, 2000).

Streptococcus pneumoniae

Another opportunistic bacterium present in the human body is *S. pneumoniae* (the pneumococcus). *S. pneumoniae* is a low-GC content Gram-positive, α-hemolytic, catalase-negative, facultatively anaerobic bacterium, which grows most often as "lancet-shaped" diplococci but it may also appear as a single coccus or in short chains (Bridy-Pappas *et al.*, 2005). *S. pneumoniae* colonizes asymptomatically the upper respiratory tract and is part of the normal nasopharyngeal and throat microbiota of healthy individuals (Sorensen and Edgar, 2018). However, pneumococci can invade and cause mucosal and upper respiratory tract infections (sinusitis and otitis media), lower respiratory tract infections (lobar pneumonia), and severe and invasive infections (bacteremia, sepsis, bacteremic pneumonia, and meningitis) (Bridy-Pappas *et al.*, 2005; Sorensen and Edgar, 2018; Steel *et al.*, 2013). Moreover, this opportunistic pathogen takes advantage of the hosts with underdeveloped, weakened, and/or deteriorating immune systems, causing illness in young children, the elderly, and individuals with underlying medical conditions. Furthermore, certain ethnic groups, lifestyle factors, immunodeficiency states such as asplenia, HIV or sickle cell disease, lower socio-economic

status, people under immunosuppressive therapies, among others, are factors that appear to increase the risk for disease (Bogaert *et al.*, 2004; Bridy-Pappas *et al.*, 2005; Butler and Schuchat, 1999; Chen *et al.*, 1998; Hofmann *et al.*, 1995; Pastor *et al.*, 1998; Redd *et al.*, 1990; Steel *et al.*, 2013). The transmission of this pathogen occurs through direct contact with respiratory secretions from an infected individual, and also through environmental surfaces owing to its ability to form biofilms and survive desiccation for many days (retaining its infectivity) (Bogaert *et al.*, 2004; Marks *et al.*, 2014; Walsh and Camilli, 2011).

In 2017, the World Health Organization (WHO) included *S. pneumoniae* as one of 12 priority pathogens, due to its ability (i) to shift from commensal to pathogen, (ii) to remodel its genome (genetic adaptability), (iii) to spread antibiotic resistance, and (iv) to evade the vaccine-induced immunity, mostly through the uptake and incorporation of exogenous DNA by natural competence (Weiser *et al.*, 2018). *S. pneumoniae* is grouped into 97 serotypes defined by the structure of its capsular polysaccharides (CPS), which are the major virulence factor (Bridy-Pappas *et al.*, 2005; Brooks and Mias, 2018; Geno *et al.*, 2015). Both the degree of pathogenicity and the ability to develop antibiotic resistance are different in each serotype. Only 23 of these serotypes cause most of the infections at all ages (Sorensen and Edgar, 2018). The polysaccharide capsule not only helps bacterial cells penetrate the mucus layer overlying mucosal epithelia but also protects the pathogen by inhibiting neutrophil phagocytosis and classic complement-mediated bacterial killing (Fine, 1975; Hostetter, 1986; Hyams *et al.*, 2010; Loughran *et al.*, 2019). Although non-encapsulated pneumococci can tolerate desiccation (not polysaccharide capsule-dependent), the ability of non-encapsulated pneumococci to cause invasive infections is low (Walsh and Camilli, 2011; Weiser *et al.*, 2018). Antibodies to CPS are highly opsonic and protective against subsequent pneumococcal challenges with the same serotype, and for these reasons they are the basis of the current pneumococcal vaccines (Loughran *et al.*, 2019).

The wide genetic diversity of *S. pneumoniae* together with its capacity to use a broad range of carbohydrates as a carbon source, such as mannose, galactose, and hyaluronic acid, among others, enable this bacterium to persist on the mucosal surface and to adapt readily to new environments niches when translocation from the nasopharynx occurs (Bidossi *et al.*, 2012; Forteza *et al.*, 2001). More than 30 carbohydrate-specific phosphotransferase systems (PTS) and ABC transporters capable of importing both a wide range of sugars and metal ions have been identified (Bidossi *et al.*, 2012). The uptake of metal ions, such as iron (Fe), manganese (Mn) and zinc (Zn), which are essential cofactors for many enzymes, is crucial for the growth and survival of *S. pneumoniae* in different niches. Interestingly, genes encoding components of these transporters are preferentially expressed in the host environment, where the availability

of free ions may be restricted (Berry and Paton, 1996; Brown *et al.*, 2001; Dintilhac *et al.*, 1997; Honsa *et al.*, 2013; Marra *et al.*, 2002; Rosch *et al.*, 2009; Weiser *et al.*, 2018).

In *S. pneumoniae*, the main characteristics that influence the colonization of a new niche are the adherence to host cells, the inherent invasive properties of each serotype, the ability of the bacterium to evade the host immune system, and the absence of serotype-specific antibodies (Bridy-Pappas *et al.*, 2005). Apart from the polysaccharide capsule, other virulence factors that contribute to the *S. pneumoniae* colonization are adhesins. For instance, pneumococci have spontaneous phase variation showing either a transparent colony phenotype or an opaque colony phenotype. The transparent phenotype, which is predominant in the nasopharynx, correlates with increased amounts of two adhesins, phosphorylcholine (ChoP) and choline-binding protein A (CbpA), while the opaque phenotype, which is prevalent in blood samples, is associated with increased levels of CPS and pneumococcal surface protein A (PspA) (Kim *et al.*, 1999; Kim and Weiser, 1998; Rosenow *et al.*, 1997; Weiser *et al.*, 1994). Experiments with strains lacking CbpA showed that they are unable to bind to the nasopharynx and, in addition, they have a reduced capacity to colonize the lower respiratory tract and cause pneumonia (Orihuela *et al.*, 2004; Rosenow *et al.*, 1997).

Nowadays, a large number of molecules contributing to pneumococcal virulence have been identified and tested. In addition to those mentioned above, the pili structures located on the cell surface play an important role. These structures help *S. pneumoniae* escape from phagocytosis and facilitate its adherence to epithelial cells within the nasopharynx. Moreover, the pili structures contribute to the production of biofilms, which increases the pneumococcal resistance to antimicrobial peptides and promotes its ability to share genetic information among bacteria (Barocchi *et al.*, 2006; Loughran *et al.*, 2019; Marks *et al.*, 2012). Additional virulence factors of *S. pneumoniae* are: (i) cell wall components (such as peptidoglycan and teichoic acids), (ii) pneumolysin Ply (pore-forming cytoplasmic toxin which causes cell lysis, promotes biofilm formation, and interferes with the host immune system), (iii) hydrogen peroxide (pneumococcal metabolism product that causes damage to host DNA and induces an innate immune response), (iv) autolysin (intracellular enzyme involved in autolysis which results in the release of pneumolysin, teichoic acids, and peptidoglycan), and (v) other pneumococcal surface proteins (*i.e.* choline-binding proteins, lipoproteins, non-classical surface proteins, and LPXTG cell wall bound protein) (Brooks and Mias, 2018; Loughran *et al.*, 2019).

2. GLOBAL REGULATION OF GENE EXPRESSION IN PATHOGENIC BACTERIA

Bacteria are constantly interacting with changeable environments, where they must be able to survive and to adapt. Mechanisms for bacterial adaptation in response to a variety of stimuli, including those related to cell stress and nutrient availability, can be divided into four principal groups: (i) *Quorum-sensing* systems, involved in cell-cell signalling and population density response; (ii) *Stand-alone* response regulators, which are comprised of signal recognition and DNA binding modules; (iii) Regulatory RNAs, which play a significant role in transcriptional and translational regulation in both eukaryotes and bacteria; and (iv) Two-component systems (TCSs), which are the most widespread regulatory systems in bacteria, and have been also found in archaea, fungi, yeast, and some plants (Cook and Federle, 2014; Dersch *et al.*, 2017; Gomez-Mejia *et al.*, 2018; McIver, 2009).

Despite the number of virulence determinants that have been identified in both *E. faecalis* and *S. pneumoniae*, the regulatory mechanisms involved in their adaptation and pathogenicity are still poorly understood. In general, the adaptation process requires coordinated control of gene expression, where global response regulators may play a key role. Global response regulators can activate and/or repress, directly or indirectly, the transcription of multiple genes in response to specific environmental signals. There is an interesting family of global response regulators that has been identified in Gram-positive bacteria: the Mga/AtxA family. This family includes the virulence regulators AtxA (from *Bacillus anthracis*), Mga (from *S. pyogenes*), Mga*Spn* (from *S. pneumoniae*) and MafR (from *E. faecalis*) (Hammerstrom *et al.*, 2015; Hondorp *et al.*, 2013; Ruiz-Cruz *et al.*, 2016; Solano-Collado *et al.*, 2012). Initially, Mga was defined as a *stand-alone* response regulator due to the absence of a membrane-bound sensor histidine kinase associated with it (McIver, 2009). Further work confirmed that members of the Mga/AtxA family have integrated a specific PTS-recognized phosphorylation domain (PRD domain) (see below).

2.1. Two-component signal transduction systems

A prototypical TCS consists of a transmembrane sensor histidine kinase and a cytoplasmic response regulator, which is usually a transcriptional regulator with nucleic acid binding activity. When a histidine kinase senses an environmental cue, it is auto-phosphorylated at a conserved histidine using intracellular ATP. Subsequently, the phosphoryl group is transferred to a conserved aspartate on the cognate response regulator; this phosphorylation induces conformational changes leading to the activation of the regulator. The activated regulator is

then able to activate or repress gene expression either at the transcriptional level (in most of the reported cases) or at the post-transcriptional level (Figure 2). The system requires divalent metal ions, such as Mg^{2+}, and can be reset by the dephosphorylation of the response regulator (Desai and Kenney, 2017; Stock *et al.*, 2000). More elaborate versions of the TCS exist, such as multiple histidine kinases phosphorylating the same response regulator, or a single histidine kinase controlling several response regulators (Stock *et al.*, 2000).

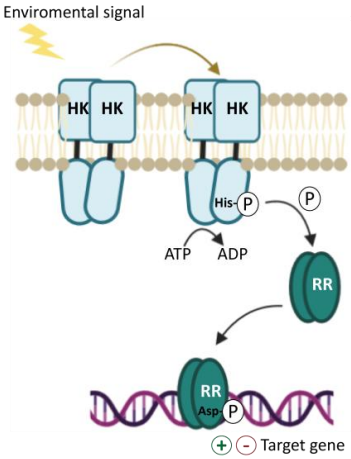


Figure 2. Scheme of a two-component signal transduction system (TCS). The environmental signal is detected by the N-terminal domain of the sensor histidine kinase (HK), leading to the activation of the kinase domain, which autophosphorylates from ATP at a conserved histidine residue. The phosphorylated sensor kinase transfers the phosphoryl group to a conserved aspartate residue of its cognate response regulator (RR). The phosphorylated RR is active, and it can perform a specific biochemical function, such as transcriptional regulation. Figure adapted from Figure 1 in (Solano-Collado, 2014).

S. pneumoniae has 13 complete TCSs and most of them are necessary for pneumococcal pathogenicity in animal models (Gomez-Mejia *et al.*, 2018; Throup *et al.*, 2000). However, the precise function for most of the TCSs in largely unknown. One of the best-characterized pneumococcal TCSs is ComDE, which uses a *quorum-sensing* mechanism in tandem with the ABC transporter ComAB to sense, activate, and regulate the pneumococcal natural transformation machinery. The environmental signal is the competence signalling peptide ComC (CSP), which is cleaved and secreted into the extracellular media by the ComAB transporter. The histidine kinase ComD is autophosphorylated when it detects the extracellular accumulated-CSP. The response regulator ComE is recruited and receives the phosphate group from ComD, therefore activating the expression of approximately 20 early competence genes. Some of these early competence genes participate in the expression of approximately 80 late competence genes (Gomez-Mejia *et al.*, 2018). As summarized recently in Wang *et al.* (2020), in addition to ComDE there are several pneumococcal TCSs whose cellular function has been studied: WalR/K is associated with cell wall peptidoglycan metabolism; LiaR/S senses cell wall

damages and inhibits autolysis; PnpR/S regulates the uptake of inorganic phosphate; CiaR/H inhibits the competence state; CbpR/S activates expression of choline-binding protein A; TCS08 is involved in the cellobiose uptake and in pilus-1 production; and BlpR/H regulates the production of bacteriocins (Wang *et al.*, 2020).

Similar to *S. pneumoniae*, the ability of *E. faecalis* to rapidly respond to environmental changes is achieved by controlling gene expression. Using a genomic-based approach, Hancock and Perego (2002) identified 17 TCSs in strain V583 (Hancock and Perego, 2002). Further mutational analysis showed that some of them are related to antibiotic resistance and environmental stress (Hancock and Perego, 2004). The mode of function of several TCSs has been investigated in detail, including the Fsr *quorum-sensing* system. The *fsr* gene cluster consists of three genes: *fsrA, fsrB* and *fsrC*. FsrA (response regulator) and FsrC (histidine kinase) constitute the FsrC/A TCS. Translation of *fsrB* was shown to generate FsrB and a short peptide (FsrD, C-terminus of FsrB) (Nakayama *et al.*, 2006). FsrB is involved in the processing and secretion of the signal peptide gelatinase biosynthesis-activating pheromone (GBAP) from FsrD. At high cell density, GBAP triggers the activation of the FsrC/A system. Then, the phosphorylated form of the FsrA regulator activates the transcription of *fsrBC*, as well as the transcription of *gelE* (gelatinase) and *sprE* (serine protease) that encode pathogenicity-related extracellular proteases (Nakayama *et al.*, 2001; Qin *et al.*, 2000; Qin *et al.*, 2001).

2.2. The Mga/AtxA family of global response regulators

AtxA FROM *Bacillus anthracis*

Bacillus anthracis is a Gram-positive, spore-forming bacterium. It is the causative agent of the anthrax disease, which primarily affects herbivores. However, the disease also affects a wide range of mammals, including humans. When the spores enter a mammalian host by cutaneous, gastrointestinal, or inhalation routes, they germinate to produce vegetative bacteria. Human can be infected by exposition to *B. anthracis* spores, as well as by contact with infected animals or contaminated animal products. *B. anthracis* pathogenesis is associated with the production of two major virulence factors, the tripartite toxin and the capsule. The anthrax toxin is composed of three proteins, known as protective antigen (PA), lethal factor (LF) and edema factor (EF), which are encoded by the *pagA*, *lef* and *cya* genes, respectively. LF in combination with PA causes cell death, while EF in combination with PA induces edema. The three genes, *pagA*, *lef* and *cya* are located on the pXO1 virulence plasmid, which also encodes the master virulence regulator anthrax toxin activator (AtxA). There is another virulence plasmid, pXO2, which encodes the capsule biosynthesis operon *capBCADE*.

AtxA positively controls the expression of genes located in pXO1 and pXO2, including the anthrax toxin genes and the capsule synthesis operon, but also affects the expression of other chromosomal genes associated with sporulation, branched-chain amino acid biosynthesis and transport and metabolic networks (Dale *et al.*, 2018; Raynor *et al.*, 2018; Uchida *et al.*, 1993). AtxA regulated-promoters do not seem to have DNA sequence similarities. Nevertheless, *in silico* modelling and *in vitro* experiments using several target gene promoter regions have shown a high degree of intrinsic curvature immediately upstream of the transcription start site, which suggests that AtxA could recognize the structural DNA topology instead of a nucleotide sequence (Hadjifrangiskou and Koehler, 2008). It has been shown that the transcription of the *atxA* gene is influenced by glucose availability via carbon catabolite repression, and the toxin synthesis is higher when the bacterium grows in a glucose-containing medium (Bier *et al.*, 2020). Furthermore, *atxA* transcription and AtxA accumulation are enhanced at 37°C compared to 28°C (Dai and Koehler, 1997). Apart from temperature, *atxA* gene expression is affected by carbohydrate availability, growth phase and redox conditions. Moreover, expression of *atxA* is regulated by the AbrB protein, a DNA-binding global regulator that binds directly to the promoter of the *atxA* gene repressing its transcription during the exponential growth phase (Saile and Koehler, 2002; Strauch *et al.*, 2005). Additionally, the regulator CodY seems to control the intracellular accumulation of AtxA post-translationally (van Schaik *et al.*, 2009). A key signal associated with AtxA-regulated gene expression is CO_2/bicarbonate. Albeit the relationship between AtxA and CO_2/bicarbonate remains unclear, it has been demonstrated that cultures grown with elevated CO_2/bicarbonate have higher levels of AtxA dimer compared to cultures grown in air, as well as higher expression of AtxA target genes (Hammerstrom *et al.*, 2011). Also, a recent study has shown that CO_2 does not influence the binding of AtxA to the promoter region of the *pagA* gene (McCall *et al.*, 2019).

AtxA is a 56 kDa protein (475 residues) usually existing as a homodimer (Hammerstrom *et al.*, 2011), which was corroborated by the publication of the crystal structure and by sedimentation velocity experiments (Hammerstrom *et al.*, 2015; McCall *et al.*, 2019). Formation of AtxA homodimers seems to be required for virulence gene activation (Hammerstrom *et al.*, 2011). Each AtxA monomer comprises five domains: two helix-turn-helix domains (HTH domains), two phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) regulation domains (PRD domains), and one Enzyme IIB (EIIB)-like domain (Figure 3) (Hammerstrom *et al.*, 2015). The HTH domains are found at the N-terminal region and have structures indicative of DNA-binding: HTH1 is a winged helix-turn-helix motif (residues 7 to 74); and HTH2, is a helix-turn-helix motif (residues 76 to 162) (Hammerstrom *et al.*, 2015). Next two domains are PRDs: PRD1 (residues 170 to 262) and PRD2 (residues 276 to 391). PRDs have usually been found in transcriptional antiterminators and transcriptional activators, whose

activity is modulated by PTS-mediated phosphorylation in response to the availability of carbon sources (Deutscher *et al.*, 2014). The PTS system can transfer a phosphate to a specific histidine residue of the PRD-containing regulators, therefore activating or inhibiting their activity (Deutscher *et al.*, 2014). AtxA phosphorylation at H199 in PRD1 facilitates its function, while phosphorylation at H379 in PRD2 impairs its dimerization, which is necessary for the activation of the protein (Hammerstrom *et al.*, 2011; Tsvetanova *et al.*, 2007). A recent study has revealed that the PTS proteins Hpr and Enzyme I (EI) regulate *atxA* gene transcription rather than AtxA protein activity (Bier *et al.*, 2020). At the C-terminal region of AtxA is found a domain similar to an EIIB-like motif (residues 391 to 475) (Hammerstrom *et al.*, 2015). According to Hammerstrom *et al*., this EIIB-like motif is involved in AtxA multimerization (Hammerstrom *et al.*, 2011).

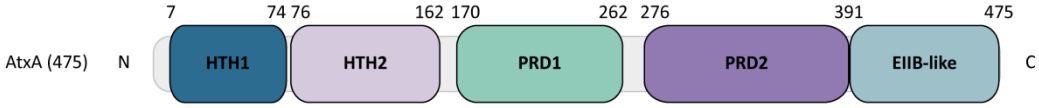


Figure 3. Organization of functional domains in AtxA. AtxA has two N-terminal DNA-binding domains (HTH1 and HTH2). The two central domains are PRDs (PRD1 and PRD2), and the C-terminal domain resembles an EIIB-like domain (Hammerstrom *et al.*, 2015).

Mga FROM *Streptococcus pyogenes*

S. pyogenes (Group A Streptococcus, GAS) is a beta-hemolytic, Gram-positive and strict human pathogen. Due to its ability to colonize a variety of tissues (throat or skin colonization), *S. pyogenes* causes life-threatening invasive disorders, such as necrotizing fasciitis (infection in fascia), Streptococcal Toxic Shock Syndrome and myonecrosis (infection in muscle), as well as self-limiting infections, such as pharyngitis, or immune sequelae (acute rheumatic fever) (Cunningham, 2000; McIver, 2009; Stevens and Bryant, 2016). Besides, *S. pyogenes* infection causes impetigo (superficial keratin layer), ecthyma, erysipelas (superficial epidermis), cellulitis (subcutaneous tissue), and myositis (Stevens and Bryant, 2016). Taken into account all the aforementioned information, an interesting feature of GAS is its ability to adapt and persist at many different tissue sites in the human host, which implicates a coordinate regulation of virulence gene expression. In contrast to other Gram-positive pathogens that use alternative sigma factors to regulate virulence gene expression during stress or growth phase, GAS seems to depend on global transcriptional regulators (McIver, 2009). One of them is Mga (multiple gene regulator of GAS), which was the first stand-alone regulator described in GAS, and controls the expression of approximately 10% of the genome during the exponential growth phase (Kreikemeyer *et al.*, 2003; McIver, 2009; McIver and Scott, 1997; Ribardo and McIver,

2006). The transcriptional activator Mga has been found in all serotypes of GAS (Kreikemeyer *et al.*, 2003) and two divergent alleles, *mga-1* and *mga-2*, associated with serum opacity factor (SOF positive or negative) have been identified (Bessen *et al.*, 1989).

Mga has a critical effect on biofilm formation, growth in whole human blood and soft tissue, adherence to host tissues (attachment to keratinocytes), resistance to phagocytosis, internalization into non-phagocytic cells, and immune evasion (Cunningham, 2000; McIver, 2009; McIver *et al.*, 1995; Perez-Casal *et al.*, 1995). The Mga-mediated control of these processes is explained by the Mga-activated regulon, which includes the virulence genes *emm* (M protein), *scpA* (cell wall-associated C5a peptidase), *sclA* (streptococcal collagen-like protein A), *fba* (fibronectin-binding proteins), *sof* (serum opacity factor) and *sic* (streptococcal inhibitor of complement) (Cunningham, 2000; McIver, 2009; McIver *et al.*, 1995; Reid *et al.*, 2001; Ribardo and McIver, 2006; Simpson *et al.*, 1990).

Additionally to the core virulence genes, Mga activates or represses (most likely indirectly) the expression of genes involved in the transport and utilization of carbohydrates as well as other metabolites such as iron and amino acids (Ribardo and McIver, 2006). Furthermore, the expression of *mga* is directly autoregulated, resulting in amplification of regulon expression (McIver *et al.*, 1999; Okada *et al.*, 1993). Thus, Mga is able to regulate not only genes involved in virulence, but also important genes for the metabolic homeostasis of GAS. Apart from the exponential growth phase, the Mga regulon has been linked to elevated CO_2, normal body temperature, increased iron levels and metabolizable sugars (Hondorp and McIver, 2007; McIver, 2009).

The interaction of Mga with its target genes has been widely studied by EMSA and DNase I protection experiments, which revealed that Mga binds to a region located upstream of its target promoters. Such assays were performed using Mga fused to a His-tag. Three categories of Mga-activated promoters have been proposed based on the differences observed in the location of Mga-binding sites and the transcription start of those promoters: class A promoters (single proximal binding site for Mga); class B promoters (single distal binding site for Mga); and class C promoters (have multiple distal sites) (Almengor and McIver, 2004). Sequence alignments of the Mga-binding sites revealed that they exhibit a very low sequence identity (Hause and McIver, 2012). Moreover, a mutational study suggested that Mga binds to DNA in a promoter-specific manner (Hause and McIver, 2012). The transcriptional regulation mechanism mediated by Mga remains unknown.

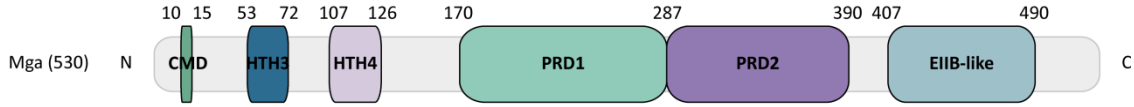


Figure 4. Organization of functional domains in Mga. Mga contains two helix-turn-helix motifs (HTH3 and HTH4) and a conserved region (CMD) at the N-terminus. The central region has two putative PRDs (PRD1 and PRD2). The C-terminal region contains a PTS EIIB-like domain (Hondorp *et al.*, 2013).

Mga is a 62 kDa protein (530 amino acids) which forms oligomers in solution. Its ability to oligomerize *in vitro* correlates with its capacity to activate transcription *in vivo* (Hondorp *et al.*, 2012). Some studies have predicted that Mga comprises six domains (Hondorp *et al.*, 2013). At the N-terminal region, there are three domains involved in DNA-binding activity and transcriptional activation: a conserved Mga domain (CMD, residues 10 to 15), a classical helix-turn-helix motif (HTH3, residues 53 to 72) and a winged-helix-turn-helix motif (HTH4, residues 107 to 126) (Figure 4). The HTH4 motif is necessary for binding to all Mga-binding sites tested so far (McIver and Myles, 2002; Vahling and McIver, 2006). In the central region of Mga, there are two PTS regulation domains, PRDs (PRD1, residues 170 to 287; PRD2, residues 288 to 390) (Figure 4). The C-terminal region of Mga contains an Enzyme IIB-like domain (EIIB-like, residues 407 to 490) (Figure 4), which seems to be important, together with the PRDs, for oligomerization *in vitro* as well as for transcriptional activation *in vivo* (Hondorp *et al.*, 2012). It has been shown that GAS PTS system can directly phosphorylate conserved PRD histidines in Mga, and this phosphorylation appears to modulate (both positively and negatively) the activity of Mga (Hondorp *et al.*, 2013). Moreover, it has been shown that Mga is differentially phosphorylated in the presence or absence of glucose; thus, glucose directly affects Mga activity and the composition of the Mga regulon (Valdes *et al.*, 2018). It has been reported that the conserved histidines within the PRDs of Mga are phosphorylated by the PTS, leading to a defect in protein oligomerization, altered gene expression, and attenuation of virulence in a mouse model of GAS infection (Hondorp *et al.*, 2013). The influence of PTS on GAS pathogenesis has revealed the GAS potential to use specific carbon metabolic pathways in the host niches (Sundar *et al.*, 2017).

Mga*Spn* FROM *Streptococcus pneumoniae*

Studies on the pathogenicity of *S. pneumoniae* have led to the identification of virulence genes, as well as to the identification of genes involved in the adaptation of this bacterium to changing environments. By signature-tagged mutagenesis in the pneumococcal TIGR4 strain, the *sp1800* gene was initially identified as a potential virulence gene (Hava and Camilli, 2002). This gene encodes a protein that was first named MgrA (Mga-like repressor A), owing to its homology to Mga of GAS (Hemsley *et al.*, 2003), and later on, it was named Mga*Spn* (Solano-

Collado *et al.*, 2012). Mga*Spn* (493 residues) encoded by the R6 genome (gene *spr1622*) differs from Mga*Spn* of TIGR4 in two amino acid residues; and exhibits a similarity/identity of 38.6/23.7%, 39.9/20.7% and 42.6/21.4% with MafR (482 residues; strain OG1RF), AtxA (475 residues; strain Ames Ancestor) and Mga (530 residues; strain MGAS10394), respectively (Table 1), according to EMBOSS needle global sequence alignment (Rice *et al.*, 2000). It was shown that Mga*Spn* plays a significant role in lung infection and nasopharyngeal colonization in a murine model (Hemsley *et al.*, 2003). Furthermore, Mga*Spn* acts as a transcriptional repressor of the *rlrA* pathogenicity islet genes (Hemsley *et al.*, 2003). In contrast to the *mgaSpn* gene, which is highly conserved among the pneumococcal genomes, the *rlrA* pathogenicity islet has been found in a small number of pneumococcal strains. For instance, strain R6 lacks the *rlrA* islet (Paterson and Mitchell, 2006; Solano-Collado *et al.*, 2012).

Mga*Spn* is a 59 kDa protein which has been confirmed to form dimers in solution by gel filtration chromatography experiments (Solano-Collado *et al.*, 2013) and analytical ultracentrifugation assays (Solano-Collado, 2014). The predicted organization of functional domains is similar to that found in the global regulators of the Mga/AtxA family. Mga*Spn* has two putative N-terminal helix-turn-helix DNA-binding domains, the so-called HTH_Mga (residues 6 to 65) and Mga (residues 71 to 158) domains (Figure 5) (Solano-Collado *et al.*, 2012). Moreover, the central region has a putative PRD domain (residues 173 to 392), and the C-terminal region contains a putative EIIB-like domain (amino acids 399 to 487) (Figure 5) (Solano-Collado *et al.*, 2016).

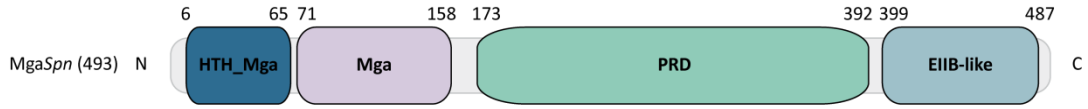


Figure 5. Organization of functional domains in Mga*Spn*. Mga*Spn* has two putative DNA-binding domains at the N-terminal region (HTH_Mga and Mga). The central region has structural homology to a PRD and the C-terminal region shows structural homology to an EIIB-like domain (Solano-Collado, 2014).

Working with the TIGR4 strain, Hemsley *et al.* reported that Mga*Spn* did not affect transcription of a neighbouring cluster of genes (Hemsley *et al.*, 2003). However, further studies using the R6 strain demonstrated that Mga*Spn* activates directly the transcription of an adjacent four-gene operon of unknown function (Solano-Collado *et al.*, 2012). Transcription of this operon is under the control of two promoters, *P1623A* and *P1623B*, which are divergent from the promoter of the *mgaSpn* gene. Mga*Spn* was shown to activate the *P1623B* promoter *in vivo*. This activation requires sequences located upstream of the target promoter (Solano-Collado *et al.*, 2012), and such sequences are recognized by Mga*Spn in vitro* (Solano-Collado *et*

al., 2013). Moreover, by gel retardation, footprinting and electron microscopy techniques, it has been demonstrated that an untagged form of Mga*Spn* interacts with linear double-stranded DNAs. In this interaction, Mga*Spn* binds to a specific site (its primary site) and then Mga*Spn* is able to spread along the adjacent DNA regions, generating multimeric protein-DNA complexes (Solano-Collado *et al.*, 2013). It has been reported that *MgaSpn* binds to DNA with little or no sequence specificity and shows a preference for DNA regions that contain a potential intrinsic curvature, thus, local DNA conformations might contribute to the DNA-binding specificity of Mga*Spn* (Solano-Collado *et al.*, 2013). Furthermore, subsequent experiments with Mga*Spn* and H-NS, an *E. coli* protein which binds to DNA in a non-specific manner but has a strong preference for intrinsically curved AT-rich DNA regions, showed that Mga*Spn* is able to recognize particular regions on extended H-NS binding sites and *vice versa* (Solano-Collado *et al.*, 2016). Thus, these studies support the hypothesis that Mga*Spn* might recognize structural characteristics in their DNA targets (Solano-Collado *et al.*, 2016).

MafR FROM *Enterococcus faecalis*

The pathogenicity of *E. faecalis* could be a consequence of the adaptation of the bacterium to particular host niches. This adaptation requires a coordinated regulation in gene expression that could be modulated by global response regulators. The limited understanding of the regulatory mechanisms involved in the pathogenicity of *E. faecalis*, together with the publication of the complete genome sequence of *E. faecalis* strain V583, led to an increased interest in the identification of new virulence regulators. In 2003, Paulsen *et al.* published the genome sequence of V583, the first vancomycin-resistant clinical isolate reported in the United States (Paulsen *et al.*, 2003). This article revealed that over a quarter of the V583 genome consists of mobile and/or exogenously acquired DNA, which is one of the highest proportions observed in a bacterial genome (Paulsen *et al.*, 2003). This includes probable integrated phage regions, insertion elements, conjugative transposons, a putative pathogenicity island, and integrated plasmid genes (Paulsen *et al.*, 2003). Only five years later, the genome sequence of *E. faecalis* strain OG1RF, a rifampicin- and fusidic acid-resistant derivative of the OG1 human isolate, was published (Bourgogne *et al.*, 2008). Plamid-free strain OG1RF is widely used in the laboratories as it has fewer mobile genetic elements and does not present resistance to common-used antibiotics (Bourgogne *et al.*, 2008). Moreover, comparing to strain V583, the OG1RF genome contains 227 unique open reading frames (Bourgogne *et al.*, 2008). Searching for homologies, Ruiz-Cruz *et al.* found that the *EF3013* gene from strain V583 could be a regulatory gene. It encodes a 482 residues protein that was first named MAEfa (Ruiz-Cruz, 2015), and later on MafR (Mga/AtxA-like *faecalis* regulator) (Ruiz-Cruz *et al.*, 2016). At present, MafR is known to be a member of the Mga/AtxA family of global response regulators (see

below). According to EMBOSS needle global sequence alignment (Rice *et al.*, 2000), MafR exhibits a similarity/identity of 40.7/22.1%, 31.3/19.6% and 38.8/23.7% with AtxA (475 residues; strain Ames Ancestor), Mga (530 residues; strain MGAS10394) and Mga*Spn* (493 residues; strain R6), respectively (Table 1).

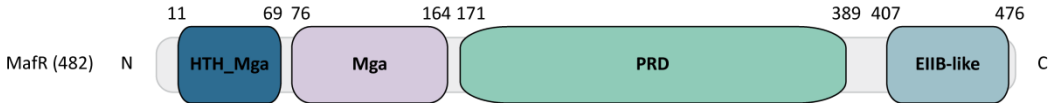


Figure 6. Organization of functional domains in MafR. MafR has two putative DNA-binding domains at the N-terminal region (HTH_Mga and Mga). The central region has structural homology to a PRD and the C-terminal region shows structural homology to an EIIB-like motif (Ruiz-Cruz, 2015).

In MafR (56.2 kDa), the predicted organization of functional domains is similar to that found in the global regulators of the Mga/AtxA family (Figure 6). MafR has two putative DNA-binding domains within the N-terminal region, the so-called HTH_Mga (residues 11 to 69) and Mga (residues 76 to 164) domains (Ruiz-Cruz *et al.*, 2016). The central region of MafR has a putative PRD domain (residues 171 to 389) (Ruiz-Cruz, 2015). Moreover, the three-dimensional structures of MafR (EF3013; PDB 3SQN; (Osipiuk *et al.*, unpublished results)) and AtxA (PDB 4R6I; (Hammerstrom *et al.*, 2015)) revealed that the closest structural homologue for the AtxA C-terminal region (EIIB-like domain) is the C-terminal region of MafR (residues 407 to 476) (Hammerstrom *et al.*, 2015). Protein MafR from strain V583 has been purified (untagged version) and its DNA-binding properties have been studied (Ruiz-Cruz *et al.*, 2018). By gel filtration chromatography and analytical ultracentrifugation experiments, MafR was shown to behave as a dimer in solution, and the frictional ratio calculated from the analytical ultracentrifugation assays indicated that the shape of the MafR dimer is an ellipsoid (Ruiz-Cruz *et al.*, 2018). Furthermore, EMSA experiments using linear dsDNAs and the untagged form of MafR revealed that multiple MafR units (likely dimers) bind sequentially to the same DNA molecule generating multimeric protein-DNA-complexes (Ruiz-Cruz *et al.*, 2018). EMSA experiments also indicated that MafR binds to DNA with little or no sequence specificity, like other regulators of the Mga/AtxA family (Ruiz-Cruz *et al.*, 2018). Further DNase I footprinting experiments using a DNA fragment that contains the promoter of the *mafR* gene (*Pma* promoter) showed that MafR binds preferentially to a region located upstream of the promoter, between positions -69 and -104. Such a MafR binding site is adjacent to the peak of a potential intrinsic curvature (Ruiz-Cruz *et al.*, 2018). The function of this interaction remains unknown, since MafR does not seem to influence the expression of its own gene *in vivo* (Ruiz-Cruz, 2015). The minimum size of DNA required for MafR binding is between 26 bp and 32 bp, while the Mga*Spn* regulator requires between 20 bp and 26 bp (Ruiz-Cruz *et al.*, 2018; Solano-

Collado *et al.*, 2013). A His-tagged MafR protein from strain V583 (MafR-His) has been also purified. This variant of MafR carries the Leu-Glu-6xHis peptide (His-tag) fused to its C-terminus. EMSA experiments revealed that the presence of the His-tag does not affect its ability to generate multimeric complexes on linear dsDNAs (Ruiz-Cruz *et al.*, 2018).

The enterococcal OG1RF genome also encodes the MafR protein (gene *OG1RF_12293*). Microarray experiments, quantitative RT-PCR assays and complementation studies have shown that MafR increases the expression of numerous genes. In such experiments, a *mafR* deletion mutant strain (OG1RFΔ*mafR*), unable to synthesize MafR, and the wild-type strain (OG1RF) were used. In MafR-lacking cells, at least 90 genes were significantly differentially expressed: 87 genes were down-regulated, and 3 genes were up-regulated (Ruiz-Cruz *et al.*, 2016). Hence, MafR activates, directly or indirectly, the expression of at least 87 genes, among them, 15 genes encode components of PTS-type membrane transporters, 18 genes encode enzymes involved in carbon source metabolism, and 9 genes encode components of ABC-type membrane transporters (Ruiz-Cruz *et al.*, 2016). Moreover, it has been demonstrated that enterococcal cells deficient in MafR induce a more moderate inflammatory response in a mouse peritonitis model, suggesting that a *mafR*-lacking strain is less virulent (Ruiz-Cruz *et al.*, 2016).

3. THE PNEUMOCOCCAL MgaP PROTEIN

In 2001, the genome sequence of the pneumococcal TIGR4 strain (serotype 4, clinical isolate) was published, revealing that about 5% of its genome is composed of insertion sequences that may contribute to genome rearrangements through the uptake of foreign DNA (Tettelin *et al.*, 2001). Later, the publication of the genomic sequence of the pneumococcal R6 strain (serotype 2, a derivative of the D39 clinical isolate) revealed that the TIGR4 and R6 genomes differ by about 10% of their genes (Brückner *et al.*, 2004; Hoskins *et al.*, 2001). For instance, unlike TIGR4, both R6 and D39 lack the *rlrA* pathogenicity islet (Lanie *et al.*, 2007). Moreover, the R6 and TIGR4 genomes differ in 18 gene clusters. Specifically, there are 6 clusters in the R6 genome that are not present in the TIGR4 genome (Brückner *et al.*, 2004). One of them is a region of 9.6 kb composed of two divergent genes, which encode a pneumococcal collagen-like protein A (*spr1403*, also named *pclA*) and a putative transcriptional regulator (*spr1404*) (Paterson *et al.*, 2008). Interestingly, by PCR techniques, Paterson *et al*. showed that 18 out of the 46 strains examined were positives for *pclA* and *spr1404*, and the analysed strains were either positive or negative for both genes (Paterson *et al.*, 2008). This cluster is always present in the same location of the genome relative to the flanking genes *spr1402* and *spr1405*

(Paterson *et al.*, 2008). Furthermore, no insertion sequences or phage elements have been found adjacent to the region (Paterson *et al.*, 2008).

Experiments to test the role in virulence of the *pclA* and *spr1404* genes have been reported, showing that they do not contribute significantly to nasopharyngeal colonization in a mouse model (Paterson *et al.*, 2008). However, PclA seems to be involved in the pneumococcal adherence and invasion of host cells (Paterson *et al.*, 2008). Moreover, it has been reported that *spr1404* does not influence the transcription of *pclA* (Paterson *et al.*, 2008). Additional studies have shown that *pclA* was variably distributed among clinical isolates and significantly associated with multilocus sequence typing (MLST). Furthermore, *pclA* correlated with Pneumococcal Molecular Epidemiology Network (PMEN) clones and antimicrobial resistance (Imai *et al.*, 2011; McGee *et al.*, 2001).

Searching for homologies, we found that the putative transcriptional regulator Spr1404 (494 residues) exhibits a 60.3% of sequence similarity to Mga*Spn* (493 residues; *S. pneumoniae* R6 strain), 40.4% to Mga (530 residues; *S. pyogenes* MGAS10394 strain), and about 36% of similarity to the others regulators of the Mga/AtxA family (Table 1). Owing to the high homology (60.3% similarity and 40.1% identity) between Mga*Spn* and Spr1404 (from now on named MgaP, Mga*Spn* Paralog), we proposed that (i) MgaP could be a member of the Mga/AtxA family of transcriptional regulators, and (ii) it might act as a transcriptional activator of its neighbouring genes.

Table 1. Similarity/Identity among the proteins of the Mga/AtxA family of global response regulators

	$MafR_{V583}$	**$AtxA_{Ames\ Ancestor}$**	**$Mga_{MGAS10394}$**	**$Mga\mathit{Spn}_{R6}$**	**$MgaP_{R6}$**
$MafR_{V583}$	–	40.7/22.1%	31.3/19.6%	38.8/23.7%	36.1/20.1%
$AtxA_{Ames\ Ancestor}$	40.7/22.1%	–	39.7/19.5%	39.9/20.7%	36.7/20.7%
$Mga_{MGAS10394}$	31.3/19.6%	39.7/19.5%	–	42.6/21.4%	40.4/19.9%
$Mga\mathit{Spn}_{R6}$	38.8/23.7%	39.9/20.7%	42.6/21.4%	–	60.3/40.1%
$MgaP_{R6}$	36.1/20.1%	36.7/20.7%	40.4/19.9%	60.3/40.1%	–

OBJECTIVES

In general, gene regulation plays an essential role in the ability of bacterial cells to adapt to new environmental conditions. The Gram-positive bacteria *E. faecalis* and *S. pneumoniae* can proliferate in various niches of the human host, either as commensals or as leading causes of serious infections. The main aim of our investigation is to identify and characterize transcriptional regulators involved in the adaptive responses of both bacteria. To this end, we have worked on the following objectives:

1. Analysis of the DNA-binding properties of MafR, a protein that causes genome-wide changes in the transcriptome of *E. faecalis* OG1RF.

2. Identification of genes regulated directly by MafR.

3. Expression analysis of the strain-specific *mgaP* gene (putative transcriptional regulator) in *S. pneumoniae* R6.

4. To study whether MgaP controls the expression of the *pclA* gene (pneumococcal collagen-like protein A).

5. To analyse whether MgaP and Mga*Spn*, a pneumococcal member of the Mga/AtxA family of transcriptional regulators, have common features in their interaction with DNA.

MATERIALS AND METHODS

MATERIALS

1. BACTERIAL STRAINS

Table 2. Bacterial strains

Strain	Features	Source
***E. faecalis* OG1RF**	*gelE*+, *sprE*+, *fsrABCD*+, GelE+, SprE+, Rif^R, Fus^R. Plasmid-free	(Dunny *et al.*, 1978)
***E. faecalis* V583**	*gelE*+, *sprE*+, *fsrABCD*+, GelE+, SprE+; Van^R	(Sahm *et al.*, 1989)
E. faecalis* OG1RFΔ*mafR	Strain derived from OG1RF. It lacks the *mafR* gene	(Ruiz-Cruz *et al.*, 2016)
***S. pneumoniae* R6**	Nonencapsulated strain derived from the serotype 2 clinical isolate D39	(Lacks, 1968)
S. pneumoniae* R6Δ*mga	Strain derived from R6. It lacks the *mgaSpn* gene	(Solano-Collado *et al.*, 2012)
***E. coli* BL21 (DE3)**	λDE3 [*lac*I, *lac*UV5-T7 gene 1, *ind*1, *sam*7, *nin*5) F- *dcm*, *omp*T, *lon*, *hsdS* (rB- mB-), *gal*	(Studier and Moffatt, 1986)

Rif^R, Fus^R, Van^R: resistance to rifampicin, fusidic acid and vancomycin, respectively.

2. CULTURE MEDIA

E. faecalis cells were grown in Brain Heart Infusion (BHI) medium unless otherwise indicated. Enterococcal cells harbouring plasmids were grown in media supplemented with tetracycline (Tc 4 μg/ml) or kanamycin (Km 250 μg/ml). *E. faecalis* competent cells were grown in SG-M17 medium (M17 medium supplemented with 0.5 M sucrose and 5-8% of glycine depending on the strain). After transformation by electroporation, cells were incubated in S-M17-MC medium (M17 medium supplemented with 0.5 M sucrose, 10 mM $MgCl_2$ and 10 mM $CaCl_2$). To select transformant cells on solid media, SR medium (Shepard and Gilmore, 1995) supplemented with the proper antibiotic was used.

S. pneumoniae cells were grown in AGCH medium (Lacks, 1966; Ruiz-Cruz *et al.*, 2010) supplemented with 0.2% yeast extract and 0.3% sucrose. Pneumococcal cells harbouring plasmids were grown in media supplemented with Tc (1 μg/ml) or Km (50 μg/ml). *S. pneumoniae* competent cells were grown in AGCH medium with 0.2% sucrose and 70 μM of $CaCl_2$. Plates for bacterial growth in solid medium were freshly prepared as it is reported in Methods, Section 1.

E. coli cells were grown in tryptone-yeast extract (TY) medium. In cells harbouring plasmid, medium was supplemented with Km (30 µg/ml). Plates for bacterial growth in solid medium were prepared with TY, 1.5% agar and supplemented with the appropriate antibiotic when was required. For competence of *E. coli*, SOB medium (Hanahan, 1983) was used. After transformation, cells were incubated in SOC medium (SOB medium supplemented with 20 mM glucose).

3. ENZYMES, CHEMICAL PRODUCTS AND REACTANTS

Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase and bovine serum albumin (BSA) were from New England BioLabs. DNase I RNase-free, Isopropyl-β-D-thiogalactopyranoside (IPTG), protease inhibitor cocktail tablets complete, High Pure Plasmid Isolation kit, phosphatase alkaline (PA) and deoxyribonucleotide triphosphates (dNTPs) were acquired from Roche Life Science. Taq DNA Polymerase (Roche Life Science and Invitrogen) and Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) were used. ThermoScript reverse transcriptase was purchased to Invitrogen. NZYDNA Ladder III and VI weight markers were from NZYTech. Proteinase K, RNase A, lysozyme, kanamycin (Km), mutanolysin, fucose, sucrose, polyethylenimine (PEI), imidazole, acrylamide:bis-acrylamide 40% solution (29:1 ratio) and dimethyl sulfoxide (DMSO) were provided by Sigma Aldrich (Merck). Rifampicin (Rif) was supplied by Sanofi-Aventis S.A. GelRed was acquired from Biotium. Ethanol absolute, glycerol 85%, ammonium sulfate, ammonium acetate, sodium acetate, hydrochloric acid, acetic acid, isopropanol, amino acids, vitamins, glucose and magnesium were provided by Merck. Phenol and tetracycline (Tc) were acquired from Panreac Applichem. Culture media components were from Pronadisa, Sigma Aldrich (Merck) and Difco-BD. DNA sequencing kit (Sequenase Version 2.0) was provided by USB Corporation. Radioactive nucleotides were purchased to Perkin-Elmer. Decon 90 from Thermo Fisher Scientific was used. Agarose, acrylamide:bis-acrylamide 30% solution (375:1 ration), ammonium persulfate (APS), Dithiothreitol (DTT), β-mercaptoethanol, TEMED (tetramethylethylenediamine), sodium dodecyl sulfate (SDS), Bio-Safe Coomassie Stain, iScript Select cDNA synthesis kit and iQ SYBER Green Supermix were from Bio-Rad. Acrylamide:bis-acrylamide 40% solution (19:1 ratio) was from National Diagnostics. The HisTrap HP columns, MicroSpinTM G-25 columns, Vivaspin 20 for protein concentration and low molecular weight protein marker (LMW) were provided by GE Healthcare. QIAquick Gel Extraction and RNeasy Mini kit were from QIAGEN. Bacterial genomic isolation kit was purchased to Norgen Biotek Corporation. Dialysis membranes were from Spectrum. 96-well plates for fluorescence measurements and nitrocellulose filters of pore size 0.22 and 0.45 µm were from Millipore-Merck. Autoradiography films were provided by Kodak (X-Omat S). Cronex Lightning Plus Intensifying X-ray screens were from Dupont. Imaging plates to visualize radioactive labelling using the Fujifilm Image

Analyzer FLA-3000 were acquired from Fuji. Centrifugal filter Microsep for protein concentration was from Pall. Electroporation cuvettes of 2 mm were from Cell Projects. RNAsin Plus RNAse Inhibitor and DTT were supplied by Promega. During RNA experiments Ambion-RNaseZAP (Thermo Fisher Scientific) was used.

4. NUCLEIC ACIDS

4.1. PLASMIDS

Table 3. Plasmids

Plasmid	Size (bp)	Description	Source
pET-24b	5,309	*E. coli* expression vector based on the Φ10 promoter of phage T7; Km^R	Novagen
pET24b-*mafR*$_{OG1RF}$-His	6,677	Derivative of pET-24b that encodes a His-tagged MafR protein from *E. faecalis* OG1RF; Km^R	This work
pET24b-*mafR*$_{OG1RF}$Δ3N-His	6,668	Derivative of pET-24b that encodes a His-tagged MafRΔ3N protein from *E. faecalis* OG1RF; Km^R	This work
pET24b-*mafR*$_{V583}$-His	6,790	Derivative of pET-24b that encodes a His-tagged MafR protein from *E. faecalis* V583; Km^R	This work
pET24b-*mgaP*-His	6,793	Derivative of pET-24b that encodes a His-tagged MgaP protein from *S. pneumoniae* R6; Km^R	(Ortuno-Camuñas, 2017) and this work
pDL287	5,740	Derivative of pVA380-1; Km^R	(LeBlanc *et al.*, 1993)
pDLF	5,911	Expression vector. Derivative of pDL287 that carries the promoter of *E. faecalis* EF2493 or *cpsC* gene; Km^R	(Ruiz-Cruz *et al.*, 2016)
pDLF*mafR*	7,425	Derivative of pDL287 that carries the *E. faecalis* OG1RF *mafR* gene under the control of the promoter *P2493*; Km^R	(Ruiz-Cruz *et al.*, 2016)
pDLF*mgaP*	7,472	Derivative of pDL287 that carries the *S. pneumoniae* R6 *mgaP* gene under the control of the promoter *P2493*; Km^R	(Ortuno-Camuñas, 2017) and this work

Plasmid	Size (bp)	Description	Source
pDLF*mgaPi*	7,472	Derivative of pDL287 that carries the inverted *mgaP* gene (*mgaPi*), from *S. pneumoniae* R6, under the control of the promoter *P2493*; Km^R	(Ortuno-Camuñas, 2017) and this work
pDL*PsulA*::*mga*	7,529	Derivative of pDL287 that carries the *S. pneumoniae* R6 *mgaSpn* gene under the control of the promoter *PsulA*; Km^R	(Solano-Collado *et al.*, 2012)
pAST	5,456	Promoter-probe vector. Derivative of pAS that carries the transcriptional termination sites *T1T2* of the *E. coli rrnB* ribosomal RNA operon downstream of the *tetL* gene; Tc^R	(Ruiz-Cruz *et al.*, 2010)
pASTT	5,732	Derivative of pAST that carries the *TrsiV* transcriptional terminator downstream of the *gfp* reporter gene; Tc^R	(Ruiz-Cruz *et al.*, 2019)
pASTT derivatives of *E. faecalis*			
pASTT-*P12294*	5,992	Derivative of pASTT that carries the promoter region of the OG1RF_*12294* or *pmr1* gene; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P12294Δ-10*	5,968	Derivative of pASTT that carries the promoter region of the OG1RF_*12294* gene, but the last 24 nucleotides have been removed, including the -10 box; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P12294Δ69*	5,924	Derivative of pASTT that carries the promoter region of the OG1RF_*12294* gene, but the first 69 nucleotides have been removed; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P12294Δ208*	5,785	Derivative of pASTT that carries the promoter region of the OG1RF_*12294* gene, but the first 208 nucleotides have been removed, including the MafR binding region; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P11486*	6,022	Derivative of pASTT that carries the promoter region of the OG1RF_*11486* gene; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work

Plasmid	Size (bp)	Description	Source
pASTT-*P11486Δ66*	5,956	Derivative of pASTT that carries the promoter region of the OG1RF_*11486* gene, but the first 66 nucleotides have been removed; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P11486Δ145*	5,877	Derivative of pASTT that carries the promoter region of the OG1RF_*11486* gene, but the first 145 nucleotides have been removed; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P11486Δ169*	5,853	Derivative of pASTT that carries the promoter region of the OG1RF_*11486* gene, but the first 169 nucleotides have been removed, including the MafR binding region; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P11486Δ188*	5,834	Derivative of pASTT that carries the promoter region of the OG1RF_*11486* gene, but the first 188 nucleotides have been removed, including the MafR binding region and the -35 box; Tc^R	(Ruiz-Cruz *et al.*, 2019) and this work
pASTT-*P10478*	6,033	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene; Tc^R	This work
pASTT-*P10478Δ-10*	6,011	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the last 23 nucleotides have been removed, including the -10 box; Tc^R	This work
pASTT-*P10478Δ122*	5,909	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the first 122 nucleotides have been removed; Tc^R	This work
pASTT-*P10478Δ210*	5,822	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the first 210 nucleotides have been removed; Tc^R	This work
pASTT-*P10478Δ228*	5,805	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the first 228 nucleotides have been removed; Tc^R	This work

Plasmid	Size (bp)	Description	Source
pASTT-*P10478Δ234*	5,799	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the first 234 nucleotides have been removed, including a region of the MafR binding site; Tc^R	This work
pASTT-*P10478Δ243*	5,790	Derivative of pASTT that carries the promoter region of the OG1RF_*10478* gene, but the first 243 nucleotides have been removed, including the MafR binding site; Tc^R	This work
pASTT derivatives of *S. pneumoniae*			
pASTT-*PmgaP*	5,922	Derivative of pASTT that carries the promoter region of the *spr1404* or *mgaP* gene; Tc^R	This work
pASTT-*PmgaPΔ-10*	5,901	Derivative of pASTT that carries the promoter region of the *mgaP* gene, but the last 22 nucleotides have been removed, including the -10 box; Tc^R	This work
pASTT-*PmgaPΔ105*	5,817	Derivative of pASTT that carries the promoter region of the *mgaP* gene, but the first 105 nucleotides have been removed; Tc^R	This work
pASTT-*PpclA*	6,024	Derivative of pASTT that carries the promoter region of the *spr1403* or *pclA* gene; Tc^R	This work
pASTT-*PpclAΔ-10*	5,987	Derivative of pASTT that carries the promoter region of the *pclA* gene, but the last 37 nucleotides have been removed, including the -10 box; Tc^R	This work
pASTT-*PpclAΔ103*	5,922	Derivative of pASTT that carries the promoter region of the *pclA* gene, but the first 103 nucleotides have been removed; Tc^R	This work
pASTT-*PpclAΔ173*	5,851	Derivative of pASTT that carries the promoter region of the *pclA* gene, but the first 173 nucleotides have been removed; Tc^R	This work
pASTT-*PpclAΔ203*	5,823	Derivative of pASTT that carries the promoter region of the *pclA* gene, but the first 203 nucleotides have been removed, including the MgaP binding site; Tc^R	This work

Plasmid	Size (bp)	Description	Source
pASTT-*PpclAΔ224*	5,802	Derivative of pASTT that carries the promoter region of the *pclA* gene, but the first 224 nucleotides have been removed, including the MgaP binding site; Tc^R	This work

Tc^R, Km^R: resistance to tetracycline and kanamycin, respectively.

4.2. OLIGONUCLEOTIDES

The oligonucleotides used in this work were synthesized at the CIB-Protein Chemistry Facility (Applied Biosystems 3400 synthesizer) and, also, by IDT Company (Integrated DNA Technology). All oligonucleotides were resuspended in sterile water to a final concentration of 100 µM. In the case of purified oligonucleotides, HPLC was used.

Table 4. Oligonucleotides

Name	Sequence (5'-3')	Applications
Up-*gfp*	CAAGAGGGCAATGGCTGATA	pASTT sequencing
Dw-*gfp*	GCTCTTCTTGGTACTGTTTTC	
Int-*gfp*	CATCACCATCTAATTCAACAAG	pASTT sequencing and primer extension
Up*mafR*	GCAAAAGGAGGTTTT**CAT**ATGTACTCCATG	Construction of pET-$mafR_{OG1RF}$-His and pET-$mafR_{V583}$-His
Dw*mafR*-His	CCTCGCTAGTTC**TCGAG**AAAATAAGAATGA	
Up*mafR*-Δ3N	GGTTTTGCCATGTAC**CAT**ATGTTAAAACGT	Construction of pET-$mafR_{OG1RF}$ Δ3N-His
mgaP-Nde	ATGGAGGAAATACCC**A**TATGAGAAACCTTT	Construction of pET-*mgaP*-His
mgaP-Xho-His	GACTTTTTTGAT**CTCGAG**TAAAGTATTGGA	
T7 promoter	TAATACGACTCACTATAGGG	pET-24b sequencing
T7 terminator	GCTAGTTATTGCTCAGCGG	
FmgaP	CTTAGACAAAAAA**GC**AT**GC**AATGAATTATGG	Construction of pDLF*mgaP* and pDLF*mgaPi*
RmgaP	GTAAAGGAAGTATAG**C**A**TGC**AGATAAGAGAA	
*gldA*F	CAAAGTTTCGAGAGCTCAGGGGAGT	EMSA
*gldA*R	GTTGATAGCCTCCTA**G**A**GCTC**GTTTACATACA	

Name	Sequence (5'-3')	Applications
3012A	AGGAATGGCTGTTGTAACCA	EMSA
3013A	AACAAACGAATTTGCCGAAGC	
F*12294*-D	GATGTCAAAGCGTTAATTGGCA	EMSA and footprinting
R*12294*-D	GACCCGTTTGCTTCGTCTTAGT	
F*12294*-S	GTTGAGTTAGCAGACCGACGT	Sequencing reactions
R*12294*-S	ACTAATGGTGGTCTTACTCGC	
F*12294*	GAAACAGCGTTGA**G**CTCTTCTAGTGAC	Construction of pASTT-*P12294*
R*12294*	CATTTCGTGTACCTCC**G**A**GC**T**C**CTTGATACCT	
R*12294Δ-10*	CCTTTATTATAGAG**CTC**ATTTTCTTCACA	Construction of pASTT-*P12294Δ-10*
F*12294Δ69*	GTAAAAATGGTGAAAG**AGC**T**C**ATGTCAAAGCGT	Construction of pASTT-*P12294Δ69*
F*12294Δ208*	CAGGGGCCACTGA**GC**T**C**ATCGACCATT	Construction of pASTT-*P12294Δ208*
F*11486*-D	GCCACAGGAAGTAGCAAAACT	EMSA and footprinting
R*11486*-D	GGTTTGTGGATTTGATGAATGA	EMSA, primer extension and footprinting
F*11486*-S	GATTTTGTAGGAGAAGGTGA	Sequencing reactions, RT-PCR and qRT-PCR OG1RF_*11485*
R*11486*-S	GGAGCGAGAAGTACAGTGACA	
F*11486*	ACACCCATGAACG**G**AGCT**C**ATTTTGTA	Construction of pASTT-*P11486*
R*11486*	ATAAAACAACGA**GC**T**C**TTTTAGTGATAACC	
F*11486Δ66*	GGGCCGTTG**A**GC**T**CAGCCACAGGAAGTA	Construction of pASTT-*P11486Δ66*
F*11486Δ145*	GGCACAGTTATGA**GC**T**C**TGATGGTGGT	Construction of pASTT-*P11486Δ145*
F*11486Δ169*	GTTGGACCATGA**GC**T**C**AAAAGACTTTACA	Construction of pASTT-*P11486Δ169*
F*11486Δ188*	GACTTTACAGAG**CTC**CTGTTCTTCAGTA	Construction of pASTT-*P11486Δ188*
F*10478*-D	GCAAACTGTTTCCAGTGATAGT	EMSA and footprinting
R*10478*-D	CCACAAACAACGTCGCTCCATCA	EMSA, primer extension and footprinting
F*10478*-S	GTCATTTTAGTTCCTCCCTATGT	Sequencing reactions
R*10478*-S	GGCTTTGATGTTTTGCTGCTCT	

Name	Sequence (5'-3')	Applications
F*10478*	CACTAATACAG**A**G**CTC**AATGTTGTCA	Construction of pASTT-*P10478*
R*10478*	GCATACGCTTA**G**A**G**CTCCTAGAACT	
R*10478Δ-10*	GAACTAAGTATAC**G**A**GC**TCAGCTTA	Construction of pASTT-*P10478Δ-10*
F*10478Δ122*	CTGTTTCCAGTGA**G**AG**C**TCTATTGTACCT	Construction of pASTT-*P10478Δ122*
F*10478Δ210*	GATAGTTTAAAAC**G**A**G**C**T**CCAAGCGTTGT	Construction of pASTT-*P10478Δ210*
F*10478Δ228*	GATAGTTTAAAACTATCACCAAGCGTTGT T**GAG**CT**C**GTCACTGT	Construction of pASTT-*P10478Δ228*
F*10478Δ234*	GATAGTTTAAAACTATCACCAAGCGTTGTTTTA CTAG**AGCTC**GTAAAAA	Construction of pASTT-*P10478Δ234*
F*10478Δ243*	CTAGTCACTGTA**G**A**GC**T**C**ATTTGTTTTT	Construction of pASTT-*P10478Δ243*
*1622*H	CGGATTAAACCTCTTGCAATTATACC	EMSA and footprinting
*1622*I	CAAATTCTTTAATTGTTGCTATTA	
*1622*A	AGTTCCTGATTGTATTCCCT	Sequencing reactions and qRT-PCR *spr1622*
*1622*B	CACAACACTGCCTACCCTCC	Sequencing reactions
Up*1404*	CTCCTAGATAGTATTTTATAGT	EMSA and footprinting
Dw*1404*	GAATTAGGGTTTCCATTAAGCGT	
Up*1404*-2	CAATGTACAACTATTTGAGGCA	Primer extension and footprinting
Dw*1404*-2	CGGCTAGTTCATGTAATTTCATCCA	
R*mgaP*-q	CAGGAAGGTCAGGAAAAGGC	Sequencing reactions and qRT-PCR *spr1404*
R*pclA*-S	CACTAGCTGCACCAAAGTAAT	Sequencing reactions
Up*pclA*	CTAATTTTTCGGC**G**AG**C**TCAT GTAATT	Construction of pASTT-*PpclA*
Dw*pclA*	CATTTTAAACTCC**G**AG**CTC**GTATTTTA	Construction of both pASTT-*PpclA* and pASTT-*PmgaP*
mgaP-Dw	CATACGGGTATT**GAGCTC**ATAATTCATT	Construction of both pASTT-*PmgaP* and pASTT-*PpclAΔ103*
mgaPΔ-10	TTCATTATATC**GAGCTC**TGTCTAAGTAG	Construction of pASTT-*PmgaPΔ-10*
mgaPΔ105	AGCCACTTATATT**GAG**CT**C**ATTTTTCCG	Construction of both pASTT-*PmgaPΔ105* and pASTT-*PpclAΔ208*

Name	Sequence (5'-3')	Applications
R*pclAΔ-10*	CAATGTACAACTATTTGAG**CTC**TTAATCAAT	Construction of pASTT-*PpclAΔ-10*
F*pclAΔ173*	GAAAATTGTCGTTAG**AGC**TCGGAAAAATT	Construction of pASTT-*PpclAΔ173*
F*pclAΔ203*	GAATAATATAAG**GA**GCT**C**TTAAGTGCAAA	Construction of pASTT-*PpclAΔ203*
F*pclAΔ224*	GTGCAAAAACAGA**GCTC**ATTAATTGTT	Construction of pASTT-*PpclAΔ224*
F*10478*-q	GAGCAGCAAAACATCAAAGCCT	qRT-PCR OG1RF_*10478*
R*10478*-q	AGATAGGGAGCGAGCATTTC	
F*10600*-q	GCGTAGAAGAGTCAGCACTA	qRT-PCR OG1RF_*10600*
R*10600*-q	GCCATTCACAACGGTACAGC	
F*zwf*-q	CGGTCAAGGGTTCAATACAACT	qRT-PCR OG1RF_*10737*
R*zwf*-q	CCAAGATTGGGCAACTTCGTCCCA	
F*mgtA2*-q	GATTTTTCCAGACGCCGCTT	qRT-PCR OG1RF_*11140*
R*mgtA2*-q	ACACACGCATCCCTTGTTTG	
F*gldA*-q	GAGGGTGGCTTTAGTGGAGA	qRT-PCR OG1RF_*11146*
R*gldA*-q	TTCACCTTCTGCTACGACTT	
R*11485*-q	CAATGGTCATAACTGTGCCA	qRT-PCR OG1RF_*11485*
F*11486*-q	TGGTTACCGCTTTGTATGTTG	qRT-PCR OG1RF_*11486* and RT-PCR
R*11486*-q	CCCTAACGTAATGGACCAGAT	
F*11602*-q	CAACACCTCATTAGCGAAAC	qRT-PCR OG1RF_*11602*
R*11602*-q	GTCAATCATACCGACTAAACCA	
F*mafR*-q	ACTTTATCAACCGTCCTTGG	qRT-PCR OG1RF_*12293*
R*mafR*-q	GTTTCGCCATAGACATTATC	
F*12294*-q	TCCTCTACCGTTGACACCTG	qRT-PCR OG1RF_*12294*
R*12294*-q	TGCCTTCGTTGACATCTCTTG	
F*recA*-q	GCAACGAAATGGTGGAACAG	qRT-PCR OG1RF_*12439*
R*recA*-q	AAGGCATCGGCAATCTCTAAG	
F*era*-q	GATTGCCATCATGAGTGACAAGG	qRT-PCR *spr0871*
R*era*-q	AGTGTCCACTTCGCGAAGGGT	

Name	Sequence (5'-3')	Applications
F*pclA*-q	GACGTGATGGTTCAGCTCCA	qRT-PCR *spr1403*
R*pclA*-q	GGATTTGTCACCGTAATTGT	
F*mgaP*-q	CCAACCTCTATCGACTGGGCA	qRT-PCR *spr1404*
*1622*J	GAATAAGGATAATCTGATTTGGCA	qRT-PCR *spr1622*
F*1623*-q	GGGGGACAGTGGTTCTATCA	qRT-PCR *spr1623*
*1623*B	CGTAAATTTACATGAACAGTTGGG	
F*RpoB*-q	CTGCTTACAACCAACGCACT	qRT-PCR *spr1777*
R*RpoB*-q	GACATCGTCCTTCACCAAGC	

Restriction sites are underlined, and the bases changes that generate restriction site are in bold.

5. BUFFER SOLUTIONS

All buffers and solutions used in this work are listed in Table 5.

Table 5. Buffers

Buffer	Composition	Application
BXF	80% Deionised formamide 10 mM NaOH 0.1% Bromophenol Blue 0.1% Xylene cyanol 1 mM EDTA	Loading-dye for DNA electrophoresis in denaturing polyacrylamide gels
BXGE 10X	0.25% Bromophenol Blue 0.25% Xylene cyanol 60% Glycerol 10 mM EDTA	Loading-dye for gel electrophoresis of DNA and DNA-protein complexes
Cracking buffer	50 mM Tris-HCl, pH 6.8 2% SDS 143 mM β-mercaptoethanol 2 mM EDTA 10% Glycerol	Analysis protein induction system
EB	200 mM NaCl 20 mM Tris-HCl, pH 8 2 mM EDTA	Elution of DNA from native polyacrylamide (5%) gels
Electroporation buffer pH7.0	500 mM Sucrose 10% Glycerol	Storage of electrocompetent cells of *E. faecalis*

Buffer	Composition	Application
LB	10 mM Tris, pH 8.0 1 mM EDTA, pH 8.0 1 mg/ml Lysozyme 150 U Mutanolysin	Lysis buffer for *E. faecalis* RNA extraction
LBP	50 mM Tris, pH 7.6 1 mM EDTA, pH 8.0 50 mM NaCl 0.1% Deoxycholate (DOC)	Lysis buffer for *S. pneumoniae* RNA extraction
PBS	10 mM Na_2HPO_4 2 mM KH_2PO_4 2.7 mM KCl 137 mM NaCl pH 7.4	Resuspension of cells for fluorescence measurements
S-His 1X	10 mM Tris-HCl, pH 7.6 5% Glycerol 300 mM NaCl 1 mM DTT	Purification of $MafR_{OG1RF}$-His, $MafR_{OG1RF}$-Δ3N-His and $MafR_{v583}$-His proteins. Buffer S-His was supplemented with imidazole (from 10 mM to 250 mM)
S-His-P 1X	10 mM Tris-HCl, pH 7.6 5% Glycerol 300 mM NaCl 0.5 mM DTT	Purification of $MgaP_{R6}$-His protein. Buffer S-His-P was supplemented with imidazole (from 10 mM to 250 mM)
SLB 5X	250 mM Tris-HCl, pH 7.2 10% SDS 3.5 M β-mercaptoethanol 50% Glycerol 0.5% Bromophenol blue	Loading-dye for protein electrophoresis (SDS-PAGE)
Solution II	0.17 M NaOH 1% SDS	Lysis Buffer for both *E faecalis* and *S. pneumoniae* plasmid DNA
TAE	40 mM Tris 20 mM Acetic acid 2 mM EDTA, pH 8.1	DNA electrophoresis in agarose gels
TBE	89 mM Tris 89 mM Boric acid 2.5 mM EDTA pH 8.3	DNA electrophoresis in polyacrylamide gels
TE	10 mM Tris, pH 8.0 1 mM EDTA, pH 8.0	Storage DNA
TG	50 mM Tris-HCl, pH 8.3 300 mM Glycine 0.1% SDS 2 mM EDTA	Protein electrophoresis: SDS-PAGE (Tris-Glycine)

Buffer	Composition	Application
VL	50 mM Tris-HCl, pH 7.6 5% Glycerol 1 mM DTT 1 mM EDTA	Purification and storage of $MafR_{OG1RF}$-Δ3N-His, $MafR_{v583}$-His and $MgaP_{R6}$-His protein. Buffer VL was supplemented with different concentrations of NaCl

6. ACRYLAMIDE SOLUTIONS

For protein electrophoresis, an acrylamide:bis-acrylamide 30% solution (37.5:1) was used. For nucleic acids, an acrylamide:bis-acrylamide 40% solution (29:1) for non-denaturing gels (native gels), and an acrylamide:bis-acrylamide 40% solution (19:1) for denaturing gels, were used.

7. SOFTWARE

Table 6. Bioinformatic tools

Application	Program	Company/webpage
Analysis DNA sequences	ApE	jorgensen.biology.utah.edu/wayned/ape
Analysis oligonucleotides	OligoAnalyzer Tm calculator	eu.idtdna.com thermofisher.com/molecular-biology.html
Analysis qRT-PCR	iQ™5 Optical System Software	Bio-Rad
Bibliography manager	EndNote X	Thomson Reuters
Data analysis and graphing	Microsoft Excel 2007	Microsoft
Homologies finder	BLAST	blast.ncbi.nlm.nih.gov/Blast.cgi
Image and figures processing	Adobe Illustrator CS3 Adobe Photoshop CS3	Adobe Systems Inc.
	BioRender	BioRender.com
	Image Lab™ Touch Software	Bio-Rad
Oligonucleotide design	Primer3 Primer Blast	primer3.ut.ee ncbi.nlm.nih.gov/tools/primer-blast
Prediction of intrinsic curvature	Bend.it	pongor.itk.ppke.hu/dna/bend_it.html
Prediction of secondary structures	SABLE program	sable.cchmc.org

Application	Program	Company/webpage
Primary protein structure analysis	ProtParam	web.expasy.org/protparam
Promoter finder	BPROM	softberry.com
Protein homology detection and structure prediction	HHpred	toolkit.tuebingen.mpg.de/tools/hhpred
Protein visualization	ChemiDoc™ Touch Imaging System	Bio-Rad
Radiolabelled DNA visualization	Image-reader MultiGauge Quantity One	Fuji Fuji Bio-Rad
Restriction maps	ApE	jorgensen.biology.utah.edu/wayned/ape
Sequence alignments	Clustal Omega EMBOSS Needle	ebi.ac.uk/Tools/msa/clustalo/ ebi.ac.uk/Tools/psa/emboss_needle/

8. AUTORADIOGRAPHY AND RADIOACTIVE MATERIAL

The radioactive DNA was visualized either by autoradiography using the X-Omat S films (Kodak) or using a Phosphor screen scanned with a Fujifilm Image Analyzer FLA-3000 (Fuji). When autoradiography was used, the radioactive signal was amplified using the intensifying screens Cronex Lightning Plus (Dupont).

METHODS

1. BACTERIAL GROWTH CONDITIONS

In liquid medium, *E. coli* was grown at 37°C in TY medium with rotary shaking using Erlenmeyer flasks. To maintain constant aeration, the flask volume was 5-times larger than the culture volume. In solid medium, bacteria cells were grown at the surface of TY-agar plates and incubated at 37°C.

E. faecalis and *S. pneumoniae* were grown in liquid medium under low or without aeration conditions at 37°C unless otherwise indicated. Bacteria were grown in flask whose volume was twice the culture volume and incubated in a static bath. In solid medium, enterococcal cells were uniformly spread over SR medium plates. In the case of pneumococci, cells and antibiotic (when required) were mixed with a basal layer (20 ml) of AGCH medium supplemented with sucrose (0.3%) and yeast extract

(0.2%) plus 1% agar. Then, an over-layer (8 ml) of AGCH medium plus 0.75% agar was added covering the basal layer. Plates were incubated at 37°C.

In all cases, bacterial growth in liquid medium was followed by turbidity at 650 nm for both *E. faecalis* and *S. pneumoniae* or at 600 nm for *E. coli*, using a Bausch & Lomb (Spectronic 20D+) spectrophotometer.

For the preservation of all bacterial cultures, cells were grown to an optical density (OD) between 0.3 and 0.4, which corresponds with the exponential phase growth. In the case of pneumococci, cells were grown to an OD between 0.2 and 0.4. Then, sterile glycerol was added to 1 ml of culture to a final concentration of 10%. The culture was kept at 37°C for 10 min and then on ice for 10 min. Finally, cultures were stored at -80°C.

1.1. BACTERIAL GROWTH CURVES

Bacterial growth curves describe the density of cell populations in liquid culture over time. There are 4 phases in the bacterial growth: (i) lag phase (bacteria are metabolically active but not dividing), (ii) exponential phase (log: exponential growth), (iii) stationary phase (growth reaches a plateau due to the number of dying cells and dividing cells is equal), and (iv) death. To this aim, cultures of enterococcal or pneumococcal cells were grown in liquid medium as described above (Methods, Section 1), and the OD was measured at 650 nm using a Bausch & Lomb (Spectronic 20D+) spectrophotometer, each 15-30 min. Data analysis and graphing were performed using Microsoft Excel 2007 (Table 6).

2. BACTERIAL TRANSFORMATION

2.1. PREPARATION OF COMPETENT CELLS

E. coli electrocompetent cells were prepared from cultures grown with rotary shaking in SOC medium to an OD_{650} of 0.5 (exponential phase). Then, the culture was cooled on ice and centrifuged at 5,000 rpm in an Eppendorf F-34-6-38 rotor for 15 min at 4°C. The cell pellet was washed several times with ice-cold sterile water. Finally, cells were resuspended in 10% glycerol, and aliquots (50 µl) were stored at -80°C.

E. faecalis electrocompetent cells were prepared as reported previously (Shepard and Gilmore, 1995). Cells were grown overnight in M17 medium (supplemented with antibiotic if required) at 37°C without aeration to an OD_{560} of 0.7-0.8. Subsequently, the culture was 100-fold diluted into SG-M17 medium (supplemented with antibiotic if required). The volume of the flask was 5-times larger than

the culture volume, and the incubation continued as before. After 20-24 h, when the cultures reached an OD_{560} of 0.1-0.2, cells were collected by centrifugation at 6,000 rpm in an Eppendorf F-34-6-38 rotor for 10 min at 4°C. Then, cells were washed twice with ice-cold electroporation buffer and resuspended in a volume of ice-cold electroporation buffer equal to 1/100 of the initial culture volume. Finally, aliquots of the cultures (40 µl) were stored at -80°C.

S. pneumoniae competent cells were prepared as reported previously (Lacks, 1966). First, cells were grown in AGCH medium supplemented with sucrose (0.3%) and antibiotic (if required) at 37°C without aeration to an OD_{650} of 0.3 (exponential phase). Then, culture was diluted 1:40 with prewarmed medium and cells were grown under the same conditions to an OD_{650} of 0.3. This dilution step was repeated twice. Finally, glycerol was added to a final concentration of 10%, and aliquots of the culture (250 µl) were stored at -80°C.

2.2 TRANSFORMATION

2.2.1. ELECTROPORATION

E. coli and *E. faecalis* electrocompetent cells were transformed as reported previously (Dower *et al.*, 1988; Shepard and Gilmore, 1995). DNA in water was mixed with electrocompetent cells in a 2 mm electroporation cuvette (Cell Projects) pre-cooled on ice. The electric pulse was generated with a MicroPulser (Bio-Rad) (2.50 kV and 3-5 ms). In the case of *E. faecalis*, after the electric pulse, 1 ml of ice-cold SG-M17-MC medium was added to the cells. The cuvette was maintained at least 5 min on ice and then the cells were incubated at 37°C for 2 h without aeration. Finally, the enterococcal cells were uniformly spread over SR plates supplemented with antibiotic for selection of transformants. Related to *E. coli* cells, after the pulse, they were transferred to 0.8 ml SOB medium containing glucose (0.4%) and incubated at 37°C with rotary shaking for 1 h. To select transformants, TY-agar plates supplemented with the appropriate antibiotic were used.

2.2.2. NATURAL TRANSFORMATION

S. pneumoniae is able to uptake foreign DNA by a process called natural transformation, which was previously reported by Lacks (Lacks, 1966). Briefly, cells were inoculated in AGCH medium supplemented with sucrose (0.2%) and $CaCl_2$ (70 µM) and incubated at 30°C for at least 20 min before addition of DNA. When strain R6 or derivatives were used, 25 ng of CSP-1 was added at the same time as the DNA. After the addition of DNA, cultures were incubated at 30° for 40 min. Then, to enable the phenotypic expression (resistance to Tc or Km), cultures were incubated at 37° for 90 min.

Transformants were selected in AGCH plates (described in Methods, Section 1) supplemented with the appropriate antibiotic.

3. CONSTRUCTION OF BACTERIAL STRAINS

3.1. CONSTRUCTION OF *E. faecalis* OG1RFΔ*mafR* STRAIN

As reported by (Ruiz-Cruz *et al.*, 2016), strain OG1RFΔ*mafR* was constructed by deleting the chromosomal region that spans coordinates 2421575 to 2422640 (Figure 7, Results, Chapter 1.1) using the plasmid pBVGh (thermosensitive replicon) (Blancato and Magni, 2010). To this end, both 560-bp (located upstream of *mafR* gene) and 494-bp (located at the end of the gene, including the *mafR* 3'-end) regions were amplified using primers which introduced *Cla*I and *Nco*I sites. Both PCR-synthesized DNAs were digested with *Cla*I, mixed in equimolecular amounts and ligated with T4 DNA ligase. The ligation product was then used as template for PCR amplification, using the same primers which introduced *Nco*I site. New 1,017-bp PCR product was digested with *Nco*I, and then the restriction fragment was cloned into the *Nco*I site of the plasmid pBVGh (pBVGhΔ*mafR*). Strain OG1RF harbouring pBVGhΔ*mafR* was used to generate strain OG1RFΔ*mafR* following the protocol (integration/excision process) reported by (Blancato and Magni, 2010). OG1RFΔ*mafR* mutant strain sequence was confirmed by dye-terminator sequencing carried out at Secugen (CIB, Madrid).

3.2. CONSTRUCTION OF *S. pneumoniae* R6Δ*mga* STRAIN

As reported by (Solano-Collado *et al.*, 2012), to construct the R6Δ*mga* mutant strain, gene replacement by homologous recombination was carried out, deleting the chromosomal region that spans coordinates 1596826 to 1598431. To this end, a 1,165-bp DNA fragment that contained the pC194 *cat* gene (chloramphenicol resistance) (Horinouchi and Weisblum, 1982) was flanked by the regions flanking the *mgaSpn* gene (543-bp and 605-bp). The *cat* cassette generated *in vitro* was used to transform competent R6 cells. Selection of transformants resistant to chloramphenicol led to the isolation of R6Δ*mga* strain. R6Δ*mga* mutant strain sequence was confirmed by dye-terminator sequencing carried out at Secugen (CIB, Madrid).

4. DNA PREPARATIONS

4.1. PLASMID DNA ISOLATION

High Pure Plasmid Isolation Kit (Roche Applied Science) was used for *E. faecalis*, *S. pneumoniae* and *E. coli* plasmid DNA isolation. Some modifications in the composition of specific buffers were done

when enterococcal and pneumococcal cells were used. The Suspension Buffer was supplemented with 50 mM glucose, 1.2 mg/ml lysozyme and 240 units/ml mutanolysin for *E. faecalis*, or 50 mM glucose and 0.1% deoxycholate in the case of *S. pneumoniae*. For both bacteria, the Lysis Buffer (Solution II, Table 5) was freshly prepared (0.17 M NaOH and 1% SDS).

4.2. GENOMIC DNA ISOLATION

E. faecalis and *S. pneumoniae* genomic DNA were kindly provided by Dr Sofía Ruiz-Cruz and Dr María Virtudes Solano-Collado, respectively. To isolate chromosomal DNA of *S. pneumoniae*, cells were grown in 40 ml AGCH medium supplemented with 0.3% sucrose and 0.2% yeast extract. Cultures were centrifuged when an OD_{650} of 0.8 was reached. Then, cells were resuspended in 1 ml TE buffer (Table 5) supplemented with 0.6 mg proteinase K and 0.6% SDS and incubated at 37°C for 30 min with gentle shaking. Samples were phenol treated, centrifuged and the supernatant was dialyzed against TE buffer. After that, 40 µg RNase A was added and the mixture was incubated at 37°C for 30 min. Finally, samples were phenol treated, ethanol precipitated and resuspended in TE. For small-scale preparations of chromosomal DNA from *E. faecalis* and *S. pneumoniae*, the Bacterial Genomic Isolation Kit (Norgen Biotek Corporation) was used.

4.3. PREPARATION OF LINEAR DOUBLE-STRANDED DNA FRAGMENTS

Linear double-stranded DNA (dsDNA) fragments used in this Thesis were obtained by digestion with restriction enzymes or by PCR (Polymerase Chain Reaction).

4.3.1. DIGESTION WITH RESTRICTION ENZYMES

Digestion of DNA using restriction enzymes was done using the conditions specified by the supplier. When needed, 10 µg/ml of BSA was included in the digestion reaction. In general, enzymes were inactivated at 65°C for 10-20 min.

4.3.2. POLYMERASE CHAIN REACTION

All PCR reactions were carried out in an iCycler Thermo Cycler (Bio-Rad). Analysis of *E. faecalis* and *S. pneumoniae* transformants by PCR (colony PCR) was performed with Taq DNA polymerase (Roche and Invitrogen). In the case of *E. faecalis*, the transformants colonies were inoculated with a sterile toothpick in both growth medium (50 µl) and sterile water (50 µl). Then, 3 µl of the cells suspended in water were used as template for PCR. In the case of *S. pneumoniae*, a single colony was inoculated into 100 µl of growth medium. After that, 10 µl were added to 90 µl of water and 1-3 µl of this

mixture were used as a template for PCR. For both bacteria, reactions (25 µl) contained 20 mM Tris-HCl pH 8.4 (Invitrogen Taq DNA polymerase) or 10 mM Tris-HCl pH 8.3 (Roche Taq DNA polymerase), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 12.5 pmol of each primer and 0.5 unit of Taq DNA polymerase. PCR conditions were: initial denaturation step at 94°C for 2-3 min, followed by 30 cycles including the next steps: denaturation at 94°C for 30-45 s, annealing of the primer to the DNA template at around 55°C (depending on the primer Tm) for 30 s followed by an extension at 72°C for 45-90 s (depending on the amplicon length). A final extension step was performed at 72°C for 7-10 min.

Phusion High-Fidelity DNA polymerase (Finnzymes) was used when high fidelity PCR products were required. Reaction mixtures (50 µl) contained 5-30 ng of template DNA, 50 pmol of each primer, 200 µM of each dNTP and 1 unit of DNA polymerase. PCR conditions were: initial denaturation step at 98°C for 1 min, followed by 30 cycles including the next steps: denaturation at 98°C for 10 s, annealing of the primer to the DNA template at around 55°C (depending on the primer Tm) for 20 s followed by an extension at 72°C for 40 s (depending on the amplicon length). A final extension step was performed at 72°C for 10 min.

4.4. DNA PURIFICATION

4.4.1. LINEAR DOUBLE-STRANDED DNAS

Linear dsDNA fragments obtained either by digestion with restriction enzymes or by PCR, were purified (when required) with the QIAquick PCR purification kit (QIAGEN). When the DNA fragments were purified from agarose gels, the QIAquick gel extraction kit (QIAGEN) was used, following the specification of the supplier. Linear dsDNA fragments radioactively labelled at 5'-ends (obtained by PCR) were purified from non-denaturing polyacrylamide gels (5%). After electrophoresis, DNA was visualized by autoradiography. The desired band was excised with a clean scalpel and the gel slice was incubated overnight in elution buffer (EB buffer, Table 5) with continuous shaking (450 rpm) at 42°C using a Thermo-shaker (TS-100, Biosan). The remains of the gel were removed using a Spin-X column (Costar). Finally, the eluted DNA was ethanol precipitated and dissolved in distilled water.

4.5. LIGATION

In general, for ligation of linear DNA fragments, the molar ratio vector to insert was between 1:5 and 1:10. When dephosphorylated vectors were used, the molar ratio vector to insert was between 1:5 and 1:10. In all cases, 400 units of T4 DNA ligase (New England Biolabs) were added to a final reaction

volume of 25-40 µl. Reaction mixtures were incubated overnight at room temperature or at 16°C for 2 h.

4.6. CONSTRUCTION OF RECOMBINANT PLASMIDS

The sequences of all the constructions described below were confirmed by dye-terminator sequencing at Secugen (CIB, Madrid).

4.6.1. CONSTRUCTION OF PLASMIDS pET24b-*mafR*$_{OG1RF}$-HIS, pET24b-*mafR*$_{OG1RF}$Δ3N-HIS AND pET24b-*mafR*$_{V583}$-HIS

To overproduce and purify the MafR$_{OG1RF}$-His, MafR$_{OG1RF}$Δ3N-His and MafR$_{V583}$-His proteins, the *mafR* or *mafR*Δ3N genes were engineered to encode a MafR or MafRΔ3N proteins, respectively, fused to a C-terminal His$_6$-tag. To this end, plasmids pET24b-*mafR*$_{OG1RF}$-His, pET24b-*mafR*$_{OG1RF}$Δ3N-His and pET24b-*mafR*$_{V583}$-His were constructed (Methods, Section 8.1). In all cases, the pET24b expression vector (Novagen) was used. To obtain pET24b-*mafR*$_{OG1RF}$-His and pET24b-*mafR*$_{V583}$-His plasmids, a 1,481-bp region of the OG1RF and V583 chromosome, respectively, containing the *mafR* gene was amplified by PCR using the Up*mafR* and Dw*mafR*-His primers. These oligonucleotides contained a single restriction site for *Nde*I and *Xho*I, respectively. The amplified products were digested with both enzymes, and the 1,448-bp digestion products were inserted into the pET24b expression vector. To obtain pET24b-*mafR*$_{OG1RF}$Δ3N-His, a 1,472-bp region of the OG1RF chromosome containing the *mafR* gene was amplified by PCR using the Up*mafR*-Δ3N and Dw*mafR*-His primers, generating target sites for *Nde*I and *Xho*I restriction enzymes, respectively. The amplified product was digested with both enzymes, and the 1,440-bp digestion product was cloned into the pET24b expression vector.

4.6.2. CONSTRUCTION OF PLASMID pET24b-*mgaP*-HIS

To overproduce and purify the MgaP-His protein, plasmid pET24b-*mgaP*-His was constructed (Methods, Section 8.2). The *mgaP* gene was engineered to encode a C-terminal tag (Leu-Glu-His$_6$) MgaP protein. Specifically, a 1,517-bp region of R6 chromosome, containing the *mgaP* gene was amplified by PCR using *mgaP-Nde* and *mgaP-Xho*-His primers. These oligonucleotides contained a single restriction site for *Nde*I and *Xho*I, respectively. Then, the amplified DNA fragment was digested with both enzymes, and the 1,484-bp digestion product was cloned into the pET24b expression vector, which enables a C-terminal His$_6$-tag fusion. The pET24b-*mgaP*-His construction was kindly provided by Alejandro Ortuno-Camuñas.

4.6.3. CONSTRUCTION OF PLASMIDS pDLF*mgaP* AND pDLF*mgaPi*

Plasmids pDLF*mgaP* and pDLF*mgaPi* are derivatives of pDLF, which carries a Km resistance gene and the *P2493* promoter (Ruiz-Cruz *et al.*, 2016). Both pDLF*mgaP* and pDLF*mgaPi* contain the promoterless *mgaP* gene from R6 strain, either under the control of the *P2493* promoter (pDLF*mgaP*) or in the opposite orientation (pDLF*mgaPi*). To construct both plasmids, a 1,594-bp region of the R6 genome was amplified by PCR using *FmgaP* and *RmgaP* primers, generating a target site for the *Sph*I restriction enzyme. Then, the amplified DNA fragment was digested with the restriction enzyme, and the 1,561-bp digestion product was cloned in both orientations into the *Sph*I site of the pDLF vector. Both pDLF*mgaP* and pDLF*mgaPi* constructions were kindly provided by Alejandro Ortuno-Camuñas.

4.6.4. CONSTRUCTION OF pASTT DERIVATIVES

Plasmid pASTT is a derivative of pAST (Ruiz-Cruz *et al.*, 2010) that carries the *TrsiV* transcriptional terminator downstream of the promoterless *gfp* reporter gene, which encodes a variant of the green fluorescent protein (D. García-Rincón, V. Solano-Collado and A. Bravo, unpublished results). To construct pASTT derivatives, different DNA fragments (promoter or promoterless sequence) were inserted into the multiple cloning site of the pASTT plasmid, just upstream of the promoterless *gfp* reporter gene.

Table 7. pASTT derivatives summary

Plasmid	Oligonucleotides	PCR product	Digestion product	Cloned coordinates
		***E. faecalis* OG1RF**		
pASTT-*P12294*	F*12294* and R*12294*	293 bp	255 bp	NC 2425885 to 2425631
pASTT-*P12294Δ-10*	F*12294* and R*12294Δ-10*	264 bp	236 bp	NC 2425885 to 2425655
pASTT-*P12294Δ69*	F*12294Δ69* and R*12294*	229 bp	192 bp	NC 2425816 to 2425631
pASTT-*P12294Δ208*	F*12294Δ208* and R*12294*	86 bp	53 bp	NC 2425677 to 2425631
pASTT-*P11486*	F*11486* and R*11486*	319 bp	284 bp	C 1542902 to 1543185
pASTT-*P11486Δ66*	F*11486Δ66* and R*11486*	248 bp	224 bp	C 1542968 to 1543185
pASTT-*P11486Δ145*	F*11486Δ145* and R*11486*	172 bp	145 bp	C 1543047 to 1543185
pASTT-*P11486Δ169*	F*11486Δ169* and R*11486*	147 bp	121 bp	C 1543071 to 1543185

Plasmid	Oligonucleotides	PCR product	Digestion product	Cloned coordinates
pASTT-*P11486Δ188*	F*11486Δ188* and R*11486*	127 bp	102 bp	C 1543090 to 1543185
pASTT-*P10478*	F*10478* and R*10478*	328 bp	301 bp	NC 498575 to 498278
pASTT-*P10478Δ-10*	F*10478* and R*10478Δ-10*	308 bp	279 bp	NC 498575 to 498301
pASTT-*P10478Δ122*	F*10478Δ122* and R*10478*	207 bp	177 bp	NC 498453 to 498278
pASTT-*P10478Δ210*	F*10478Δ210* and R*10478*	120 bp	90 bp	NC 498365 to 498278
pASTT-*P10478Δ228*	F*10478Δ228* and R*10478*	120 bp	73 bp	NC 498347 to 498278
pASTT-*P10478Δ234*	F*10478Δ234* and R*10478*	120 bp	67 bp	NC 498341 to 498278
pASTT-*P10478Δ243*	F*10478Δ243* and R*10478*	87 bp	58 bp	NC 498332 to 498278
	***S. pneumoniae* R6**			
pASTT-*PmgaP*	Dw*pclA* and *mgaP*-Dw	221 bp	190 bp	C 1387937 to 1388121
pASTT-*PmgaPΔ-10*	Dw*pclA* and *mgaPΔ-10*	199 bp	169 bp	C 1387937 to 1388099
pASTT-*PmgaPΔ105*	*mgaPΔ105* and *mgaP*-Dw	116 bp	85 bp	C 1388042 to 1388121
pASTT-*PpclA*	Up*pclA* and Dw*pclA*	324 bp	292 bp	NC 1388224 to 1387937
pASTT-*PpclAΔ-10*	Up*pclA* and R*pclAΔ-10*	290 bp	255 bp	NC 1388224 to 1387974
pASTT-*PpclAΔ103*	*mgaP*-Dw and Dw*pclA*	221 bp	190 bp	NC 1388121 to 1387937
pASTT-*PpclAΔ173*	F*pclAΔ173* and Dw*pclA*	152 bp	119 bp	NC 1388051 to 1387937
pASTT-*PpclAΔ203*	F*pclAΔ203* and Dw*pclA*	122 bp	91 bp	NC 1388021 to 1387937
pASTT-*PpclAΔ224*	F*pclAΔ224* and Dw*pclA*	100 bp	70 bp	NC 1388000 to 1387937

CONSTRUCTION OF PLASMIDS pASTT-*P12294*, pASTT-*P12294Δ-10*, pASTT-*P12294Δ69* AND pASTT-*P12294Δ208*

The pASTT-*P12294* recombinant plasmid was used to characterize the *P12294* promoter. To construct pASTT-*P12294*, a 293-bp region of the OG1RF enterococcal genome was amplified by PCR using the F*12294* and R*12294* primers, which contained a restriction site for *Sac*I. The amplified DNA was digested with *Sac*I, and the 255-bp digestion product (coordinates 2425885 to 2425631) was inserted into the *Sac*I site of pASTT. To construct pASTT-*P12294Δ-10*, pASTT-*P12294Δ69* and pASTT-

P12294Δ208, different regions of the OG1RF chromosome were amplified by PCR using the primers indicated in brackets: pASTT-*P12294Δ-10* (F*12294* and R*12294Δ-10*, 264-bp DNA fragment), pASTT-*P12294Δ69* (F*12294Δ69* and R*12294*, 229-bp DNA fragment) and pASTT-*P12294Δ208* (F*12294Δ208* and R*12294*, 86-bp DNA fragment). Next, all the amplified fragments were digested with *Sac*I and the 236-bp (coordinates 2425885 to 2425655), 192-bp (coordinates 2425816 to 2425631) and 53-bp (coordinates 2425677 to 2425631) restriction fragments, respectively, were cloned into the *Sac*I site of pASTT.

CONSTRUCTION OF PLASMIDS pASTT-*P11486,* pASTT-*P11486Δ66,* pASTT-*P11486Δ145,* pASTT-*P11486Δ169* AND pASTT-*P11486Δ188*

The pASTT-*P11486* recombinant plasmid was used to characterize the *P11486* promoter. To construct pASTT-*P11486*, a 319-bp region of the OG1RF enterococcal genome was amplified by PCR using the F*11486* and R*11486* primers, which contained a restriction site for *Sac*I. The amplified DNA was digested with *Sac*I, and the 284-bp digestion product (coordinates 1542902 to 1543185) was inserted into the *Sac*I site of pASTT. To construct pASTT-*P11486Δ66*, pASTT-*P11486Δ145*, pASTT-*P11486Δ169* and pASTT-*P11486Δ188*, different regions of the OG1RF chromosome were amplified by PCR using the primers indicated in brackets: pASTT-*P11486Δ66* (F*11486Δ66* and R*11486*, 248-bp DNA fragment), pASTT-*P11486Δ145* (F*11486Δ145* and R*11486*, 172-bp DNA fragment), pASTT-*P11486Δ169* (F*11486Δ169* and R*11486*, 147-bp DNA fragment) and pASTT-*P11486Δ188* (F*11486Δ188* and R*11486*, 127-bp DNA fragment). Next, all the amplified fragments were digested with *Sac*I and the 224-bp (coordinates 1542968 to 1543185), 145-bp (coordinates 1543047 to 1543185), 121-bp (coordinates 1543071 to 1543185) and 102-bp (coordinates 1543090 to 1543185) restriction fragments, respectively, were cloned into the *Sac*I site of pASTT.

CONSTRUCTION OF PLASMIDS pASTT-*P10478,* pASTT-*P10478Δ-10,* pASTT-*P10478Δ122,* pASTT-*P10478Δ210,* pASTT-*P10478Δ228,* pASTT-*P10478Δ234* AND pASTT-*P10478Δ243*

The pASTT-*P10478* recombinant plasmid was used to characterize the *P10478* promoter. To construct pASTT-*P10478*, a 328-bp region of the OG1RF enterococcal genome was amplified by PCR using the F*10478* and R*10478* primers, which contained a restriction site for *Sac*I. The amplified DNA was digested with *Sac*I, and the 301-bp digestion product (coordinates 498575 to 498278) was inserted into the *Sac*I site of pASTT. To construct pASTT-*P10478Δ-10,* pASTT-*P10478Δ122*, pASTT-*P10478Δ210*, pASTT-*P10478Δ228,* pASTT-*P10478Δ234* and pASTT-*P10478Δ243*, different regions of the OG1RF chromosome were amplified by PCR using the primers indicated in brackets: pASTT-*P10478Δ-10* (F*10478* and R*10478Δ-10*, 308-bp DNA fragment), pASTT-*P10478Δ122* (F*10478Δ122* and

R*10478*, 207-bp DNA fragment), pASTT-*P10478Δ210* (F*10478Δ210* and R*10478*, 120-bp DNA fragment), pASTT-*P10478Δ228* (F*10478Δ228* and R*10478*, 120-bp DNA fragment), pASTT-*P10478Δ234* (F*10478Δ234* and R*10478*, 120-bp DNA fragment) and pASTT-*P10478Δ243* (F*10478Δ243* and R*10478*, 87-bp DNA fragment). Next, all the amplified fragments were digested with *Sac*I and the 279-bp (coordinates 498575 to 498301), 177-bp (coordinates 498453 to 498278), 90-bp (coordinates 498365 to 498278), 73-bp (coordinates 498347 to 498278), 67-bp (coordinates 498341 to 498278) and 58-bp (coordinates 498332 to 498278) restriction fragments, respectively, were cloned into the *Sac*I site of pASTT.

CONSTRUCTION OF PLASMIDS pASTT-*PmgaP,* pASTT-*PmgaPΔ-10* AND pASTT-*PmgaPΔ105*

The pASTT-*PmgaP* recombinant plasmid was used to characterize the *PmgaP* promoter. To construct pASTT-*PmgaP*, a 221-bp region of the R6 streptococcal chromosome was amplified by PCR using the Dw*pclA* and *mgaP*-Dw primers, which contained a restriction site for *Sac*I. The amplified DNA was digested with *Sac*I, and the 190-bp digestion product (coordinates 1387937 to 1388121) was inserted into the *Sac*I site of pASTT. To construct pASTT-*PmgaPΔ-10* and pASTT-*PmgaPΔ105*, different regions of the R6 genome were amplified by PCR using the primers indicated in brackets: pASTT-*PmgaPΔ-10* (Dw*pclA* and *mgaPΔ-10*, 199-bp DNA fragment) and pASTT-*PmgaPΔ105* (*mgaPΔ105* and *mgaP*-Dw, 116-bp DNA fragment). The amplified fragments were digested with *Sac*I and the 169-bp (coordinates 1387937 to 1388099) and 85-bp (coordinates 1388042 to 1388121) restriction fragments, respectively, were cloned into the *Sac*I site of pASTT.

CONSTRUCTION OF PLASMIDS pASTT-*PpclA,* pASTT-*PpclAΔ-10,* pASTT-*PpclAΔ103,* pASTT-*PpclAΔ173,* pASTT-*PpclAΔ203* AND pASTT-*PpclAΔ224*

The pASTT-*PpclA* recombinant plasmid was used to characterize the *PpclA* promoter. To construct pASTT-*PpclA*, a 324-bp region of the R6 streptococcal chromosome was amplified by PCR using the Up*pclA* and Dw*pclA* primers, which contained a restriction site for *Sac*I. The amplified DNA was digested with *Sac*I, and the 292-bp digestion product (coordinates 1388224 to 1387937) was inserted into the *Sac*I site of pASTT. To construct pASTT-*PpclAΔ-10*, pASTT-*PpclAΔ103*, pASTT-*PpclAΔ173*, pASTT-*PpclAΔ20*3 and pASTT-*PpclAΔ224*, different regions of the R6 genome were amplified by PCR using the primers indicated in brackets: pASTT-*PpclAΔ-10* (Up*pclA* and R*pclAΔ-10*, 290-bp DNA fragment), pASTT-*PpclAΔ103* (*mgaP*-Dw and Dw*pclA*, 221-bp DNA fragment), pASTT-*PpclAΔ173* (F*pclAΔ173* and Dw*pclA*, 152-bp DNA fragment), pASTT-*PpclAΔ203* (F*pclAΔ203* and Dw*pclA*, 122-bp DNA fragment) and pASTT-*PpclAΔ224* (F*pclAΔ224* and Dw*pclA*, 100-bp DNA fragment). Next, all the amplified fragments were digested with *Sac*I and the 255-bp (coordinates 1387974 to 1388224), 190-

bp (coordinates 1387937 to 1388121), 119-bp (coordinates 1387937 to 1388051), 91-bp (coordinates 1387937 to 1388021) and 70-bp (coordinates 1387937 to 1388000) restriction fragments, respectively, were cloned into the *Sac*I site of pASTT.

4.7. RADIOACTIVE LABELLING OF DNA

Radiolabelled DNA was visualized either by autoradiography or using a Fujifilm Image Analyzer FLA-3000 (Phosphorimager). The intensity of the labelled DNA bands was quantified using the Quantity One software (Bio-Rad).

4.7.1. 5'-END LABELLING

Oligonucleotides were radioactively labelled at the 5'-end using [γ-^{32}P]-ATP and the T4 PNK (Polynucleotide kinase, New England Biolabs). Reaction mixtures (25 µl) contained 25 pmol of oligonucleotide, 2.5 µl of 10X kinase buffer (provided by the supplier), 41.5 pmol of [γ-^{32}P]-ATP (3,000 Ci/mmol; 10 µCi/µl) and 10 units of T4 PNK. After incubation at 37°C for 30 min, additional T4 PNK (10 units) was added and the reaction mixture was incubated at 37°C for 30 min. Finally, to inactivate the enzyme, reaction mixtures were incubated at 65°C for 20 min. Non-incorporated nucleotide was removed using MicroSpin™ G-25 columns (GE Healthcare).

The 5'-labelled oligonucleotides were used to obtain labelled-dsDNA by PCR amplification. Moreover, 5'-labelled oligonucleotides were used in primer extension reactions and for manual sequencing.

4.7.2. INTERNAL LABELLING

The incorporation of a radiolabelled nucleotide ([α-^{32}P]-dATP, 3,000 Ci/mmol; 10 µCi/µl, Hartmann) was used for primer extension reactions and manual sequencing (using the Sequenase version 2.0 DNA sequencing kit, according to the indications of the supplier).

5. ANALYSIS OF DNA

5.1. DNA QUANTIFICATION

To quantify non-radiolabelled DNA, agarose gels were stained with GelRed 1X (Biotium). Then DNA bands were visualized using a Gel Doc XR system (Bio-Rad) and quantified using a molecular weight marker (NZYDNA Ladder III) with the Quantity One program. In addition, the concentration of DNA samples was determined using a NanoDrop ND-2000 Spectrophotometer (Thermofisher).

To quantify 5'-labelled DNA, the ionizing radiation was measured with a scintillation counter (Wallac1450 MicroBeta, TriLux). Knowing that 125 µCi of [γ-^{32}P]-ATP (41.5 pmol) are equivalent to 1.37x10^{8} cpm, we estimated the incorporation of [γ-^{32}P]-ATP in the labelling reaction and then the DNA concentration using the total cpm obtained.

5.2. DNA ELECTROPHORESIS

5.2.1. AGAROSE GELS

To analyse DNA, horizontal agarose gel electrophoresis in TAE buffer (Table 5) was used. The agarose concentration used (generally 0.6-1.5 %) depended on the size of the DNA analysed. DNA samples (chromosomal DNA, plasmid DNA and linear dsDNA fragments) were mixed with BXGE buffer (Table 5) and loaded onto the gel. After electrophoresis, gels were stained in a solution of GelRed (1X in water supplemented with 0.1 M NaCl; Biotium) for 15-30 min at room temperature. Finally, DNA was visualized using a short wavelength UV light (254 nm) with a Gel Doc XR system (Bio-Rad). The image obtained was captured with the Quantity One software and quantified when it was necessary.

5.2.2. NATIVE POLYACRYLAMIDE GELS

Vertical polyacrylamide gels were run in TBE buffer (Table 5) using a Mini Protean-III system (Bio-Rad). The concentration of acrylamide used (usually 6 %) depended on the size of the DNA analysed. DNA samples were mixed with BXGE buffer (Table 5) and loaded onto the gels. Gels were run at 100 V at room temperature or 4°C. After electrophoresis, gels were stained in GelRed 1X and visualized as described above. When the DNA was radioactively labelled, after electrophoresis, gels were fixed with acetic acid (10%) and dried using a gel dryer (model 583, Bio-Rad). The radiolabelled DNA was visualized by autoradiography or using a Fujifilm Image Analyzer FLA-3000.

5.2.3. DENATURING POLYACRYLAMIDE GELS

Sequencing reactions, primer extension products and footprinting reactions were run onto vertical denaturing urea (8 M) polyacrylamide (6%) gels, preheated at 50°C. A Sequi-Gen GT sequencing system and 21x50 cm gels (Bio-Rad) were used. Gels were run at 50 W to maintain the gel temperature and the time varied depending on the experiment. After electrophoresis, gels were allowed to cool down, then, the siliconized glass plate was carefully removed from the gel assembly and, immediately, a large piece of Whatman™ 3MM paper filter was employed to cover the exposed gel. The gel was peeled off along the paper and covered with plastic Saran™ wrap, the vacuum-dried, and exposed to Phosphor screen. Bands were visualized using a Fujifilm Image Analyzer FLA-3000.

5.3. *IN SILICO* PREDICTION OF INTRINSIC DNA CURVATURE

The intrinsic curvature of the DNA was predicted with the bend.it server (pongor.itk.ppke.hu/dna/bend_it.html) (Vlahoviček *et al.*, 2003). The server predicts DNA curvature from DNA sequences. The curvature is calculated as a vector sum of dinucleotide geometries (roll, tilt and twist angles) using the BEND algorithm of Godsell and Dickerson (DNase I + nucleosome positioning data), and expressed as degrees per helical turn (10.5°/helical turn = 1°/basepair) (Goodsell and Dickerson, 1994).

6. DNA SEQUENCING

6.1. MANUAL DNA SEQUENCING: DIDEOXY CHAIN-TERMINATION SEQUENCING METHOD (SANGER METHOD)

In general, the dideoxy chain-termination method was chosen for manual DNA sequencing (Sanger *et al.*, 1977). The Sequenase Version 2.0 DNA Sequencing kit (USB Corporation) was used following the supplier indications. When linear dsDNA was used some modifications were performed. The sequencing reactions which were run alongside the primer extension products were carried out using DNA from M13mp18 (Yanisch-Perron *et al.*, 1985), otherwise indicated.

6.2. AUTOMATED DNA SEQUENCING

All the plasmid constructions, as well as the chromosomal modifications, were confirmed by dye-terminator sequencing at Secugen (Automated DNA Sequencing Service, CIB).

7. RNA TECHNIQUES

7.1. TOTAL RNA ISOLATION FROM *E. faecalis*

For primer extension, RT-PCR and qRT-PCR experiments, total RNA from enterococcal cells was isolated using RNeasy Mini Kit (QIAGEN). Cells were grown to an OD_{650} of 0.4 (exponential phase) and 2 ml of culture were centrifuged at 4°C. When necessary, cells were grown to an OD_{650} of 0.8 (stationary phase) and 1 ml of culture was centrifuged at 4°C. Then, cells were resuspended in 100 µl of LB buffer (Table 5) and 150 units of mutanolysin were added. The samples were incubated at 37°C for 10 min (cell lysis). After this step, the RNeasy Mini Kit (QIAGEN) was used following the supplier's recommendations. To remove any residual DNA in the RNA preparations, the samples were treated with DNase I recombinant RNase free (Roche). After DNase I digestion, the RNA preparations were

cleaned up using the kit columns. In some cases, an additional DNase I digestion and the subsequent RNA purification (kit columns) were performed. The integrity and yield of rRNAs were checked by agarose gel electrophoresis in all the RNA preparations. Finally, the concentration of the RNA preparations was determined using a NanoDrop ND-2000 Spectrophotometer (Thermofisher).

7.2. TOTAL RNA ISOLATION FROM *S. pneumoniae*

For primer extension and qRT-PCR experiments, total RNA from pneumococcal cells was isolated using RNeasy Mini Kit (QIAGEN). Cells were grown to an OD_{650} of 0.2-0.4 (exponential phase) and 2 ml of culture were centrifuged at 4°C. When necessary, cells were grown to an OD_{650} of 0.99 (stationary phase) and 1 ml of culture was centrifuged at 4°C. Then, cells were resuspended in 100 µl of LBP buffer (Table 5) and incubated at 37°C for 5 min (cell lysis). After this step, the RNeasy Mini Kit (QIAGEN) was used following the supplier's recommendations. To remove any residual DNA in the RNA preparations, the samples were treated with DNase I recombinant RNase free (Roche). After DNase I digestion, the RNA preparations were cleaned up using the kit columns. In some cases, an additional DNase I digestion and the subsequent RNA purification (kit columns) were performed. The integrity and yield of rRNAs were checked by agarose gel electrophoresis in all the RNA preparations. Finally, the concentration of the RNA preparations was determined using a NanoDrop ND-2000 Spectrophotometer (Thermofisher).

7.3. PRIMER EXTENSION

Primer extension assays were carried out using the ThermoScript Reverse Transcriptase kit (Invitrogen) and total RNA isolated from different strains. Specific primers ^{32}P-labelled at the 5' end (described in Methods, Section 4.7.1) and non-radiolabelled primers ([α-^{32}P]-dATP, described in Methods, Section 4.7.2) were used in primer extension experiments. Generally, reaction mixtures (12 µl) contained 1-2 pmol of primer, 1 mM of dNTPs and 2.5-15 µg of total RNA. The mixture was incubated at 65°C for 5 min (annealing). Then, cDNA Synthesis buffer (supplied in the kit), 15 units of ThermoScript Reverse Transcriptase and 5 mM DTT were added. Extension reactions were carried out at 55°C for 45 min. When non-radiolabelled primers were used, 9.75 µM of dATP and 0.25 µM [α-^{32}P]-dATP, 100 µM of dCTP, dGTP and dTTP were added to the extension reactions. Reactions were stopped by heating at 85°C for 5 min, and when the extension reaction had been carried out with non-radiolabelled primers, the non-incorporated nucleotides were removed using Illustra MicroSpin™ G-25 columns (GE Healthcare). Finally, samples were ethanol precipitated and dissolved in BXF buffer (Table 5). The cDNA products were subjected to electrophoresis in an 8 M urea and 6% polyacrylamide sequencing gel. To estimate the length of the extension products, sequencing

reactions obtained by the Sanger method were run in the same gel. Labelled products were visualized using a Fujifilm Image Analyzer FLA-3000.

7.4. REVERSE TRANSCRIPTION-POLYMERASE CHAIN REACTION (RT-PCR)

RT-PCR assays were carried out using both, the ThermoScript Reverse Transcriptase kit (Thermofisher) or the iScript Select cDNA Synthesis kit (Bio-Rad), as recommended by the suppliers. Specifically, when the ThermoScript Reverse Transcriptase kit was used, 20 pmol of the specific primer were mixed with 200 ng of total RNA and 1 mM of dNTPs. The reaction mixture (12 µl) was incubated at 65°C for 5 min. Then, 5 mM DTT, cDNA Synthesis buffer (provided by the supplier) and the ThermoScript Reverse Transcriptase (15 units) were added. Reaction mixtures (20 µl) were incubated at 55°C for 45 min. After the extension step, the enzyme was inactivated (incubation at 85°C for 5 min).

When the iScript Select cDNA Synthesis kit was used, the reaction mixture (20 µl) contained 8 pmol of the specific primer, 250 ng of total RNA, 4 µl of 5X iScript Select Reaction mix, 2 µl GSP enhancer solution and 1 µl iScript Reverse Transcriptase enzyme. The reaction mixture was incubated at 25°C for 5 min and, then at 42°C for 30 min. After the extension step, the enzyme was inactivated (incubation at 85°C for 5 min). The resulting cDNA product was 3.5-fold diluted with sterile nuclease-free water.

PCRs were carried out as described in Section 4.3.2, using the cDNA as template (5-10% of the first-strand reaction) and Phusion High-Fidelity DNA Polymerase (Finnzymes). To ensure the absence of genomic DNA in the RNA preparations, the same reactions were performed without adding reverse transcriptase (negative control). As a positive control, PCRs were performed with chromosomal DNA as template. PCR products were analysed by agarose gel electrophoresis.

7.5. QUANTITATIVE REVERSE TRANSCRIPTION-POLYMERASE CHAIN REACTION (qRT-PCR)

Quantitative reverse polymerase chain reaction (qRT-PCR) was carried out in two steps. First, we performed cDNA synthesis from total RNA samples and, then, cDNA was used as a template in a quantitative PCR based on SYBR® Green I detection. The cDNA synthesis was carried out with the iScript Select cDNA Synthesis kit (Bio-Rad), as recommended by the supplier. The reaction mixture (20 µl) contained 1 µg of total RNA, 4 µl of 5X iScript Select Reaction mix, 2 µl of random primers and 1 µl iScript Reverse Transcriptase enzyme. The reaction mixture was incubated at 25°C for 5 min and, then at 42°C for 30 min. After the extension step, the enzyme was inactivated (incubation at 85°C for

5 min). The resulting cDNA product was 3.5-fold diluted with sterile nuclease-free water and stored at -80°C. Negative controls (same reactions without adding Reverse Transcriptase enzyme) were performed to ensure the absence of DNA.

The oligonucleotides for quantitative PCR (qPCR) were designed using the published genome sequences of the strains *E. faecalis* OG1RF (Bourgogne *et al.*, 2008) and *S. pneumoniae* R6 (Hoskins *et al.*, 2001) and the Primer3 software (Rozen and Skaletsky, 2000) or Primer-BLAST program (Ye *et al.*, 2012) to synthesize amplicons of similar lengths (100-150 bp). The analysis of each primer was performed with the OligoAnalizer 3.1 tool available at https://eu.idtdna.com and its specificity was verified using the Basic Local Alignment Search Tool BLAST NCBI tool (Johnson *et al.*, 2008).

iQ SYBR Green Supermix (Bio-Rad) was used to perform qPCRs, on an iCycler Thermal Cycler (Bio-Rad). PCR reactions were carried out using as template 3 µl of cDNA synthesized with random primers. Moreover, negative controls of the reverse transcriptions reactions were used as template in PCR reactions. According to the manufacturer's instruction, the reaction mixtures (20 µl) contained 10 µl of iQ SYBR Green Supermix 2X and 500 nM of each specific primer. The protocol conditions were: initial denaturation step at 95°C for 5 min, followed by 40 cycles that included the next steps: (i) denaturation at 95°C for 30 s; (ii) annealing at 55°C or 58°C for 30 s; and (iii) extension at 72°C for 20 s. The final step was a melt curve analysis to confirm the specificity of the selected primers. Data were analysed with the software iQ^{TM}5 Optical System Software. Relative quantification of gene expression was performed following the comparative method C_T (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008). *E. faecalis* internal control genes used were *recA* gene (OG1RF_*12439*) and *zwf* gene (OG1RF_*10737*). *S. pneumoniae* internal control genes used were *rpoB* gene (*spr1777*) and *era* gene (*spr0871*). The threshold cycles values (C_T) of the genes of interest and the control gene were used to calculate $2^{-\Delta C_T}$, where ΔC_T = (C_T gene of interest – C_T internal control). For a particular gene, the fold change in expression (FC) between two strains was obtained dividing the corresponding $2^{-\Delta C_T}$ values.

8. PROTEIN PURIFICATION

All proteins were overproduced and purified using the commercially available pET24b expression vector (Novagen). In this vector, the gene of interest is expressed under the control of the Φ10 promoter of the phage T7 using the *E. coli* BL21 (DE3) strain. This strain contains a chromosomal copy of the *lacI* gene and the T7 RNA polymerase (RNAP) encoding gene (T7 gene) under the control of the *lac*UV5 promoter. In the absence of IPTG, the LacI repressor binds to the operator region (*lac*O) of the *lac*UV5 promoter and represses the transcription of the T7 gene. In the presence of IPTG, LacI is

blocked and, thus, T7 RNAP is synthesized. T7 RNAP transcribes the gene of interest from the Φ10 promoter (Figure 8).

8.1. PURIFICATION OF MafR$_{OG1RF}$-HIS, MafR$_{OG1RF}$Δ3N-HIS AND MafR$_{V583}$-HIS

E. coli BL21 (DE3) cells harbouring the pET24b-*mafR*$_{OG1RF}$-His, pET24b-*mafR*$_{OG1RF}$Δ3N-His and pET24b-*mafR*$_{V583}$-His plasmids were used to overproduce the MafR$_{OG1RF}$-His, MafR$_{OG1RF}$Δ3N-His and MafR$_{V583}$-His proteins, respectively. Cells carrying the plasmids were grown at 37°C with rotary shaking in TY broth containing Km (30 μg/ml). When the culture reached an OD_{600} of 0.4-0.5, the gene of interest (*mafR*$_{OG1RF}$-His, *mafR*$_{OG1RF}$Δ3N-His or *mafR*$_{V583}$-His) was induced by the addition of 1 mM IPTG. After 25 min at 37°C, the culture was treated with rifampicin (200 μg/ml) for 60 min, which specifically inhibits bacterial RNAP. Cells were collected by centrifugation (8,000 rpm in an SLA-3000 rotor for 15 min at 4°C), washed twice with buffer VL (Table 5) containing 200 mM NaCl and stored at -80°C. The cell pellet was concentrated (40X) in buffer VL containing 200 mM NaCl and protease inhibitor cocktail (Roche) was added. Cells were lysed by two passages through a pre-chilled cell-pressure French Press, and the whole-cell extract was centrifuged (10,000 rpm in an Eppendorf F-34-6-38 rotor for 40 min at 4°C) to remove cell debris. The clarified extract was mixed with 0.2% polyethyleneimine (PEI), kept on ice for 30 min, and centrifuged at 9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4°C. Under these conditions, nucleic acids and some proteins, including the protein of interest, precipitated. Then, pellet was washed twice with buffer VL containing 200 mM NaCl to remove any trapped protein (low ionic strength). After that, the pellet was washed with buffer VL containing 500 mM NaCl to elute the MafR$_{OG1RF}$-His protein or MafR$_{OG1RF}$Δ3N-His protein or MafR$_{V583}$-His protein in each case. Afterwards, proteins were precipitated from the supernatant with 70% (w/v) saturated ammonium sulphate, which was added slowly. The mixture was kept on ice with constant stirring for 60 min, followed by a centrifugation step (9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4°C). The precipitate was dissolved in buffer S-His 1X (Table 5) and then dialyzed (Spectra/Por® 4, Repligen) at 4°C against the same buffer to remove both ammonium sulphate and EDTA. Following, dialyzed samples were centrifuged at 10,000 rpm in an Eppendorf F-34-6-38 rotor for 30 min at 4°C. Imidazole (10 mM) was added to the sample supernatant, then it was loaded onto a nickel affinity column (HisTrap HP column, GE Healthcare) pre-equilibrated with buffer S-His 1X containing 10 mM imidazole and subjected to fast-pressure liquid chromatography (FPLC; Biologic DuoFlow; Bio-Rad). After washing with the same buffer, the protein of interest (MafR$_{OG1RF}$-His protein or MafR$_{OG1RF}$Δ3N-His protein or MafR$_{V583}$-His) was eluted using a linear gradient of buffer S-His 1X containing from 10 mM to 250 mM imidazole. Fractions were analysed by electrophoresis on SDS polyacrylamide gels (12%), followed by staining with Coomassie Brilliant Blue. Fractions containing

the protein of interest were pooled, dialyzed (Spectra/Por® 4, Repligen) against VL containing 200 mM NaCl, concentrated by filtering through a 10-kDa-cutoff (VivaSpin 20, Fisher Scientific) and stored at -80°C.

8.2. PURIFICATION OF MgaP-HIS

E. coli BL21 (DE3) cells harbouring the pET24b-*mgaP*-His plasmid were used to overproduce the MgaP-His protein. Cells carrying the pET24b-*mgaP*-His plasmid were grown at 37°C with rotary shaking in TY broth containing Km (30 μg/ml). When the culture reached an OD_{600} of 0.4, the *mgaP*-His gene was induced by addition of 0.5 mM IPTG. After 25 min at 37°C, the culture was treated with rifampicin (200 μg/ml) for 60 min, which specifically inhibits bacterial RNAP. Cells were collected by centrifugation (8,000 rpm in an SLA-3000 rotor for 15 min at 4°C), washed twice with buffer VL (Table 5) containing 200 mM NaCl and stored at -80°C. The cell pellet was concentrated (40X) in buffer VL containing 300 mM NaCl and protease inhibitor cocktail (Roche) was added. Cells were lysed by two passages through a pre-chilled cell-pressure French Press, and the whole-cell extract was centrifuged (10,000 rpm in an Eppendorf F-34-6-38 rotor for 40 min at 4°C) to remove cell debris. The clarified extract was mixed with 0.2% polyethyleneimine (PEI), kept on ice for 30 min, and centrifuged at 9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4°C. Under these conditions, nucleic acids and some proteins, including the protein of interest, precipitated. Then, pellet was washed twice with buffer VL containing 300 mM NaCl to remove any trapped protein (low ionic strength). After that, the pellet was washed with buffer VL containing 700 mM NaCl to elute the MgaP-His protein. Afterwards, proteins were precipitated from the supernatant with 70% (w/v) saturated ammonium sulphate, which was added slowly. The mixture was kept on ice with constant stirring for 60 min, followed by a centrifugation step (9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4°C). The precipitate was dissolved in buffer S-His-P 1X (Table 5) and then dialyzed (Spectra/Por® 4, Repligen) at 4°C against the same buffer to remove both ammonium sulphate and EDTA. After dialysis, samples were centrifuged at 10,000 rpm in an Eppendorf F-34-6-38 rotor for 30 min at 4°C. Imidazole (10 mM) was added to the supernatant and, it was loaded onto a nickel affinity column (HisTrap HP column, GE Healthcare) pre-equilibrated with buffer S-His-P 1X containing 10 mM imidazole and subjected to fast-pressure liquid chromatography (FPLC; Biologic DuoFlow; Bio-Rad). After washing with the same buffer, MgaP-His protein was eluted using a linear gradient of buffer S-His-P 1X containing from 10 mM to 250 mM imidazole. Fractions were analysed by electrophoresis on SDS polyacrylamide gels (12%), followed by staining with Coomassie Brilliant Blue. Fractions containing MgaP-His protein were pooled, dialyzed (Spectra/Por® 4, Repligen) against VL

containing 200 mM NaCl, concentrated by filtering through a 3-kDa-cutoff (VivaSpin 20, Fisher Scientific; Nanosep, Pall) and stored at -80°C.

9. PROTEIN ANALYSIS

9.1. DETERMINATION OF PROTEIN CONCENTRATION

The protein concentrations were measured using a NanoDrop ND-2000 Spectrophotometer (Thermofisher). The theoretical molecular weight (Da) and the molar extinction coefficient (M^{-1} cm^{-1}) were calculated from the amino acid sequence of the corresponding protein.

9.2. N-TERMINAL SEQUENCING

Both purified $MafR_{OG1RF}$-His protein and $MafR_{OG1RF}$Δ3N-His protein (in buffer VL containing 200 mM NaCl) were subjected to amino terminal-sequencing by Edman degradation using a Procise 494 Sequencer (Perkin-Elmer) (Protein Chemistry, CIB).

9.3. PROTEIN ELECTROPHORESIS: TRIS-GLYCINE SDS-PAGE

Protein samples were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The stacking gel contained 5% polyacrylamide in 0.125 M Tris-HCl, pH 6.8, and 0.1% SDS. The separating gel contained 12% polyacrylamide, otherwise indicated, in 0.39 M Tris-HCl, pH 8.8, and 0.1% SDS. Temed and APS were used to catalyze polyacrylamide gel polymerization. Before electrophoresis, protein samples were mixed with SLB (Table 5) and heated at 95°C for 5 min. Gels were run in a Mini-Protean III Electrophoresis System (Bio-Rad) using TG buffer (Table 5). Electrophoresis was carried out at 80 V until the dye (bromophenol blue) migrated down to the bottom of the stacking gel. Then, the voltage was increased to 180 V. Gels were stained with Coomassie Brilliant Blue R-250 (Bio-Rad).

10. DNA-PROTEIN INTERACTIONS

10.1. ELECTROPHORETIC MOBILITY SHIFT ASSAYS (EMSA)

Generally, standard binding reactions (16-20 µl) were performed mixing 10 nM of unlabelled DNA or 2-4 nM of 5'-labelled DNA with different amounts of the purified protein ($MafR_{OG1RF}$-His or $MafR_{OG1RF}$Δ3N-His or $MafR_{V583}$-His or MgaP-His protein). The binding buffer contained 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 50 mM NaCl, 1% glycerol, and 0.5 mg/ml BSA. After 20 min incubation at room temperature, BXGE buffer was added and free and bound DNA forms were

separated by electrophoresis on a native polyacrylamide gel (6% polyacrylamide) using TBE buffer (Table 5). Gels were pre-run (20 min) and run at 100 V at room temperature. Labelled DNA was visualized using a Fujifilm Image Analyzer FLA-3000. EMSA gels containing unlabelled DNA were stained in a solution of GelRed (1X in water supplemented with 0.1 M NaCl; Biotium) for 15 min at room temperature. Finally, DNA was visualized using a short wavelength UV light (254 nm) with a Gel Doc XR system (Bio-Rad). The image obtained was captured with the Quantity One software (Bio-Rad).

In dissociation experiments, different amounts of the non-labelled competitor calf thymus DNA or heparin glycosaminoglycan were added after the formation of protein-DNA complexes (binding reaction under standard conditions). Then, the reaction mixtures were incubated for 5 min (dissociation of protein-DNA complexes).

10.2. DNase I FOOTPRINTING ASSAYS

The DNase I footprinting assay is based on the enzymatic digestion of the DNA using DNase I in the presence of a DNA binding protein (Galas and Schmitz, 1978). Binding reactions (8 µl) contained 2-4 nM of the 5'-labelled DNA fragment, 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50 mM NaCl, 0.5 mg/ml BSA, 1 mM $CaCl_2$, 10 mM $MgCl_2$ and different concentrations of the $MafR_{OG1RF}$-His, MgaP-His or Mga*Spn*-His proteins. The reaction mixtures were incubated at room temperature for 20 min. DNase I (1 unit/µl) was diluted in 50% glycerol to 1:750 or 1:300. Then, 1 µl of the corresponding dilution was added to the reaction mixtures. After incubation for 5 min at room temperature, reactions were stopped by adding 1 µl of EDTA 0.25 M, following by 4 µl of BXF (Table 5). Then samples were heated at 95°C for 5 min and loaded onto sequencing gels (6% polyacrylamide, 8 M urea).

11. FLUORESCENCE MEASUREMENTS

Enterococcal and pneumococcal cells harbouring derivatives of pASTT or pAST (containing the *gfp* reporter gene) were grown to an OD_{650} of 0.3-0.4. Then, different volumes of the culture (from 0.2 ml to 1 ml) were centrifuged at 4°C for 10 min. An additional centrifugation at 4°C for 1 min was carried out to remove all the supernatant. Cells were resuspended in 200 µl of PBS buffer (Table 5). Finally, fluorescence was measured on a Thermo Scientific Varioskan Flash instrument (Perkin-Elmer) by excitation at 488 nm and detection of emission at 515 nm. In all cases, at least three independent cultures were analysed and the fluorescence corresponding to 200 µl of PBS buffer without cells was also measured.

RESULTS

CHAPTER 1

FIRST STUDIES ON THE DNA-BINDING PROPERTIES OF MafR ENCODED BY THE *E. faecalis* OG1RF GENOME

In this Chapter, we describe the procedure used to overproduce and purify (i) a His-tagged version of the MafR protein encoded by the *E. faecalis* OG1RF genome ($MafR_{OG1RF}$-His), and (ii) a variant of $MafR_{OG1RF}$-His that lacks the first three amino acid residues ($MafR_{OG1RF}$Δ3N-His). By gel retardation experiments, we show that $MafR_{OG1RF}$-His, but not $MafR_{OG1RF}$Δ3N-His, generates multimeric complexes on linear dsDNAs. In addition, we show that MafR is highly conserved among the strains whose genomes have been totally or partially sequenced. These results were included in a manuscript focused on the DNA-binding properties of MafR. The manuscript was published in 2018 (Ruiz-Cruz *et al.*, 2018).

1.1. Protein $MafR_{OG1RF}$-His

The genome sequence of the *E. faecalis* strain V583, the first vancomycin-resistant clinical isolate reported in the United States, was published in 2003 (Paulsen *et al.*, 2003). This article revealed that over a quarter of the V583 genome consists of mobile and/or exogenously acquired DNA, which is one of the highest proportions observed in a bacterial genome. Five years later, the genome sequence of the *E. faecalis* strain OG1RF, a rifampicin- and fusidic acid-resistant derivative of the human isolate OG1, was published (Bourgogne *et al.*, 2008). The OG1RF strain is widely used in the laboratories. Both genomes, V583 and OG1RF, encode the MafR protein, a member of the Mga/AtxA family of global regulators (Ruiz-Cruz *et al.*, 2016). Previous *in vitro* studies performed in our laboratory showed that MafR from strain V583 (here named $MafR_{V583}$) binds to linear dsDNAs with little or no sequence specificity. Moreover, $MafR_{V583}$ is able to generate multimeric protein-DNA complexes (Ruiz-Cruz, 2015).

In the OG1RF genome, the *mafR* gene (locus_tag *OG1RF_12293*) spans coordinates 2421605 to 2423053 (Figure 7). Translation from the ATG codon at coordinate 2421605 would generate a protein of 482 residues (here named $MafR_{OG1RF}$), as there is a TAA stop codon at coordinate 2423053. Downstream of the stop codon, there is a putative Rho-independent transcriptional terminator (coordinates 2423059 to 2423094). The *mafR* gene is flanked by two uncharacterized genes: *OG1RF_12292* (hypothetical membrane protein, coordinates 2420308 to 2421513) and *OG1RF_12294* (putative phosphorylated intermediate-type ATPase (P-type ATPase) transporter, coordinates 2425611 to 2423059). The *OG1RF_12294* and *mafR* genes are located in different strands.

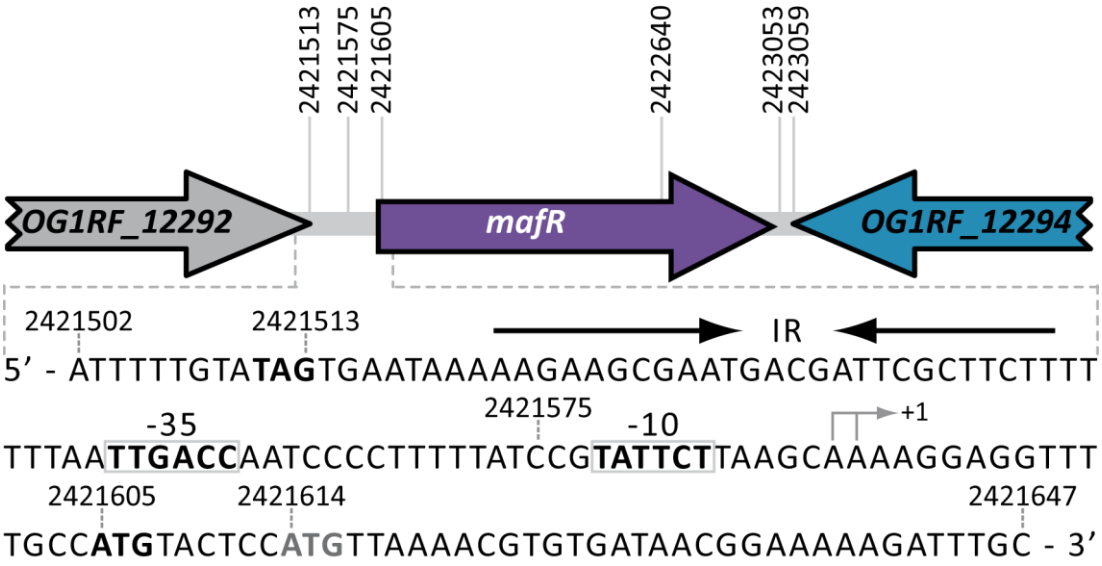


Figure 7. Genetic organization of the *mafR* region in the *E. faecalis* OG1RF genome. The nucleotide sequence of the region spanning coordinates 2421502 to 2421647 is shown. The coordinates of the predicted *mafR* translation start codons (first ATG start codon in black; second ATG start codon in grey), and the *OG1RF_12292* translation stop codon (TAG), are indicated in boldface letters. The transcription start site (+1 position) of the *mafR* gene and the main sequence elements (-35 box and -10 box) of the *Pma* promoter (Ruiz-Cruz *et al.*, 2016) are indicated. IR: inverted-repeat. The chromosomal region that spans coordinates 2421575 to 2422640 was deleted to construct the OG1RFΔ*mafR* strain (see Methods, Section 3.1) (Ruiz-Cruz *et al.*, 2016).

The OG1RF genome, compared to the V583 genome, contains 227 unique open reading frames but has fewer mobile genetic elements (Bourgogne *et al.*, 2008). Despite this difference between both genomes, the nucleotide sequence of the region that spans coordinates 2421450 to 2421621 in OG1RF (Figure 7) is identical in V583 (coordinates 2888932 to 2889103). Such a region contains the promoter of the *mafR* gene (*Pma* promoter). However, compared to $MafR_{V583}$, the $MafR_{OG1RF}$ protein has five amino acid changes (Ala37Thr, Gln131Leu, Met145Thr, Ser193Asn, and Ile388Ser), and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the predicted DNA-binding domain (residues 11 to 164). To analyse whether these changes affected the formation of *in vitro* protein-DNA complexes, we purified a His-tagged $MafR_{OG1RF}$ protein ($MafR_{OG1RF}$-His). This variant of $MafR_{OG1RF}$ carries the Leu-Glu-6xHis peptide (His-tag) fused to its C-terminus. We used this experimental approach because previous studies allowed us to conclude that the presence of a His-tag at the C-terminal end of $MafR_{V583}$ does not affect its ability to generate multimeric complexes (Ruiz-Cruz, 2015). The procedure used to overproduce and purify $MafR_{OG1RF}$-His was set up previously in our laboratory (Moreno-Blanco, 2016) and it has been extensively described in Methods, Section 8.1. Briefly, to overproduce $MafR_{OG1RF}$-His, the *mafR* gene from strain OG1RF was cloned into the *E. coli* inducible expression vector pET24b, which is based on a promoter recognized by the bacteriophage T7 RNA polymerase (T7 RNAP) (Figure 8). The recombinant plasmid, pET24b-$mafR_{OG1RF}$-His was then introduced into the *E. coli* BL21 (DE3) strain, which carries the T7 RNAP-encoding gene (T7 gene)

fused to the *lac*UV5 promoter. Both plasmid pET24b and the BL21 (DE3) chromosome carry the *lacI* repressor gene. In this system, expression of the T7 gene, and consequently expression of the *mafR*$_{OG1RF}$-His gene, is induced when IPTG is added to the bacterial culture. The method used for large-scale purification of MafR$_{OG1RF}$-His involved essentially three steps: (i) precipitation of nucleic acids and MafR$_{OG1RF}$-His with PEI using a low ionic strength buffer (0.2 M NaCl); (ii) elution of MafR$_{OG1RF}$-His from the PEI pellet using a higher ionic strength buffer (0.5 M NaCl), and (iii) fast-pressure liquid chromatography on a nickel affinity column. In each step of the purification process, protein fractions were analysed by electrophoresis using Coomassie-stained SDS polyacrylamide (12%) gels (Figure 9A). Under these conditions, MafR$_{OG1RF}$-His migrates between the 45 and 66 kDa bands of the molecular weight marker (Figure 9A), which agrees with the theoretical molecular weight of the MafR$_{OG1RF}$-His monomer (57.3 kDa; 490 residues). Determination of the N-terminal amino acid sequence (the first eight residues) by Edman degradation showed that the first Met residue of MafR$_{OG1RF}$-His was not removed after protein synthesis.

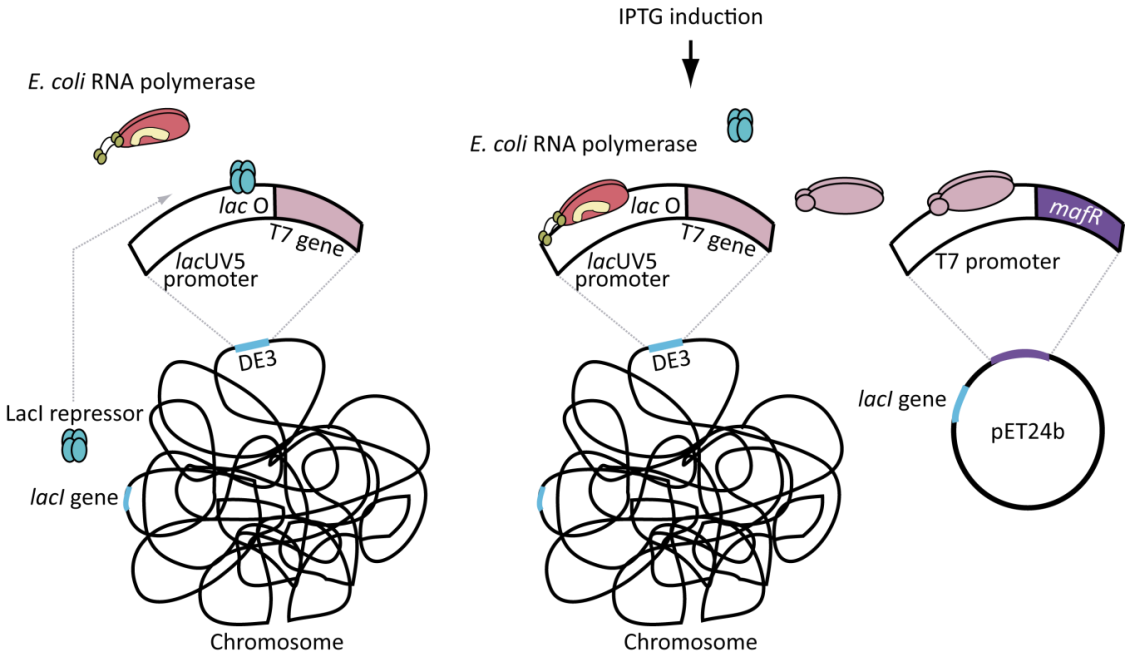


Figure 8. Experimental design to overproduce MafR$_{OG1RF}$-His. The *mafR* gene from strain OG1RF was cloned into the *E. coli* inducible expression vector pET24b, which is based on the Φ10 promoter of the bacteriophage T7 (T7 promoter). The T7 RNAP recognizes specifically this promoter. The recombinant plasmid, pET24b-*mafR*$_{OG1RF}$-His, was introduced into the *E. coli* BL21 (DE3) strain, which carries the T7 RNAP-encoding gene (T7 gene) under the control of the IPTG-inducible *lac*UV5 promoter. Besides, both plasmid pET24b and host strain, BL21 (DE3), carry a copy of the *lacI* gene. In the absence of IPTG, the LacI repressor binds to the operator region (*lac*O) of the *lac*UV5 promoter and represses the transcription of the T7 gene. When the inductor, IPTG, is added, it displaces the LacI repressor from the *lac*O region, allowing the binding of the *E. coli* RNAP and, therefore, the expression of the T7 gene. The T7 RNAP recognizes the T7 promoter and the *mafR*$_{OG1RF}$ gene is transcribed (Methods, Section 8.1). Figure adapted from Figure 10 (Solano-Collado, 2014).

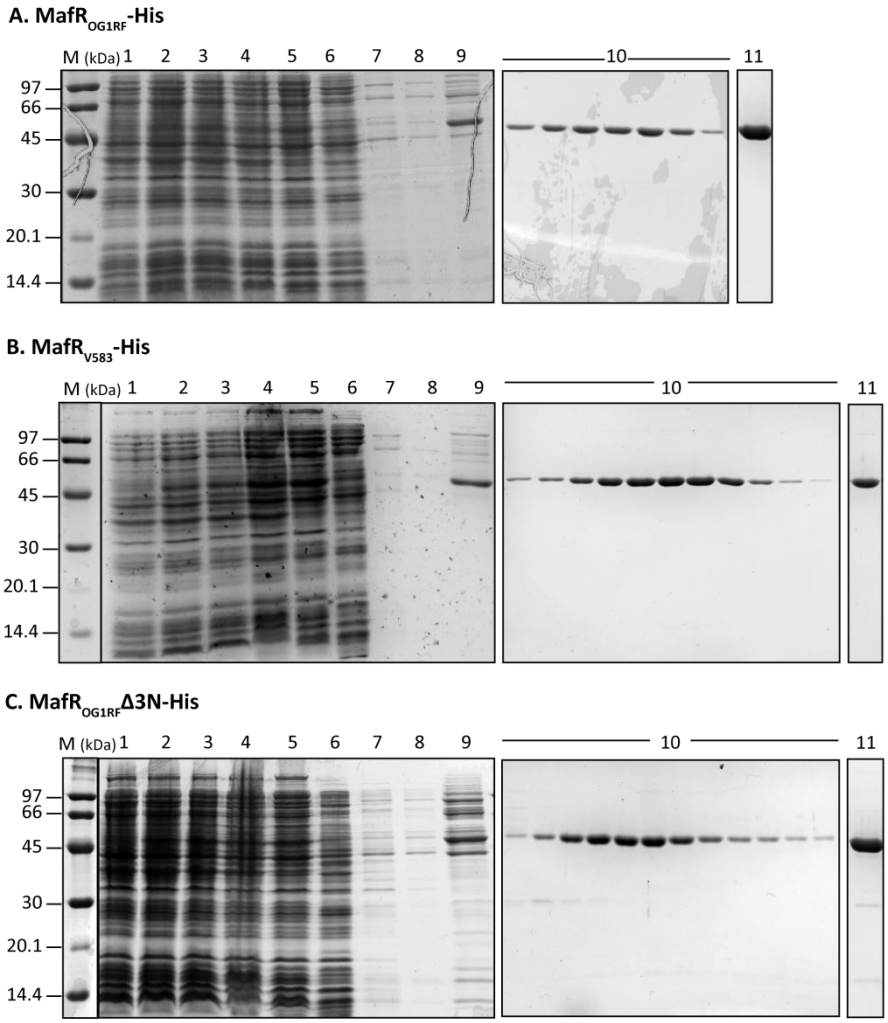


Figure 9. Purification of MafR-His proteins. (A) MafR$_{OG1RF}$-His, (B) MafR$_{V583}$-His, and (C) MafR$_{OG1RF}$Δ3N-His. Protein fractions were analysed by SDS-polyacrylamide (12%) gel electrophoresis. **Lanes 1 to 3**: Induction of *mafR* gene expression: (1) uninduced cultures; (2) with IPTG for 25 min; (3) after treatment with rifampicin for 60 min. **Lanes 4 to 11**: purification steps: (4) supernatant of a clear lysate; (5) supernatant after PEI precipitation at low ionic strength; (6-8) proteins eluted from the PEI pellet using the same low ionic strength buffer; (9) proteins eluted from the PEI pellet using a higher ionic strength buffer; (lanes group 10) nickel affinity chromatography: proteins were eluted with a 10 mM to 250 mM imidazole gradient; (11) final protein preparation after concentration. M indicates the molecular weight standards (in kDa) (LMW Marker, GE Healthcare).

Protein MafR$_{V583}$-His was purified using the same procedure described for MafR$_{OG1RF}$-His (Methods, Section 8.1). As shown in Figure 9B, MafR$_{V583}$-His migrates between the 45 and 66 kDa bands of the molecular weight marker, which agrees with the theoretical molecular weight of the MafR$_{V583}$-His monomer (57.34 kDa; 490 residues). This protein was used as a control in the following EMSA experiments (Methods, Section 10.1):

(a) Since MafR$_{OG1RF}$ activates *in vivo* the transcription of the *gldA* gene (glycerol metabolism) (Ruiz-Cruz *et al.*, 2016), we performed EMSA experiments using a 345-bp DNA fragment that contains the

gldA promoter (coordinates 1197161 to 1197505 of the OG1RF genome). This DNA fragment was incubated with increasing concentrations of $MafR_{OG1RF}$-His. Then, free and bound DNA forms were separated by electrophoresis on native polyacrylamide (6%) gels (Figure 10A). At 50 nM of $MafR_{OG1RF}$-His, free DNA and a protein-DNA complex (C1) were observed. However, as the protein concentration was increased, complexes of lower electrophoretic mobility appeared sequentially and faster-moving complexes disappeared gradually. This pattern of complexes indicated that multiple units of $MafR_{OG1RF}$-His bind orderly on the same linear DNA molecule generating multimeric complexes. As shown in Figure 10B, $MafR_{V583}$-His (used as control) was also able to form multimeric complexes on the 345-bp DNA fragment.

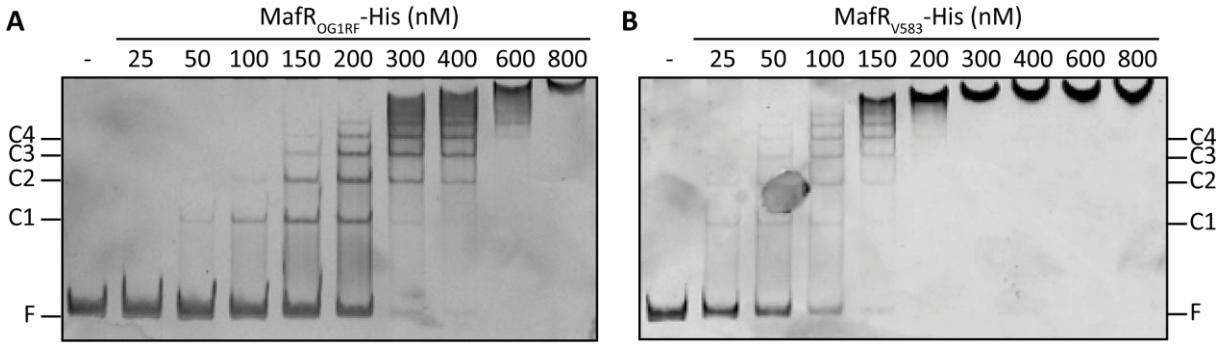


Figure 10. Formation of multimeric protein-DNA complexes. (A) EMSA using $MafR_{OG1RF}$-His, and (B) EMSA using $MafR_{V583}$-His. A 345-bp DNA fragment (*gldA* promoter, coordinates 1197161 to 1197505) was used in both A and B experiments. The DNA fragment (10 nM) was mixed with increasing concentrations (25 to 800 nM) of $MafR_{OG1RF}$-His (A) or $MafR_{V583}$-His (B). Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated.

(b) Since $MafR_{V583}$ and $MafR_{V583}$-His generate multimeric complexes on linear dsDNAs that contain the *Pma* promoter (Ruiz-Cruz, 2015), we also performed EMSA experiments with the $MafR_{OG1RF}$-His protein using a 217-bp DNA fragment that contains the *Pma* promoter (coordinates 2421450 to 2421666 of the OG1RF genome) (Figure 11). This experiment confirmed that $MafR_{OG1RF}$-His is able to form multimeric complexes.

Taken together, our EMSA results showed that the DNA-binding behaviour of $MafR_{OG1RF}$-His is similar to the one described for $MafR_{V583}$ and $MafR_{V583}$-His (Ruiz-Cruz, 2015). Hence, we conclude that the three amino acids substitutions found within the predicted DNA-binding domain of $MafR_{OG1RF}$ (Ala37Thr, Gln131Leu and Met145Thr; compared to $MafR_{V583}$) do not affect its interaction with DNA. Further DNase I footprinting experiments using linear DNA fragments that contain particular promoters showed that $MafR_{OG1RF}$-His recognizes specific DNA sites (see Results, Chapter 2).

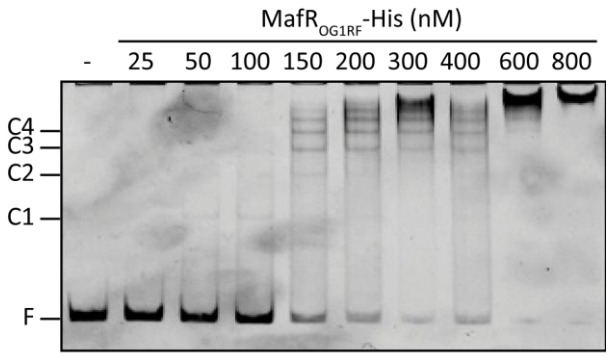


Figure 11. Binding of MafR$_{OG1RF}$-His to a DNA fragment that contains the *Pma* promoter. EMSA using MafR$_{OG1RF}$-His. 10 nM of the 217-bp DNA fragment (*Pma* promoter, coordinates 2421450 to 2421666) was mixed with increasing concentrations (25 to 800 nM) of MafR$_{OG1RF}$-His. Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated.

1.2. Protein MafR$_{OG1RF}$Δ3N-His

Previous studies performed in our laboratory identified the transcription initiation site of the *mafR*$_{OG1RF}$ gene at coordinate 2421589, and suggested that the first ATG codon at coordinate 2421605 is likely the translation start site (Figure 7) (Ruiz-Cruz, 2015). Translation from this ATG generates a protein of 482 residues (MafR$_{OG1RF}$). However, there is a second ATG codon (Figure 7, grey start codon) at coordinate 2421614 that might function as a translation start site. Translation from this second ATG would result in a variant (here named MafR$_{OG1RF}$Δ3N) that lacks the first three amino acid residues (Met-Tyr-Ser). To analyse whether the lack of these three residues impaired the formation of protein-DNA complexes, we purified a His-tagged MafR$_{OG1RF}$Δ3N protein (MafR$_{OG1RF}$Δ3N-His) using the same procedure described for MafR$_{OG1RF}$-His (Methods, Section 8.1; Results, Chapter 1.1). The MafR$_{OG1RF}$Δ3N-His protein migrates between the 45 and 66 kDa bands of the molecular weight marker (Figure 9C), which agrees with the theoretical molecular weight of the MafR$_{OG1RF}$Δ3N-His monomer (56.91 kDa; 487 residues). Determination of the N-terminal amino acid sequence by Edman degradation showed that the first Met residue of MafR$_{OG1RF}$Δ3N-His was not removed after protein synthesis.

Next, we carried out EMSA experiments using (i) the 345-bp DNA fragment that contains the *gldA* promoter (Figure 12A), and (ii) the 217-bp DNA fragment that contains the *Pma* promoter (Figure 12B). In both experiments increasing concentrations of MafR$_{OG1RF}$-His protein (100 to 800 nM) and MafR$_{OG1RF}$Δ3N-His protein (100 to 1600 nM) were used. As expected, multiple DNA-protein complexes were observed when the DNA fragments were incubated with MafR$_{OG1RF}$-His. However, no DNA-protein complexes were detected when the DNA fragments were incubated with MafR$_{OG1RF}$Δ3N-

His. These results indicated that MafR$_{OG1RF}$Δ3N-His has lost the capacity to interact with DNA. Thus, we conclude that (i) the first ATG codon (coordinate 2421605) is the translation start site, and (ii) the first three amino acid residues of MafR$_{OG1RF}$ are crucial for its structure and/or function.

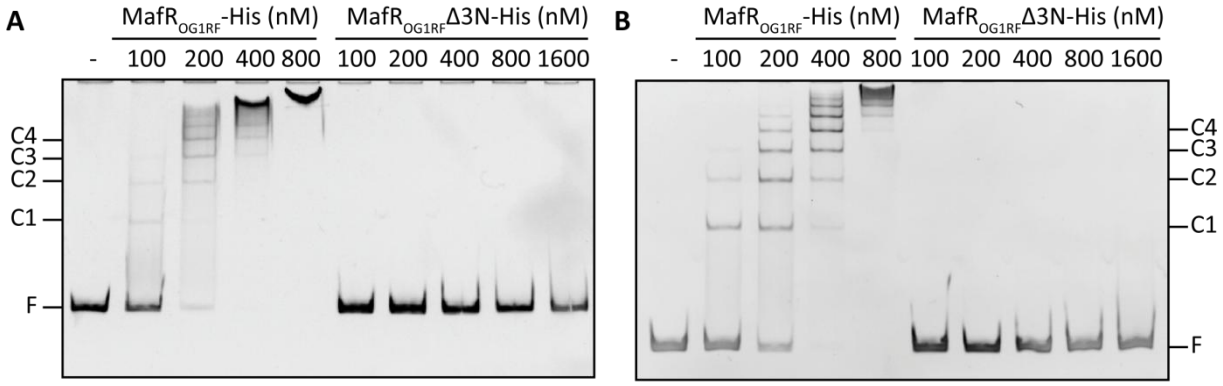


Figure 12. MafR$_{OG1RF}$Δ3N-His does not bind to DNA. EMSA analysis of the ability of MafR$_{OG1RF}$Δ3N-His to interact with DNA using **(A)** the 345-bp DNA fragment (*gldA* promoter, coordinates 1197161 to 1197505), and **(B)** the 217-bp DNA fragment (*Pma* promoter, coordinates 2421450 to 2421666). Each DNA fragment (10 nM) was mixed with the indicated concentration of MafR$_{OG1RF}$Δ3N-His or MafR$_{OG1RF}$-His. Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). Bands corresponding to free DNA (F) and to several MafR$_{OG1RF}$-DNA complexes (C1 to C4) are indicated.

1.3. MafR encoded by other *E. faecalis* genomes

As mentioned above, the amino acid sequence of MafR$_{V583}$ differs from that of MafR$_{OG1RF}$ in five residues (Ala37Thr, Gln131Leu, Met145Thr, Ser193Asn, and Ile388Ser), and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the predicted DNA-binding domain (See Figure 6 and Table 8). Despite this difference, we have shown that both proteins generate multimeric complexes on linear dsDNAs, which could be an indication of functional conservation. Similar DNA-binding properties are expected for MafR encoded by the genomes of the *E. faecalis* strains that belong to the groups GA2, KS19, MTmid8, B653, X98, AZ19 and Com1 (Supplementary material, Table S1). Compared to MafR$_{V583}$, MafR from such strains has one to four amino acid substitutions, and all of them are present in MafR$_{OG1RF}$ (Table 8).

At the time of writing this Thesis, 1,521 *E. faecalis* genomes have been totally or partially sequenced (National Center for Biotechnology Information, NCBI, Genome Assembly and Annotation Report, 05/04/2020). Among them, we have found that 196 *E. faecalis* genomes encode a protein that is identical to MafR$_{V583}$ (Supplementary material, Table S2). Moreover, we have found that 162 *E. faecalis* genomes encode a protein that is identical to MafR$_{OG1RF}$ (Supplementary material, Table

S3). Hence, the MafR regulator seems to be highly conserved among the enterococcal genomes that have been sequenced.

Table 8. Amino acid substitutions in the indicated MafR proteins compared to $MafR_{V583}$. Data from the National Center for Biotechnology Information (Identical Protein Groups) (22/12/2017 and 05/04/2020)

Strain[a]	Amino Acid Residue					Identical[b]	
	37	**131**	**145**	**193**	**388**	22/12/2017	05/04/2020
V583	A	Q	M	S	I	146	196
GA2	T					2	5
KS19	T		T			6	8
MTmid8			T		S	2	19
B653	T	L	T			1	2
X98	T		T		S	15	53
AZ19	T	L	T	N		8	15
Com1	T	L	T		S	7	32
OG1RF	T	L	T	N	S	50	162

[a]Name of the *E. faecalis* strain that represents the group.
[b]Number of strains that belong to the group (identical MafR) (Tables S1, S2, and S3).

CHAPTER 2

GENES REGULATED DIRECTLY BY THE *E. faecalis* MafR PROTEIN

The *E. faecalis* MafR protein causes genome-wide changes in the transcriptome (Ruiz-Cruz *et al.*, 2016). In this Chapter, we show by qRT-PCR, primer extension, transcriptional fusions and DNase I footprinting assays, that MafR activates directly the transcription of three genes: *OG1RF_12294*, *OG1RF_11486* and *OG1RF_10478*. These genes are predicted to have a role in calcium transport, uptake of a queuosine precursor and cellobiose transport, respectively. The results concerning the *OG1RF_12294* and *OG1RF_11486* genes were published in 2019 (Ruiz-Cruz *et al.*, 2019) and showed, for the first time, that MafR functions as a transcription activator.

2.1. Expression of *mafR* under laboratory conditions

To know whether the regulatory *mafR* gene was expressed under laboratory conditions, we determined its relative expression in OG1RF cells by qRT-PCR assays (Methods, Section 7.5) and using the comparative C_T method (Schmittgen and Livak, 2008). To this aim, total RNA was extracted from OG1RF cells grown under standard conditions (BHI broth, 37°C, without aeration) to both logarithmic and stationary phases (Methods, Section 1). Figure 13 shows the corresponding bacterial growth curve. Transcription of *mafR* was found to be higher at the logarithmic phase (OD_{650} of 0.4). Compared to stationary phase (two hours after reaching an OD_{650} of 0.8), the fold change (log_2FC) in *mafR* expression was ~4. Thus, since the expression of the *mafR* gene was higher at exponential phase, all the experiments shown in Chapter 2 were performed at such a phase.

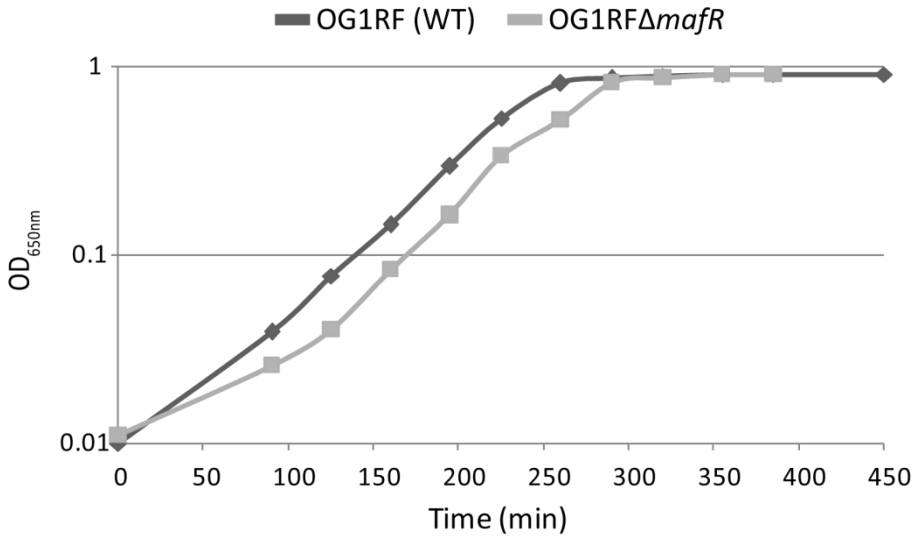


Figure 13. *E. faecalis* growth curves. Bacteria were grown in BHI liquid medium at 37°C and without aeration. Optical density (OD) was measured at 650 nm using a Bausch & Lomb (Spectronic 20D+) spectrophotometer, at intervals of 30 min. The dark grey line corresponds to strain OG1RF (wild-type) and the light grey line corresponds to strain OG1RFΔ*mafR*.

2.2. Selection of potential MafR target genes

Previous genome-wide microarray assays performed in our laboratory showed that MafR is involved in global regulation of gene expression (Ruiz-Cruz *et al.*, 2016). In such experiments, the transcriptional profiles of strains OG1RF (wild-type) and OG1RFΔ*mafR* (*mafR* deletion mutant) grown to mid-log phase under standard laboratory conditions (see Figure 13) were compared. Numerous genes were significantly differentially expressed (*P*-value < 0.05) in the presence of MafR (Ruiz-Cruz *et al.*, 2016). Among those genes, the *OG1RF_10478* gene of unknown function was up-regulated in cells that synthesize MafR: the fold change in gene expression (log_2FC) was ~2. In this Thesis, we have investigated whether MafR stimulates directly the transcription of the *OG1RF_10478* gene (Results, Chapter 2.5).

Furthermore, additional genome-wide microarray experiments using an OG1RFΔ*mafR* derivative that produces high levels of MafR (plasmid-encoded MafR) allowed us to identify two new potential MafR target genes: *OG1RF_12294* and *OG1RF_11486*. In the presence of plasmid-encoded MafR, the highest increase in gene expression corresponded to both genes (S. Ruiz-Cruz and A. Bravo, unpublished results). In this Thesis, we have addressed the validation of such a finding by several experimental approaches, both *in vivo* and *in vitro* (Results, Chapter 2.3 and 2.4).

2.3. Gene *OG1RF_12294*

According to DNA microarray assays (S. Ruiz-Cruz and A. Bravo, unpublished results), MafR could activate, directly or indirectly, the expression of the *OG1RF_12294* gene. This gene is adjacent to *mafR* (Figure 7; Results, Chapter 1.1) and encodes a putative P-type ATPase cation transporter. P-type ATPases constitute a large superfamily of cation and lipid pumps that uses ATP hydrolysis for energy. They are integral, multispanning membrane proteins that are found in bacteria and in many eukaryotic plasma membranes and organelles (Palmgren and Nissen, 2011). The *OG1RF_12294* gene has been annotated as *pmr1* (GenID: 12289043) because it encodes a protein of 850 residues that has sequence similarity (~52%) to eukaryotic PMR1 (plasma membrane ATPase related) P-type ATPases. Some PMR1-type pumps can transport calcium or manganese into the Golgi apparatus (Sorin *et al.*, 1997; Van Baelen *et al.*, 2001). Thus, protein OG1RF_12294, together with other calcium-transporters, might contribute to the maintenance of calcium homeostasis in enterococcal cells.

The OG1RF genome encodes others putative calcium-transporting ATPases, such as OG1RF_10600 and OG1RF_11602. We found that the OG1RF_12294 protein has sequence similarity (~53-57%) to

both ATPases, and also to several prokaryotic proteins characterized as calcium-P-type ATPases (Supplementary material, Table S4).

MafR INFLUENCES POSITIVELY THE TRANSCRIPTION OF THE *OG1RF_12294* GENE

To analyse whether MafR regulates the expression of the *OG1RF_12294* gene, we performed qRT-PCR assays (Methods, Section 7.5) using two strains, OG1RF (wild-type) and OG1RFΔ*mafR* (deletion mutant). The results indicated that the expression of the *OG1RF_12294* gene was up-regulated in the presence of MafR. The fold change in gene expression (log_2FC) was ~3. Further experiments confirmed this result. Specifically, we determined the relative expression of the *OG1RF_12294* gene in two strains: OG1RFΔ*mafR* harbouring plasmid pDLF (absence of MafR) and OG1RFΔ*mafR* harbouring plasmid pDLF*mafR* (plasmid-encoded MafR). As an internal control, we also determined the relative expression of two genes that encode putative calcium-transporting ATPases, the *OG1RF_10600* and *OG1RF_11602* genes. In the presence of plasmid-encoded MafR, the transcription of *OG1RF_12294* was increased (log_2FC ~4), whereas the expression of *OG1RF_10600* and *OG1RF_11602* was not affected. Hence, MafR influences positively and specifically the transcription of the *OG1RF_12294* gene.

MafR ACTIVATES THE *P12294* PROMOTER *IN VIVO*

To study whether MafR activated the promoter of the *OG1RF_12294* gene, we analysed the genetic organization of the chromosome region that contains this gene (Figure 14A). The ATG codon at coordinate 2425611 is likely the translation start site of the *OG1RF_12294* gene since it is preceded by a putative ribosome binding site sequence (AGGAGG). Upstream of such a sequence there is a putative promoter (here named *P12294*) that has a canonical -10 element (**TATAAT**) but lacks a potential -35 element (consensus **TTGACA**) at the optimal length of 17 nucleotides. However, there is a near-consensus -35 element (**T**C**GAC**C) at the suboptimal spacer length of 22 nucleotides. These characteristics suggested that (i) the *P12294* promoter could be recognized by a σ factor similar to the housekeeping σ^{70} factor of *E. coli*, and (ii) a regulatory protein could enhance the activity of the *P12294* promoter. Moreover, sequence analysis of the region located between the TAA stop codon of the *OG1RF_12295* gene (coordinate 2425761) and the *P12294* promoter revealed the existence of an inverted-repeat (IR) that may function as a Rho-independent transcriptional terminator (Figure 14A).

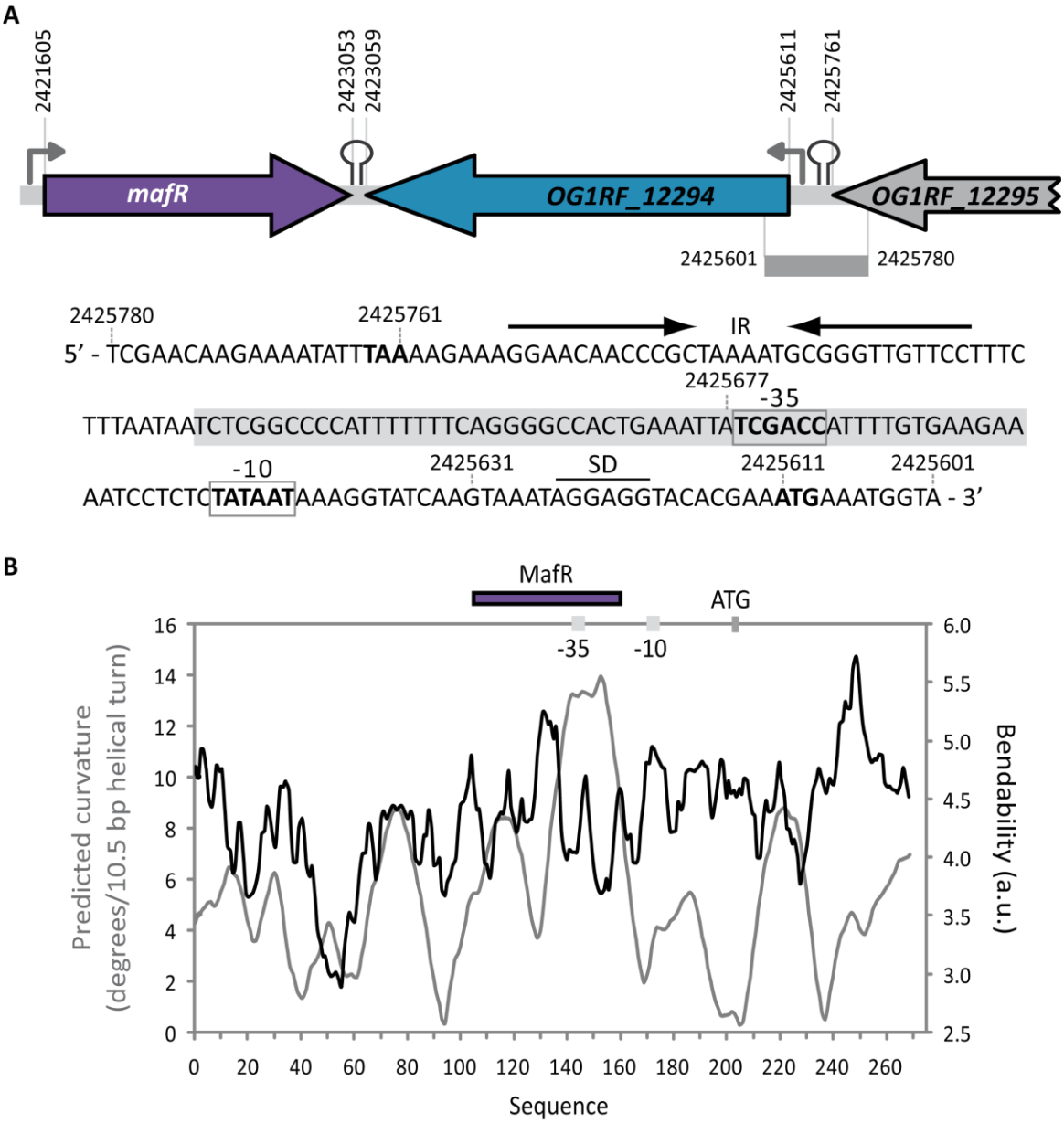


Figure 14. Relevant features of the *P12294* promoter region.

A. The *OG1RF_12294* gene. Coordinates of the translation start and stop codons are indicated. Stem-loop elements represent potential transcriptional terminators. Arrows upstream of the genes represent promoters. The nucleotide sequence of the region spanning coordinates 2425780 to 2425601 is shown. The stop codon (TAA) of the *OG1RF_12295* and the start codon (ATG) of the *OG1RF_12294* are indicated in boldface letters. IR: inverted-repeat. SD: Shine-Dalgarno sequence. The main sequence elements (-35 box and -10 box) of the *P12294* promoter are indicated. The MafR-His binding site defined in this Chapter is shown (grey shadowed box). **B. Bendability/curvature propensity plot.** The region spanning coordinates 2425817 to 2425548 is represented in the plot. The location of the *P12294* core promoter, the start codon of the *OG1RF_12294* gene and the MafR-His binding site are indicated.

To characterize the *P12294* promoter, a 255-bp DNA fragment (coordinates 2425885 to 2425631) (Figure 15) was inserted into the pASTT promoter-probe vector (D. García-Rincón, V. Solano-Collado, A. Bravo, unpublished results), which is based on a *gfp* reporter gene that encodes a variant of the green fluorescent protein (Methods, Section 4.6.4). The recombinant plasmid (pASTT-*P12294*, Tables

3 and 7) was first introduced into OG1RF and OG1RFΔ*mafR*, and the expression of *gfp* (0.32 ± 0.02 and 0.26 ± 0.04 units, respectively) was found to be similar to the basal level (OG1RF harbouring pASTT; 0.38 ± 0.02 units). However, different results were obtained when pASTT-*P12294* was introduced into OG1RFΔ*mafR* harbouring either pDLF or pDLF*mafR* (Figure 15). The expression of *gfp* was ~2.5-fold higher in the presence of plasmid-encoded MafR, indicating that the 255-bp DNA fragment contains a MafR-dependent promoter activity. Furthermore, we constructed a recombinant plasmid (pASTT-*P12294Δ-10*, Tables 3 and 7) in which the -10 element of the *P12294* promoter was removed, losing the MafR-dependent promoter activity. A further deletion analysis allowed us to conclude that the 186-bp region between coordinates 2425816 and 2425631 contains both the *P12294* promoter and the site required for its activation by MafR (plasmids pASTT-*P12294Δ-69* and pASTT-*P12294Δ-208*, Tables 3 and 7) (Figure 15). Thus, we conclude that MafR activates the *P12294* promoter *in vivo*.

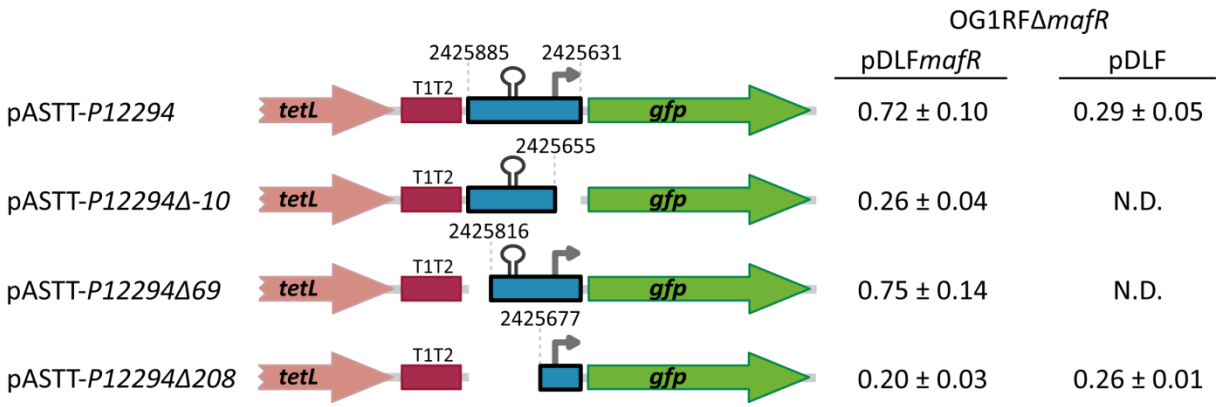


Figure 15. Deletion analysis of the *OG1RF_12294* promoter region. The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The T1T2 box represents the tandem transcriptional terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Four regions from the OG1RF chromosome were inserted independently into the *Sac*I site of the promoter-probe vector pASTT (D. García-Rincón, V. Solano-Collado, A. Bravo, unpublished results). The coordinates of the OG1RF DNA regions are indicated. The stem-loop structure represents the inverted-repeat (IR) identified upstream of the *P12294* promoter (Figure 14A). The arrow represents the canonical -10 element of the *P12294* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} of 0.4). In each case, three independent cultures were analysed. N.D.: non-determined.

MafR BINDS TO THE *P12294* PROMOTER REGION *IN VITRO*

To investigate whether MafR binds to a 270-bp DNA fragment (coordinates 2425817 to 2425548) that contains the *P12294* promoter and the site required for its activation by MafR *in vivo* (Figures 14 and 15), we performed EMSA assays. In these experiments, the 270-bp DNA fragment (10 nM) was incubated with increasing concentrations of $MafR_{OG1RF}$-His (100 to 800 nM) or $MafR_{V583}$-His (50 to 800

nM) (Methods, Section 10.1). As shown in Figure 16, at high protein concentration (400 nM), free DNA was not detected and protein-DNA complexes (C1) moving slowly were observed, which suggested that $MafR_{OG1RF}$-His and $MafR_{V583}$-His could generate multimeric complexes as previously described (see Results, Chapter 1.1).

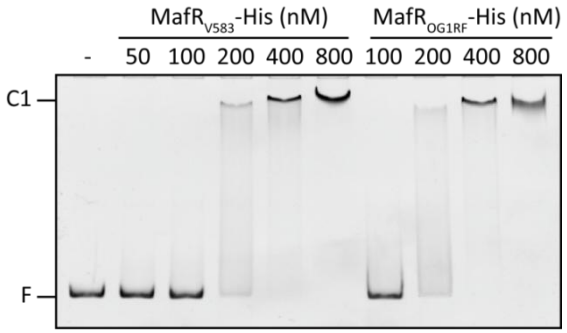


Figure 16. Binding of $MafR_{OG1RF}$-His to the *OG1RF_12294* promoter region. EMSA analysis of the $MafR_{V583}$-His and $MafR_{OG1RF}$-His-DNA complexes. 10 nM of the non-labelled 270-bp DNA fragment (*P12294* promoter, coordinates 2425817 to 2425548) was mixed with increasing concentrations of $MafR_{V583}$-His (50 to 800 nM) and $MafR_{OG1RF}$-His (100 to 800 nM). Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). Bands corresponding to free DNA (F) and protein-DNA complexes (C1) are indicated.

Then, to know whether $MafR_{OG1RF}$-His recognized particular sites on the 270-bp DNA fragment, we performed DNase I footprinting experiments (Methods, Section 10.2). The 270-bp DNA fragment was radioactively labelled either at the 5'-end of the coding strand or at the 5'-end of the non-coding strand (Methods, Section 4.7.1) (Figure 17). On the coding strand and at 100 nM of $MafR_{OG1RF}$-His, protections against DNase I digestion were observed within the region spanning coordinates 2425708 and 2425658. On the non-coding strand and at 125 nM of $MafR_{OG1RF}$-His diminished cleavages were observed between coordinates 2425712 and 2425686. Thus, $MafR_{OG1RF}$-His recognizes a site overlapping the -35 element of the *P12294* promoter (Figure 17). This result allowed us to conclude that MafR activates directly the transcription of the *OG1RF_12294* gene.

The bendability/curvature propensity plot of the 270-bp DNA fragment according to the bend.it program (Table 6) (Vlahoviček *et al.*, 2003) is shown in Figure 14B. The profile contains an intrinsic curvature of high magnitude (~13 degrees per helical turn), which is adjacent to the MafR binding site. Besides, the site recognized by MafR contains a region of potential bendability (~5.2 units).

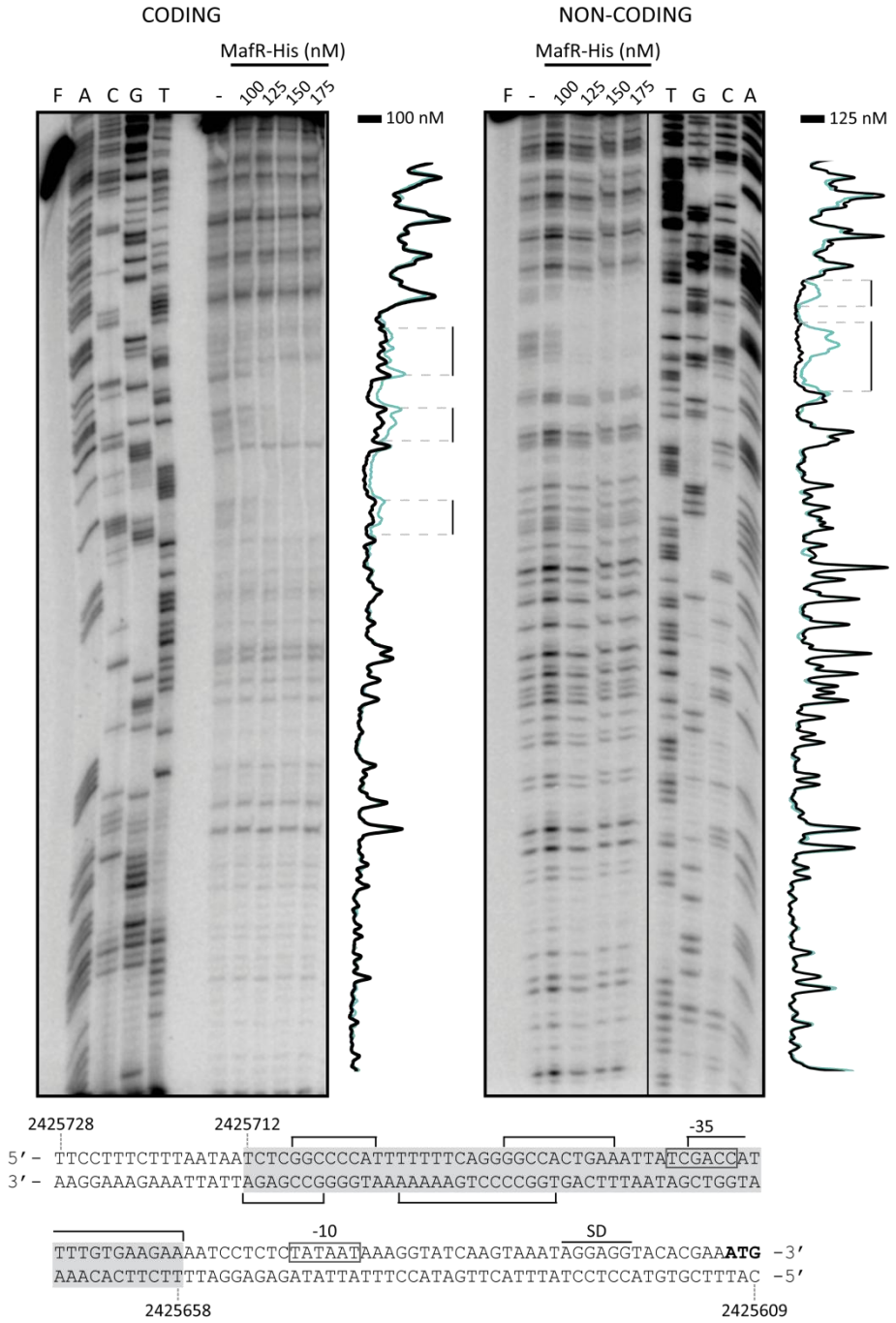


Figure 17. DNase I footprints of complexes formed by Mafr$_{OG1RF}$-His on the *OG1RF_12294* promoter region. Coding and non-coding strands relative to the *P12294* promoter (270-bp DNA fragment) were ^{32}P-labelled at the 5'-end. The DNA (2 nM) was incubated with the indicated concentrations of Mafr$_{OG1RF}$-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 439-bp PCR-amplified DNA fragment (F*12294*-S and R*12294*-S primers) and the ^{32}P-labelled F*12294*-D oligonucleotide (coding) or the ^{32}P-labelled R*12294*-D oligonucleotide (non-coding). All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to free DNA (turquoise line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 2425728 to 2425609 is shown. The -35 and -10 boxes of the *P12294* promoter are indicated. SD: Shine-Dalgarno sequence. Brackets indicate regions protected against DNase I digestion. The site recognized by Mafr$_{OG1RF}$-His (coordinates 2425712 to 2425658) is indicated with a grey box.

2.4. Gene *OG1RF_11486*

As indicated above, another potential MafR target gene identified by genome-wide microarray experiments was *OG1RF_11486* (S. Ruiz-Cruz and A. Bravo, unpublished results). This gene encodes a putative QueT transporter family protein (GenBank AEA94173.1), which could be a membrane-embedded S-component of an energy-coupling factor (ECF) transporter. The ECF transporters are a family of ATP-binding cassette (ABC) transporters that are responsible for the uptake of vitamins and trace elements in prokaryotes. They consist of a membrane-embedded S-component that provides substrate specificity and a three-subunit ECF module that couples ATP hydrolysis to transport. ECF transporters are classified into two groups: (i) group I transporters, genes for the S-component and the ECF module are clustered in the same operon, and (ii) group II transporters, only the genes for the three-subunit ECF module are located in an operon. Moreover, different S-components can share the same ECF module (Majsnerowska *et al.*, 2015; Rodionov *et al.*, 2009). We have found that 759 *E. faecalis* genomes encode a protein identical to OG1RF_11486 (173 residues) (National Center for Biotechnology Information, Genome Assembly and Annotation Report, 13/04/2020) (Altschul *et al.*, 1997). Furthermore, proteins identical to OG1RF_11486 have been found in *Mycobacterium abscenssus* (CPW17925.1), *Listeria monocytogenes* (CWW42654.1; 172 up to 173 residues are identical, V160L) and *S. agalactiae* (KLL29182.1). In these bacteria, the corresponding protein has been annotated as a queuosine precursor ECF transporter S-component QueT. Therefore, protein OG1RF_11486 could be involved in the uptake of a queuosine biosynthetic intermediate. Using the BLASTP program (Altschul *et al.*, 1997) (Table 6), we found that the OG1RF genome encodes an additional QueT transporter family protein (OG1RF_12031; 168 residues; AEA94718.1). It has 55% of sequence similarity to the OG1RF_11486 protein.

TRANSCRIPTION OF THE *OG1RF_11486* GENE

The *OG1RF_11485* (*fabG2*) and *OG1RF_11486* genes appear to be organized into an operon (Figure 18A). The ATG codon at coordinate 1543217 is likely the translation start site of the *OG1RF_11486* gene. This ATG codon is preceded by a putative ribosome binding site sequence (AGGAGAA). The adjacent gene, *fabG2*, encodes a 3-oxoacyl-[acyl-carrier-protein] reductase (254 residues; GenBank WP_002414021.1), which catalyzes the first NADPH-dependent reduction in the elongation cycle of fatty acid biosynthesis.

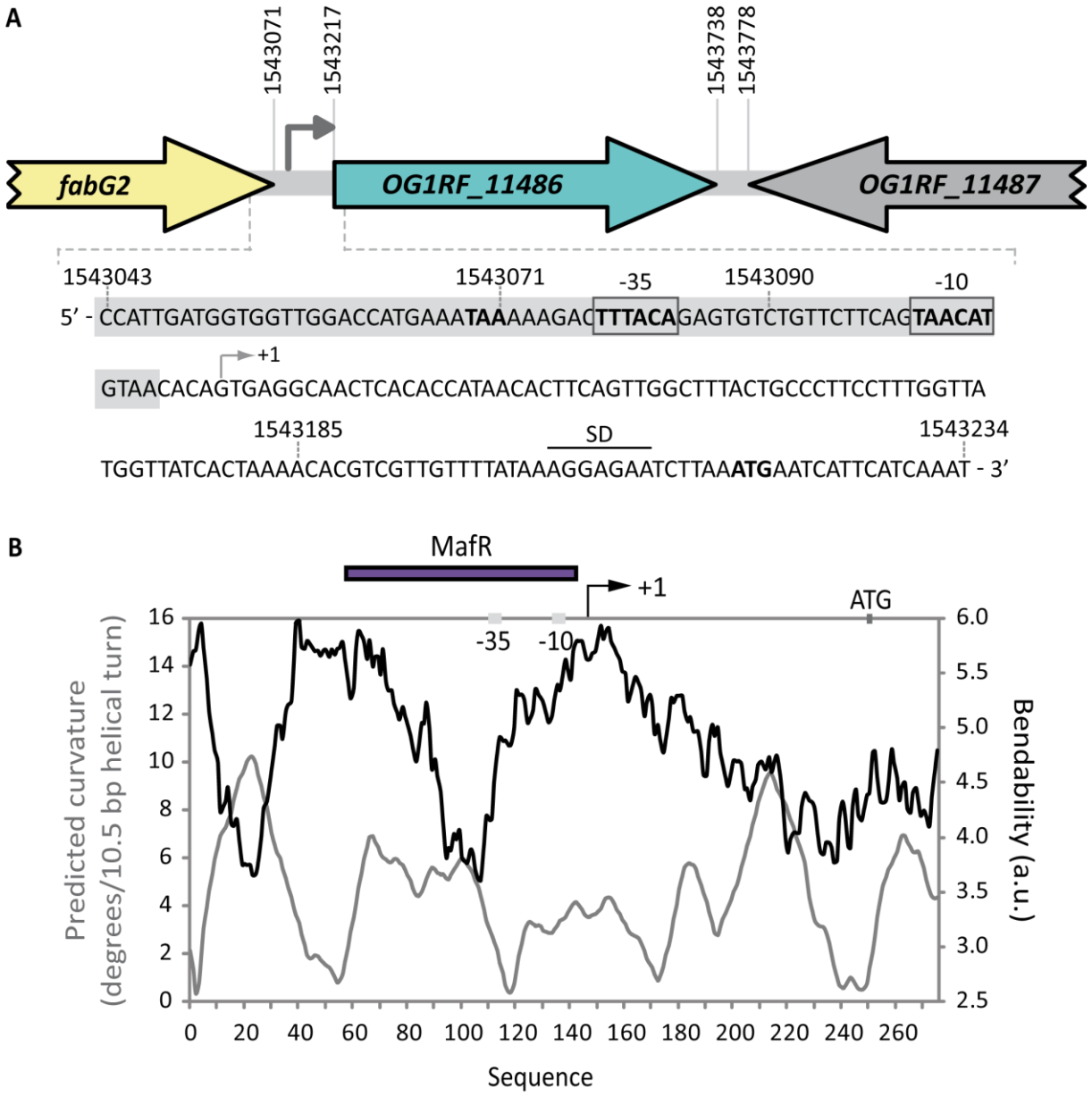


Figure 18. Relevant features of the *P11486* promoter region.

A. Genetic organization of the *OG1RF_11486* gene. Coordinates of the translation start and stop codons are indicated. The nucleotide sequence of the region spanning coordinates 1543043 to 1543234 is shown. The arrow upstream of the *OG1RF_11486* gene represents its promoter. The stop codon (TAA) of the *fabG2* gene and the start codon (ATG) of the *OG1RF_11486* gene are indicated in boldface letters. SD: Shine-Dalgarno sequence. The transcription start site (+1 position) of the *OG1RF_11486* gene and the main sequence elements (-35 box and -10 box) of the *P11486* promoter are indicated. The MafR-His binding site defined in this work is shown (grey shadowed box). **B. Bendability/curvature propensity plot.** The region spanning coordinates 1542969 to 1543243 is represented in the plot. The location of the *P11486* core promoter, the start codon of the *OG1RF_11486* and the MafR-His binding site are indicated.

To investigate whether the *fabG2* and *OG1RF_11486* genes were co-transcribed under our experimental conditions, we performed RT-PCR assays (Methods, Section 7.4). The R*11486*-q primer (primer A in Figure 19), which anneals with the *OG1RF_11486* transcript, was used for extension on

total RNA isolated from OG1RF. The cDNA extension product was then amplified by PCR using either the A and F*11486*-q (primer B in Figure 19) oligonucleotides or the A and F*11486*-S (primer C in Figure 19) oligonucleotides (Table 4). In both cases, PCR fragments with the expected mobility were detected: a 137-bp DNA fragment using the A and B primers, and a 517-bp DNA fragment using the A and C primers. No PCR products were visualized in the negative controls (total RNA as template). Positive controls (chromosomal DNA as template) were run in the same agarose gel electrophoresis. These results indicated that the *OG1RF_11486* gene is transcribed under our experimental conditions. Moreover, they suggested that *fabG2* and *OG1RF_11486* could be transcribed into a polycistronic mRNA molecule from a site located upstream of coordinate 1542901. Nevertheless, these results did not rule out that *OG1RF_11486* could be transcribed from a promoter located between the *fabG2* and *OG1RF_11486* genes (see below.)

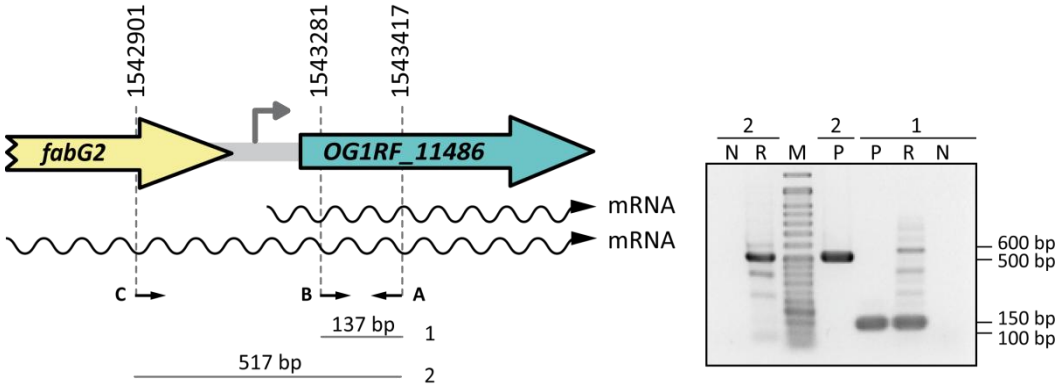


Figure 19. The *fabG2* and *OG1RF_11486* genes may constitute an operon. RT-PCR experiments were performed using total RNA isolated from OG1RF cells. The positions of the oligonucleotides used (A: R*11486*-q; B: R*11486*-q; C: F*11486*-S) are shown. RT-PCRs (lanes R) were subjected to agarose (1.5%) gel electrophoresis, stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). As negative controls (lanes N), RT-PCRs were carried out without adding the reverse transcriptase. The sizes of the PCR-amplified DNA fragments (1 and 2) using genomic DNA as a template (lanes P, positive control) are indicated. The sizes (in bp) of DNA fragments (lane M) used as molecular weight markers (NZYDNA ladder VI, Nzytech) are indicated on the right of the gel.

Next, to analyse whether MafR regulated the expression of the *fabG2* and *OG1RF_11486* genes, we performed qRT-PCR assays (Methods, Section 7.5) using two strains: OG1RFΔ*mafR* harbouring plasmid pDLF (absence of MafR) and OG1RFΔ*mafR* harbouring plasmid pDLF*mafR* (plasmid-encoded MafR). In the presence of plasmid-encoded MafR, transcription of the *OG1RF_11486* gene was increased (log_2FC ~2.4), whereas it did not affect the expression of *fabG2*. Hence, MafR influences positively and specifically the transcription of the *OG1RF_11486* gene. This finding indicated that transcription of *OG1RF_11486*, but not *fabG2*, is under the control of a MafR-dependent promoter. To identify such a promoter, we carried out sequence analysis of the intergenic region using the

BPROM prediction program (*Softberry, Inc.*; Table 6). This analysis predicted a promoter sequence upstream of the *OG1RF_11486* gene (promoter *P11486*) (Figure 18A). The -35 (**TTT****ACA**) and -10 (**TA**AC**AT**) elements of this promoter are separated by 17 nucleotides (optimal length) (Figure 18A). By primer extension (Methods, Section 7.3) using total RNA from OG1RF cells, we demonstrated that the *P11486* promoter is functional *in vivo* (Figure 20). When the oligonucleotide R*11486*-D was used as a primer (Table 4), a cDNA product of 130 nucleotides was detected, indicating that transcription of *OG1RF_11486* starts at coordinate 1543115 (Figure 18A).

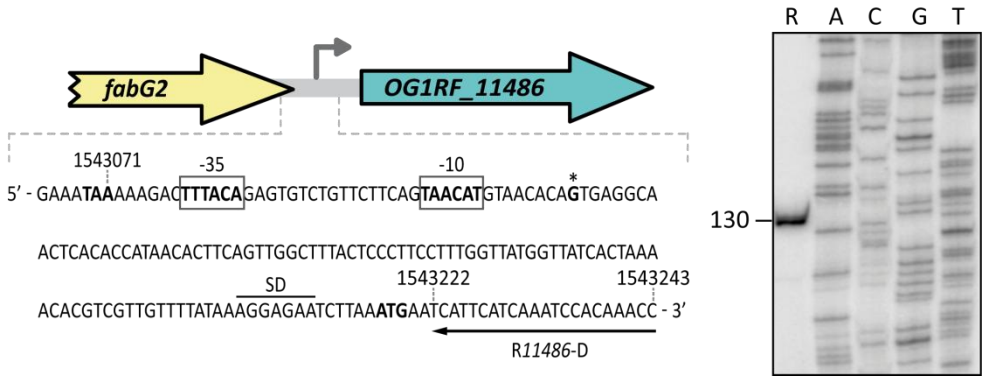


Figure 20. The *OG1RF_11486* gene is transcribed from the *P11486* promoter. Primer extension reactions were carried out using total RNA from OG1RF cells. The asterisk indicates the 3'-end of the cDNA product synthesized using the oligonucleotide R*11486*-D (coordinates 1543222 - 1543243). The size (in nucleotides) of the cDNA product (lane R) is indicated on the left of the gel. Dideoxy-mediated chain termination sequencing reactions were used as DNA size markers (lanes A, C, G, T). Sequencing reactions were prepared using a 423-bp PCR-amplified DNA fragment (F*11486*-S and R*11486*-S primers) and the ^{32}P-labelled R*11486*-D oligonucleotide. The stop codon (TAA) of *fabG2*, the start codon (ATG) of the *OG1RF_11486* and the main sequence elements (-35 box and -10 box) of the *P11486* promoter are indicated in boldface letters. SD: Shine-Dalgarno sequence.

MafR ACTIVATES THE *P11486* PROMOTER *IN VIVO*

To further characterize the *P11486* promoter, we constructed several transcriptional fusions (Figure 21). A 284-bp DNA fragment (coordinates 1542902 to 1543185) was inserted into the pASTT promoter-probe vector (Methods, Section 4.6.4). The recombinant plasmid (pASTT-*P11486*) was first introduced into OG1RF and OG1RFΔ*mafR*. In both strains, *gfp* expression (1.48 ± 0.10 and 1.51 ± 0.16 units, respectively) was ~4-fold higher than the basal level (OG1RF harbouring pASTT). This result indicated that the 284-bp DNA fragment has promoter activity, however, the chromosomal copy of *mafR* is not sufficient to activate such a promoter located on pASTT (multicopy plasmid). Next, we introduced pASTT-*P11486* into OG1RFΔ*mafR* harbouring plasmid pDLF*mafR* (plasmid-encoded MafR). In this strain, *gfp* expression was ~3-fold higher than in the control strain (OG1RFΔ*mafR* harbouring plasmid pDLF) (Figure 21). Similar results were obtained with plasmids pASTT-*P11486Δ66* and pASTT-

P11486Δ145 (Table 7), which allowed us to conclude that the 139-bp region between coordinates 1543047 and 1543185 contains both the *P11486* promoter and the site required for its activation by MafR. A further deletion analysis showed that sequences between coordinates 1543047 and 1543071 (plasmid pASTT-*P11486Δ169*) are needed for MafR-mediated activation of the *P11486* promoter but not for promoter activity. Furthermore, deletion of the region that spans coordinates 1543071 and 1543090 (pASTT-*P11486Δ188*) removes the -35 element of the *P11486* promoter and, consequently, reduces the expression of *gfp* to basal levels (Figure 21).

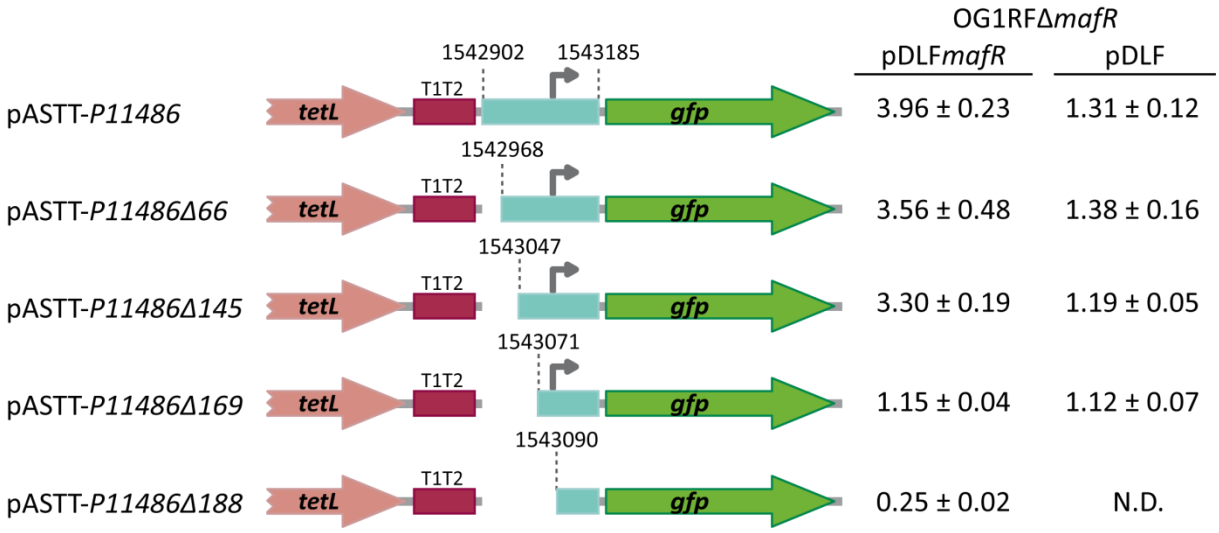


Figure 21. Deletion analysis of the *OG1RF_11486* promoter region. The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The T1T2 box represents the tandem transcriptional terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Five regions from the OG1RF chromosome were inserted independently into the *Sac*I site of the promoter-probe vector pASTT (D. García-Rincón, V. Solano-Collado and A. Bravo, unpublished results). The coordinates of such regions are indicated. The arrow represents the -35 element of the *P11486* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} of 0.4). In each case, three independent cultures were analysed. N.D.: non-determined.

MafR BINDS TO THE *P11486* PROMOTER REGION *IN VITRO*

To investigate whether MafR binds to a 275-bp DNA fragment (coordinates 1542969 to 1543243) that contains both the *P11486* promoter and the site required for its activation by MafR *in vivo* (Figure 21), we performed EMSA assays. In these experiments, the 275-bp DNA fragment was radioactively labelled at the 5'-end of the non-coding strand. The labelled DNA (4 nM) was incubated with increasing concentrations of $MafR_{OG1RF}$-His (200 to 1000 nM) (Figure 22). Free and bound DNAs were separated by electrophoresis on a native polyacrylamide (6%) gel. Labelled DNA was visualized using a Fujifilm Image Analyzer (FLA-3000). As shown in Figure 22, DNA-protein complexes (C1 to C4)

were observed at the lower concentration of $MafR_{OG1RF}$-His (200 nM). However, as the protein concentration was increased, higher-order complexes appeared and faster-moving complexes disappeared. Thus, in agreement with our previous results (Results, Chapter 1.1), $MafR_{OG1RF}$-His generated multimeric complexes on the 275-bp DNA. The interaction of $MafR_{OG1RF}$-His with the 275-bp DNA fragment was further analysed by DNase I footprinting experiments (Methods, Section 10.2). The 275-bp DNA fragment was radioactively labelled either at the 5'-end of the coding strand or at the 5'-end of the non-coding strand (Methods, Section 4.7.1) (Figure 23). On the coding strand and at 350 nM of $MafR_{OG1RF}$-His, changes in DNase I sensitivity (diminished cleavages) were observed within the region spanning coordinates 1543047 and 1543110. On the non-coding strand and at 300 nM of $MafR_{OG1RF}$-His, diminished cleavages were observed between coordinates 1543043 and 1543110. On both strands and at 400 nM of $MafR_{OG1RF}$-His, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the ability of $MafR_{OG1RF}$-His to generate multimeric complexes. Similar results were obtained when the binding reactions contained heparin as a competitor (not shown). Thus, $MafR_{OG1RF}$-His recognizes preferentially a DNA site overlapping the *P11486* core promoter. Such a DNA site includes sequences needed for MafR-mediated activation of the *P11486* promoter *in vivo* (Figure 21). This result allowed us to conclude that MafR activates directly the transcription of the *OG1RF_11486* gene. According to the bendability/curvature propensity plot (Vlahoviček *et al.*, 2003) of the 275-bp DNA fragment, the MafR binding site contains regions of potential bendability (Figure 18B).

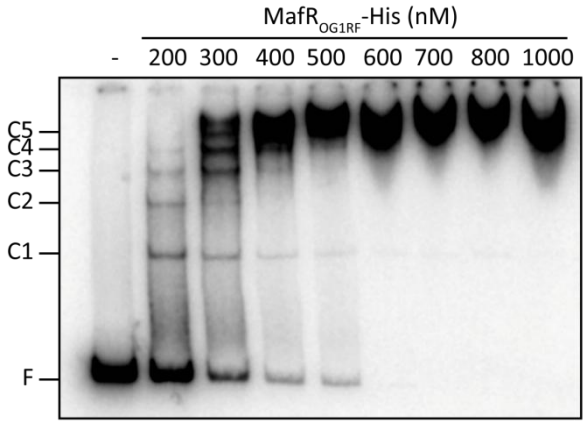


Figure 22. Binding of $MafR_{OG1RF}$-His to the *OG1RF_11486* promoter region. EMSA was carried out incubating $MafR_{OG1RF}$-His (200 to 1000 nM) and 4 nM of the ^{32}P-labelled 275-bp DNA fragment (*P11486* promoter, coordinates 1542969 to 1543243). Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis and labelled-DNA bands were visualized using a Fujifilm Image Analyzer (FLA-3000). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C5) are indicated.

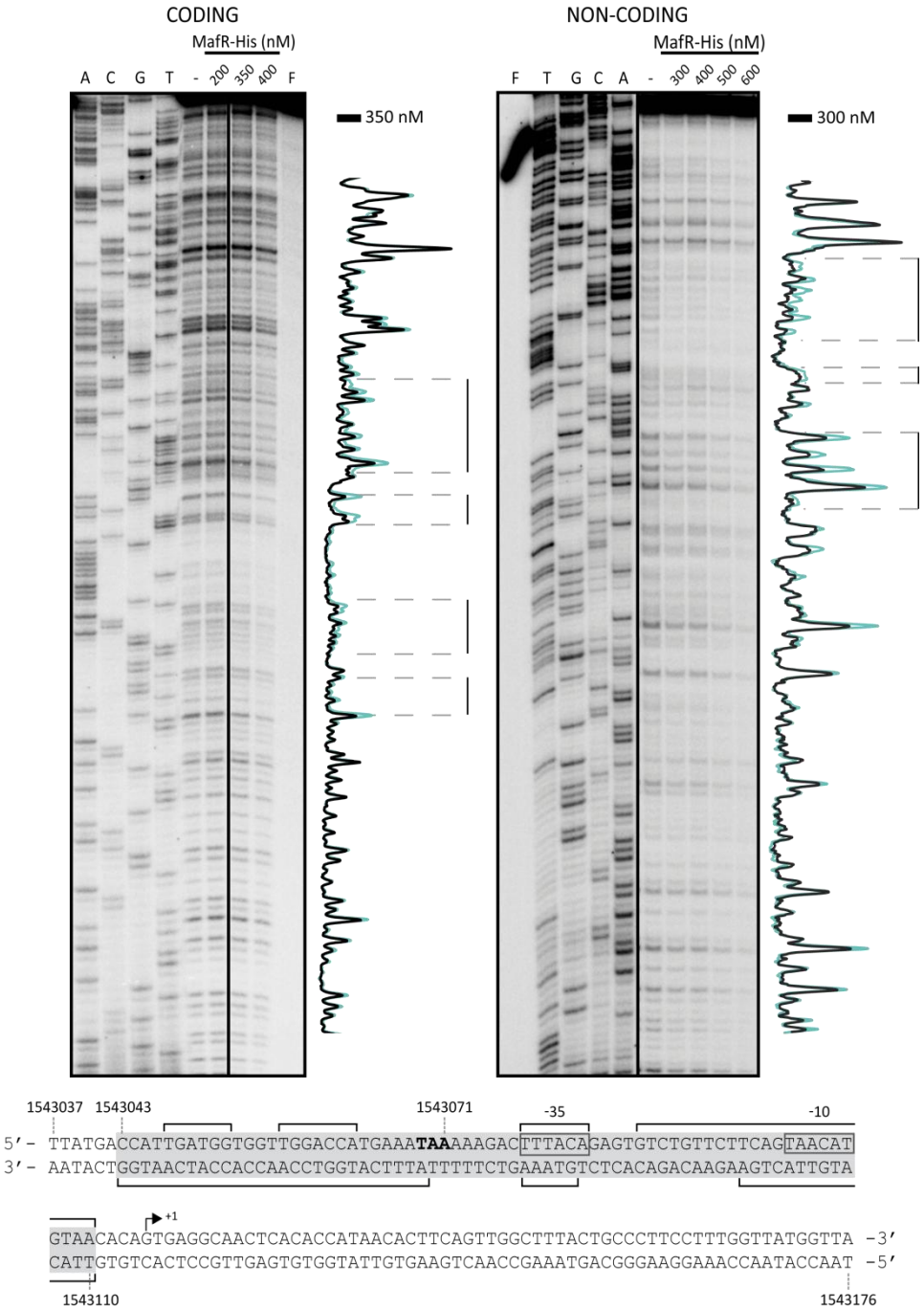


Figure 23. DNase I footprints of complexes formed by MafR$_{OG1RF}$-His on the *OG1RF_11486* promoter region. Coding and non-coding strands relative to the *P11486* promoter (275-bp DNA fragment) were ^{32}P-labelled at the 5'-end. The labelled DNA (4 nM) was incubated with the indicated concentrations of MafR$_{OG1RF}$-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 423-bp PCR-amplified DNA fragment (F*11486*-S and R*11486*-S primers) and the ^{32}P-labelled F*11486*-D oligonucleotide (coding) or the ^{32}P-labelled R*11486*-D oligonucleotide (non-coding). All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to free DNA (turquoise line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 1543037 to 1543176 is shown. The transcription initiation site (+1 position) of the *OG1RF_11486* gene and the -35 and -10 boxes of the *P11486* promoter are indicated. Brackets indicate regions protected against DNase I digestion. The site recognized by MafR$_{OG1RF}$-His (coordinates 1543043 to 1543110) is indicated with a grey box.

2.5. Gene *OG1RF_10478*

Genome-wide microarray assays designed for strains OG1RF and OG1RFΔ*mafR* (Ruiz-Cruz *et al.*, 2016) revealed that the *OG1RF_10478* gene was up-regulated in the presence of MafR (log_2FC ~2). Here we validate these results by qRT-PCR assays and, in addition, we demonstrate that MafR activates directly the transcription of the *OG1RF_10478* gene. This gene encodes a hypothetical protein (GenBank AEA93165.1) of unknown function and is flanked by the *OG1RF_10477* and *OG1RF_10479* genes, which encode a transposase and a sodium/dicarboxylate symporter family protein, respectively (Figure 24). We have found that 244 *E. faecalis* genomes encode a protein identical to OG1RF_10478 (National Center for Biotechnology Information, Genome Assembly and Annotation Report, 16/04/2020) (Altschul *et al.*, 1997). Moreover, proteins identical or almost identical to OG1RF_10478 have been found in *M. abscenssus* (CPW46010.1), *L. monocytogenes* (EAC5386049.1; 110 up to 111 residues are identical, G25A), *S. agalactiae* (KLL21071.1; 109 up to 111 residues are identical, T16M and F18Y), *S. pneumoniae* (CWJ89864.1; 109 up to 111 residues are identical, T16M and F18Y) and *Sphingobacterium faecium* (SJN36152.1; 109 up to 111 residues are identical, T16M and F18Y). In all of these bacteria, the function of the equivalent protein remains also unknown. However, using the HHpred program for protein homology detection and structure prediction (Söding *et al.*, 2005), we have found that the OG1RF_10478 protein has homology (~96% probability) to the PTS EIIB cellobiose-specific component of the *E. coli* K12 strain (van Montfort *et al.*, 1997). PTS uses phosphoenolpyruvate (PEP) rather than ATP as an energy source to drive translocation, and it chemically modifies its substrate by phosphorylation. PTS is composed of two general proteins, enzyme I (EI) and histidine-containing phosphocarrier protein (Hpr), and the carbohydrate-specific enzyme II complex (EII). The EII of a PTS usually consists of three functional domains (two cytoplasmic domains: IIA and IIB; a transmembrane channel: IIC). The three functional domains can be encoded by a single gene, IIABC (for instance, the mannitol-specific EII from *E. coli*), or by separated ones, as the cellobiose-specific EII from *E. coli*. Therefore, protein OG1RF_10478 could be involved in the transport of cellobiose.

MafR INFLUENCES POSITIVELY THE TRANSCRIPTION OF THE *OG1RF_10478* GENE

To analyse whether MafR regulated the expression of the *OG1RF_10478* gene, we performed qRT-PCR assays (Methods, Section 7.5) using two strains, OG1RF (wild-type) and OG1RFΔ*mafR* (deletion mutant). The results indicated that expression of the OG1RF_*10478* gene was up-regulated in the presence of MafR. The fold change in gene expression (log_2FC) was ~3. Further experiments confirmed this result. Specifically, we determined the relative expression of OG1RF_*10478* in two strains: OG1RFΔ*mafR* harbouring plasmid pDLF (absence of MafR) and OG1RFΔ*mafR* harbouring

plasmid pDLF*mafR* (plasmid-encoded MafR). In the presence of plasmid-encoded MafR, the transcription of *OG1RF_10478* was increased (log_2FC ~3.6). Hence, MafR influences positively the transcription of the *OG1RF_10478* gene.

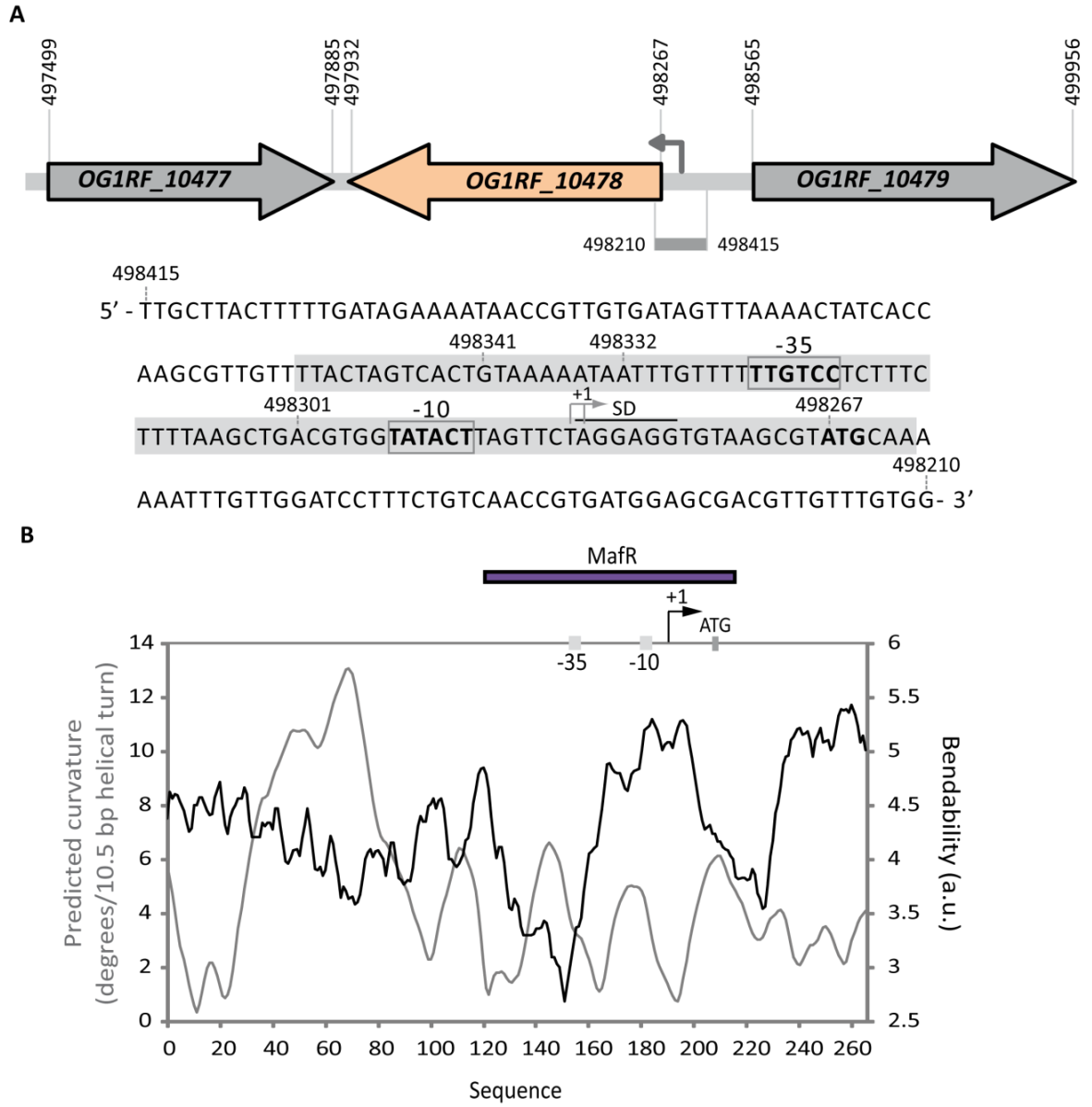


Figure 24. Relevant features of the *P10478* promoter region. A. Genetic organization of the chromosome region that contains the *OG1RF_10478* gene. Coordinates of the translation start and stop codons are indicated. The nucleotide sequence of the region spanning coordinates 498415 to 498210 is shown. The arrow upstream of the *OG1RF_10478* gene represents its promoter. The start codon (ATG) of the *OG1RF_10478* and the main sequence elements (-35 box and -10 box) of the *P10478* promoter are indicated in boldface letters. SD: Shine-Dalgarno sequence. The transcription start site (+1 position) of the *OG1RF_10478* gene is indicated. The MafR-His binding site defined in this work is shown (grey shadowed box). **B. Bendability/curvature propensity plot.** The region spanning coordinates 498210 to 498475 is represented in the plot. The location of the *P10478* core promoter, the translation start codon of *OG1RF_10478* and the MafR-His binding site are indicated.

MafR ACTIVATES THE *P10478* PROMOTER *IN VIVO*

To identify the MafR-dependent promoter of the *OG1RF_10478* gene, we carried out a sequence analysis of the region spanning coordinates 498210 to 498415, which includes part of the intergenic region between *OG1RF_10478* and *OG1RF_10479* (see Figure 24A). The ATG codon at coordinate 498267 is likely the translation start site of the *OG1RF_10478* gene. This ATG codon is preceded by a putative ribosome binding site sequence (AGGAGG). Using the BPROM prediction program (*Softberry, Inc.*; Table 6), we found a potential promoter sequence upstream of the *OG1RF_10478* gene (promoter *P10478*) (Figure 24A). This promoter has a near-consensus -10 element (**TATAC**T) but lacks a potential -35 element (consensus **TTGACA**) at the optimal length of 17 nucleotides. However, there is a possible -35 element (**TTG**T**C**C) at the suboptimal spacer length of 22 nucleotides (Figure 24A). By primer extension (Methods, Section 7.3) using total RNA from OG1RF cells and from OG1RFΔ*mafR* cells harbouring plasmid pDLF*mafR* (plasmid-encoded MafR), we demonstrated that the *P10478* promoter is functional *in vivo* (Figure 25). When the oligonucleotide R*10478*-D was used as a primer (Table 4), cDNA products of 73-74 nucleotides were detected, indicating that transcription of *OG1RF_10478* starts at coordinate 498283 (Figure 24A).

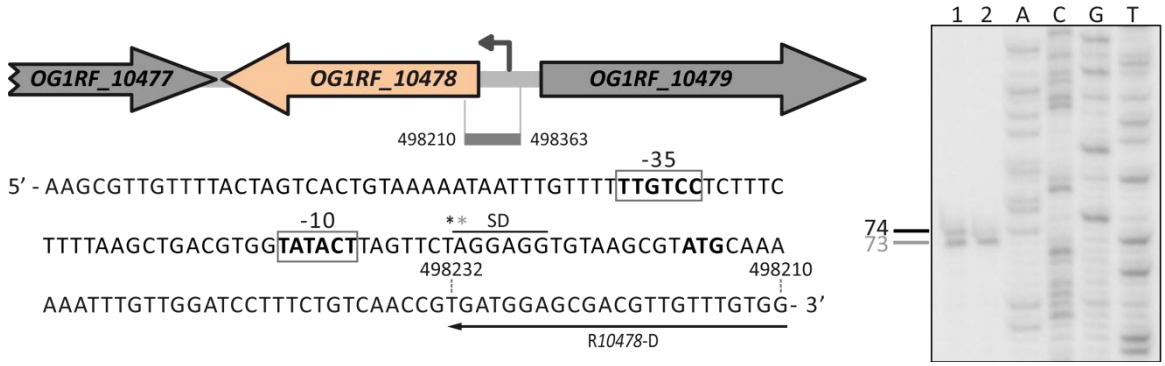


Figure 25. The *OG1RF_10478* gene is transcribed from the *P10478* promoter. Primer extension reactions were carried out using total RNA from OG1RF cells (lane 2) and from OG1RFΔ*mafR* cells harbouring plasmid pDLF*mafR* (lane 1). The asterisks indicate the 3'-end of the cDNA products synthesized using the oligonucleotide R*10478*-D (coordinates 498210 - 498232). The size (in nucleotides) of the cDNA products (lanes 1 and 2) is indicated on the left of the gel. Dideoxy-mediated chain termination sequencing reactions were used as DNA size markers (lanes A, C, G, T). Sequencing reactions were prepared using a 422-bp PCR-amplified DNA fragment (F*10478*-S and R*10478*-S primers) and the ^{32}P-labelled R*10478*-D oligonucleotide. The start codon (ATG) of the *OG1RF_10478* gene and the main sequence elements (-35 box and -10 box) of the *P10478* promoter are indicated in boldface letters. SD: Shine-Dalgarno sequence.

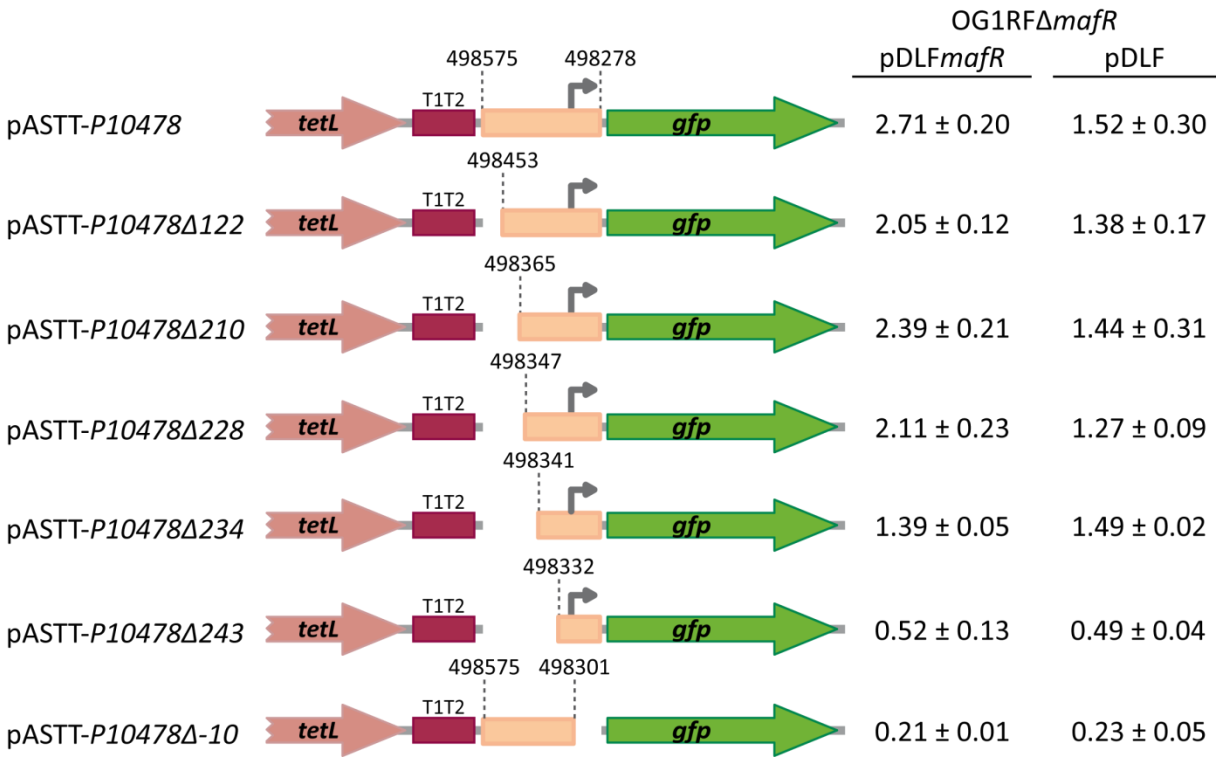


Figure 26. Deletion analysis of the *OG1RF_10478* promoter region. The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The T1T2 box represents the tandem transcriptional terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Seven regions from the OG1RF chromosome were inserted independently into the *Sac*I site of the promoter-probe vector pASTT. The coordinates of such regions are indicated. The arrow represents the -10 element of the *P10478* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} of 0.4). In each case, three independent cultures were analysed.

To further characterize the *P10478* promoter, we constructed several transcriptional fusions (Figure 26). A 298-bp DNA fragment (coordinates 498575 to 498278) was inserted into the pASTT promoter-probe vector (Methods, Section 4.6.4). The recombinant plasmid (pASTT-*P10478*) was first introduced into OG1RF and OG1RFΔ*mafR*. In both strains, *gfp* expression (1.54 ± 0.23 and 1.60 ± 0.16 units, respectively) was ~4-fold higher than the basal level (OG1RF harbouring pASTT, 0.38 ± 0.02). This result indicated that the 298-bp DNA fragment has promoter activity, however, the chromosomal copy of *mafR* is not sufficient to activate such a promoter located on pASTT (multicopy plasmid). Next, we introduced pASTT-*P10478* into OG1RFΔ*mafR* harbouring plasmid pDLF*mafR* (plasmid-encoded MafR). In this strain, *gfp* expression was ~1.8-fold higher than in the control strain (OG1RFΔ*mafR* harbouring plasmid pDLF) (Figure 26). Similar results were obtained with plasmids pASTT-*P10478Δ122,* pASTT-*P10478Δ210* and pASTT-*P10478Δ228* (Table 7) (Figure 26), which allowed us to conclude that the 70-bp region between coordinates 498347 and 498278 contains both the *P10478* promoter and the site required for its activation by MafR. A further deletion analysis showed that sequences between coordinates 498347 and 498341 (plasmid pASTT-*P10478Δ234*) are needed

for MafR-mediated activation of the *P10478* promoter but not for promoter activity. Furthermore, deletion of the region that spans coordinates 498341 and 498332 (pASTT-*P10478Δ243*) reduces the activity of the *P10478* promoter and, consequently, reduces to some extent the expression of *gfp*. Finally, deletion of the region that spans coordinates 498301 and 498278 (pASTT-*P10478Δ-10*) removes the -10 element of the *P10478* promoter and, consequently, reduces the expression of *gfp* to basal levels (Figure 26). Concluding, the 70-bp region between coordinates 498347 and 498278 contains both the *P10478* promoter and the site required for its activation by MafR.

MafR BINDS TO THE *P10478* PROMOTER REGION *IN VITRO*

In a first approach, we analysed the interaction of $MafR_{OG1RF}$-His with the *P10478* promoter region by EMSA assays using a 266-bp DNA fragment (coordinates 498475 to 498210) that contains both the *P10478* promoter and the site required for its activation by MafR *in vivo* (Figure 26). The 266-bp DNA fragment was radioactively labelled at the 5'-end of the non-coding strand. The labelled DNA (4 nM) was incubated with increasing concentrations of $MafR_{OG1RF}$-His (100 to 800 nM) (Figure 27A). The binding conditions were those described in Methods, Section 10.1. Free and bound DNAs were separated by electrophoresis on a native polyacrylamide (6%) gel. As shown in Figure 27A, at 200 nM of $MafR_{OG1RF}$-His, three DNA-protein complexes were observed (C1, C2 and C3), as well as free DNA. As the protein concentration was increased, higher-order complexes appeared and faster-moving complexes disappeared. This pattern of complexes is compatible with the formation of multimeric protein-DNA complexes (see Results, Chapter 1.1). Moreover, by EMSA assays, we also analysed the effect of heparin on the binding of $MafR_{OG1RF}$-His to the 266-bp DNA (Figure 27B). The labelled 266-bp DNA fragment (4 nM) was incubated with 600 nM of $MafR_{OG1RF}$-His in the presence of different concentrations of heparin (0.25-4.5 µg/ml). As shown in Figure 27B, $MafR_{OG1RF}$-His-DNA interactions were disrupted at heparin concentrations above 0.75 µg/ml.

To determine whether MafR recognized specific sites on the *P10478* promoter region, we performed DNase I footprinting experiments (Methods, Section 10.2) using $MafR_{OG1RF}$-His and the aforementioned 266-bp DNA fragment (coordinates 498475 to 498210). This fragment was radioactively labelled either at the 5'-end of the coding strand or at the 5'-end of the non-coding strand (Methods, Section 4.7.1) (Figure 28). On the coding strand and at 300 nM of $MafR_{OG1RF}$-His, changes in DNase I sensitivity (diminished cleavages) were observed from position 498352 to 498340, and from 498279 to 498262. On the non-coding strand and at 300 nM of $MafR_{OG1RF}$-His, diminished cleavages were observed from 498353 to 498317, and from 498285 to 498264. Therefore, the interaction of $MafR_{OG1RF}$-His with the 266-bp DNA fragment protected two regions against DNase I digestion: (i) a region that includes both the -35 element and sequences (498347-

498341) needed for MafR-mediated activation of the *P10478* promoter (Figure 26), and (ii) a region just downstream from the -10 element. On both strands and at 600 nM of $MafR_{OG1RF}$-His, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the ability of $MafR_{OG1RF}$-His to generate multimeric complexes. From these results, we conclude that $MafR_{OG1RF}$-His enhances the efficiency of the *P10478* promoter by binding to a site that overlaps the core promoter. According to the bendability/curvature propensity plot (Vlahoviček *et al.*, 2003) of the 266-bp DNA fragment, the -10 element of the *P10478* promoter is flanked by regions of potential bendability (Figure 24B).

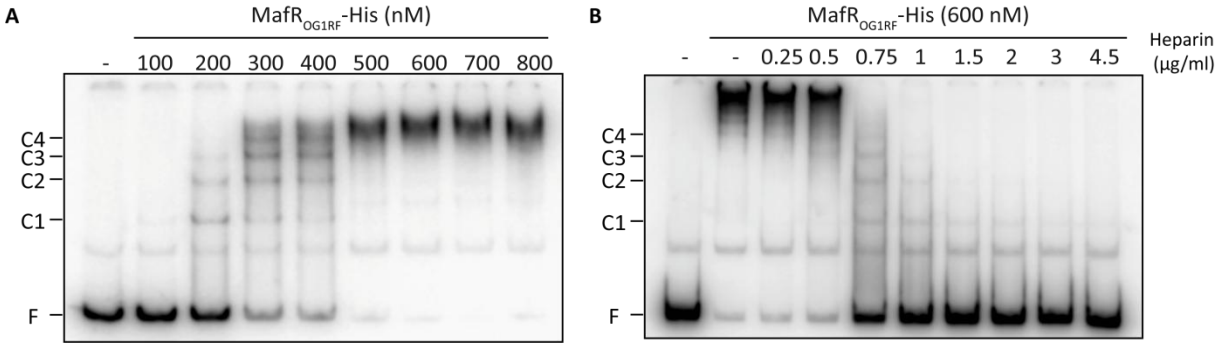


Figure 27. Binding of $MafR_{OG1RF}$-His to the *OG1RF_10478* promoter region. The 266-bp DNA fragment (*P10478* promoter, coordinates 498475 to 498210) was radiolabelled at the 5'-end of the non-coding strand. Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis and labelled DNA was visualized using a Fujifilm Image Analyzer (FLA-3000). **A. Binding of $MafR_{OG1RF}$-His to the *OG1RF_10478* promoter region in the absence of heparin.** The labelled DNA fragment (4 nM) was incubated with increasing concentrations of $MafR_{OG1RF}$-His (100 to 800 nM). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated. **B. Binding of $MafR_{OG1RF}$-His to the *OG1RF_10478* promoter region in the presence of heparin.** The labelled DNA fragment (4 nM) was mixed with 600 nM of $MafR_{OG1RF}$-His (formation of higher-order complexes, Figure 27A) in the presence of different concentrations of heparin (0.25 to 4.5 µg/ml). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated.

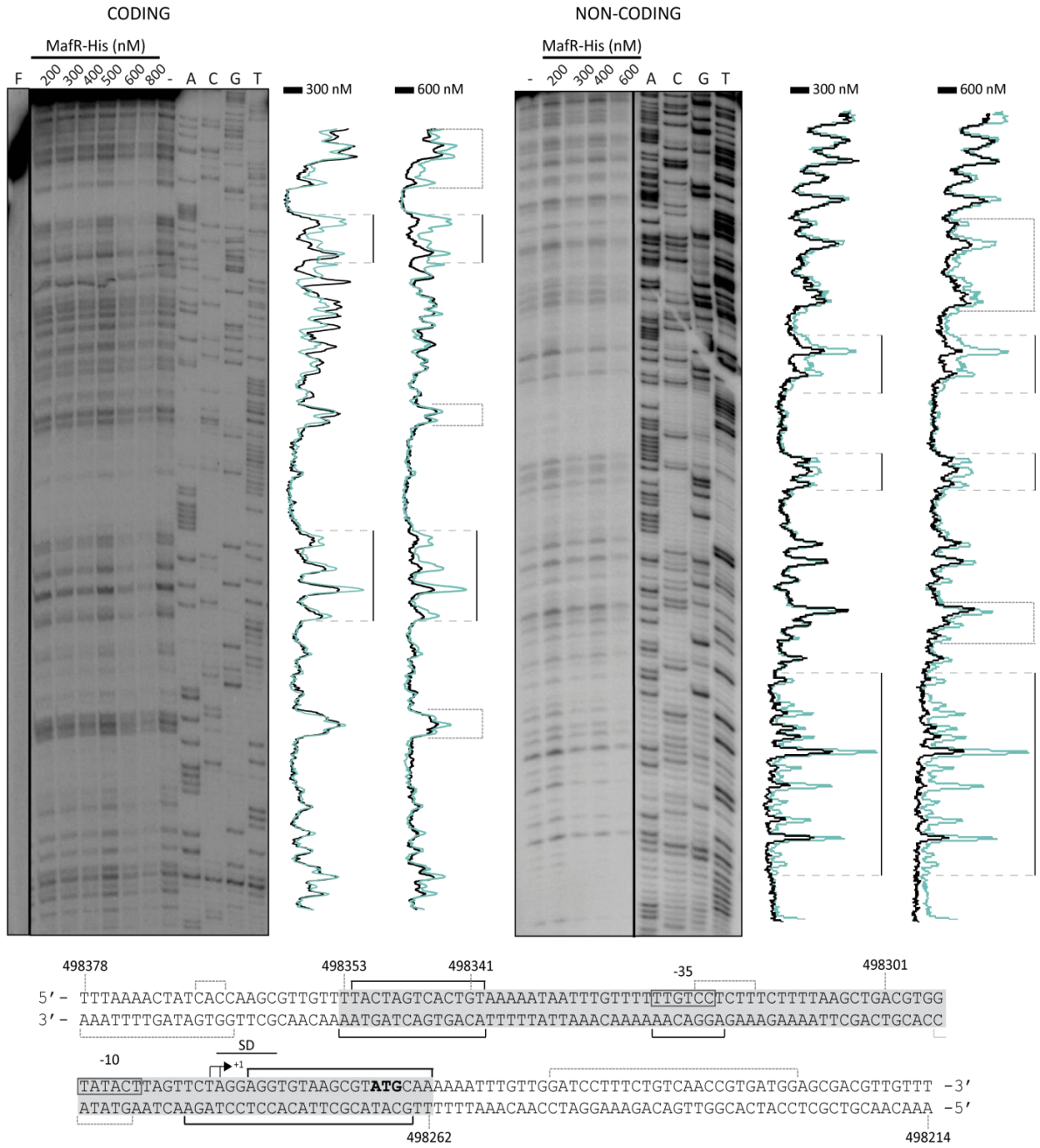


Figure 28. DNase I footprints of complexes formed by MafR$_{OG1RF}$-His on the *OG1RF_10478* promoter region. Coding and non-coding strands relative to the *P10478* promoter (266-bp DNA fragment) were ^{32}P-labelled at the 5'-end. The labelled DNA (4 nM) was incubated with the indicated concentrations of MafR$_{OG1RF}$-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 422-bp PCR-amplified DNA fragment (F*10478*-S and R*10478*-S primers) and the ^{32}P-labelled F*10478*-D oligonucleotide (coding) or the ^{32}P-labelled R*10478*-D oligonucleotide (non-coding). All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to free DNA (turquoise line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 498378 to 498214 is shown. The transcription initiation site (+1 position) of the *OG1RF_10478* gene and the -35 and -10 boxes of the *P10478* promoter are indicated. SD: Shine-Dalgarno sequence. Black brackets indicate the main regions protected against DNase I digestion. The site recognized by MafR$_{OG1RF}$-His (coordinates 498353 to 498262) is indicated with a grey box.

CHAPTER 3

FUNCTIONAL CHARACTERIZATION OF THE MgaP PROTEIN ENCODED BY THE *S. pneumoniae* R6 GENOME

The strain-specific MgaP protein of *S. pneumoniae* shares sequence similarity with the regulators of the Mga/AtxA family, particularly with the pneumococcal Mga*Spn* protein. In this Chapter, we describe the identification of the promoter of the *mgaP* gene (*PmgaP*) using primer extension and promoter-reporter fusions. We show that the *mgaP* gene is expressed under laboratory conditions and that its promoter does not seem to be autoregulated. Moreover, by qRT-PCR, primer extension, transcriptional fusions and DNase I footprinting assays, we demonstrate that MgaP activates directly the transcription of its adjacent gene, *pclA* (collagen-like protein A). Furthermore, by gel retardation experiments, we show that MgaP generates multimeric complexes on linear DNA fragments and, by DNase I footprinting assays, we demonstrate that MgaP and Mga*Spn* have different DNA-binding specificities. This study was co-supervised by S. Ruiz-Cruz and A. Bravo.

3.1. Expression of *mgaP* under laboratory conditions

In *S. pneumoniae*, there is a high degree of genetic diversity among strains. In general, homologous recombination and horizontal gene transfer mechanisms determine a substantial reorganization of the chromosome of individual strains. The pneumococcal TIGR4 strain (serotype 4) was isolated in Norway in the 1990s and its genome sequence was published in 2001 (Tettelin et al., 2001). The pneumococcal R6 strain derives from the D39 strain (serotype 2), which was isolated in the United States early in the 20th century. The complete genome sequence of the R6 strain was also published in 2001 (Hoskins *et al.*, 2001). A comparison between both genomes revealed that the R6 genome has six gene clusters that are absent from the TIGR4 genome (Brückner *et al.*, 2004). One of the R6-specific clusters (9,634 bp) contains two divergent genes, *spr1403* (also known as *pclA*; collagen-like protein) and *spr1404* (named *mgaP* in this Thesis) (Paterson *et al.*, 2008) (see Figure 29). Paterson *et al.* reported that this gene cluster is present in particular clinical isolates (Paterson *et al.*, 2008). Searching for homologies, we found that the MgaP protein has homology (60.3% similarity) to the Mga*Spn* transcriptional activator, which is present in both pneumococcal strains, TIGR4 and R6 (Hemsley *et al.*, 2003; Solano-Collado *et al.*, 2012). The following sections are focused on the functional characterization of the MgaP protein.

To know whether the *mgaP* gene of the pneumococcal R6 strain was transcribed under laboratory conditions, we determined its relative expression by qRT-PCR assays (Methods, Section 7.5) and using the comparative C_T method (Schmittgen and Livak, 2008). To this aim, total RNA was isolated from R6 cells grown under standard laboratory conditions (AGCH medium supplemented with 0.2% yeast extract and 0.3% sucrose, 37°C and without aeration) to both logarithmic and stationary phases (Methods, Section 1). Figure 30 shows the corresponding bacterial growth curve. Transcription of

mgaP was found to be higher at logarithmic phase (sample taken at an OD_{650} of 0.2). Compared to the stationary phase (sample taken at an OD_{650} of 0.99), the fold change (log_2FC) in *mgaP* expression was ~1.7. Therefore, since the expression of the *mgaP* gene was higher at exponential phase, all the experiments shown in the next sections were performed at such a phase.

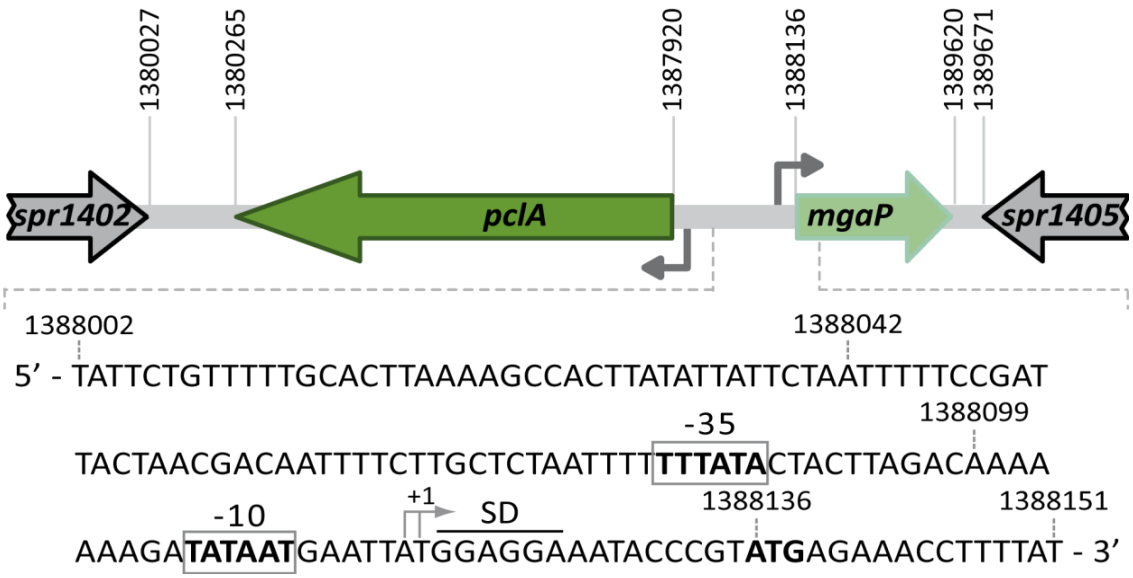


Figure 29. Genetic organization of the region that contains the *mgaP* gene in the *S. pneumoniae* R6 genome. The coordinates of the translation start and stop codons of the *pclA* and *mgaP* genes are indicated. The nucleotide sequence of the region spanning coordinates 1388002 to 1388151 is shown. The possible translation start codon (ATG; coordinate 1388136) of *mgaP* is indicated in boldface letters. The transcription start site (+1 position; identified in this work) of the *mgaP* gene and the main sequence elements (-35 box and -10 box) of the *PmgaP* promoter are indicated. SD: Shine-Dalgarno sequence.

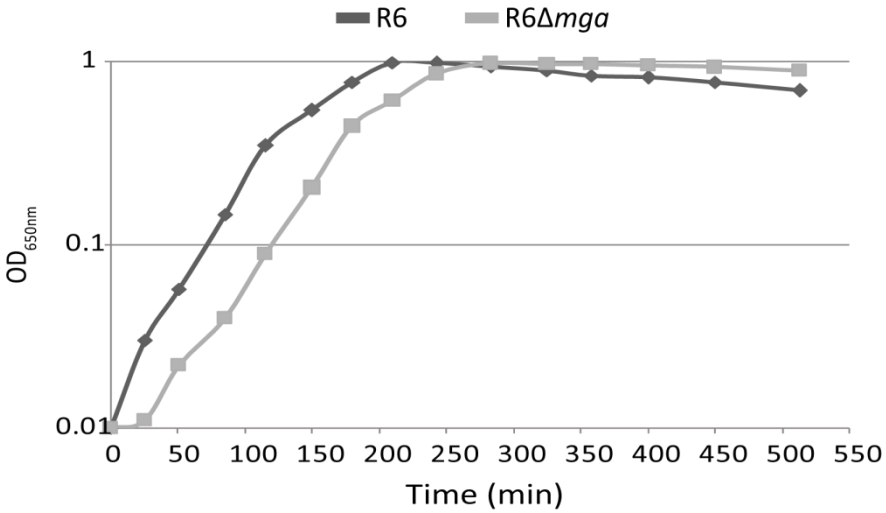


Figure 30. *S. pneumoniae* growth curves. Bacteria were grown in AGCH medium supplemented with 0.2% yeast extract and 0.3% sucrose, at 37°C and without aeration. Optical density (OD) was measured at 650 nm using a Bausch & Lomb (Spectronic 20D+) spectrophotometer, at intervals of 15-30 min. The dark grey line corresponds to strain R6 (wild-type) and the light grey line corresponds to strain R6Δ*mga*, which lacks the regulatory *mgaSpn* gene (Solano-Collado *et al.*, 2012).

We have not yet constructed an R6 mutant strain lacking the *mgaP* gene. Nonetheless, to investigate whether MgaP influenced the activity of particular promoters *in vivo* (See Results, Chapter 3.4), we constructed a pneumococcal strain designed to produce higher levels of the MgaP protein. Specifically, we inserted the promoterless *mgaP* gene into the pDLF constitutive expression vector (Ruiz-Cruz *et al.*, 2016) in both orientations, generating the recombinant plasmids pDLF*mgaP* (constitutive expression of *mgaP* located on the plasmid) and pDLF*mgaPi* (no expression of *mgaP* located on the plasmid). Subsequently, we introduced each recombinant plasmid into the R6 strain. By qRT-PCR assays, we determined the relative expression of the *mgaP* gene in both strains: R6 harbouring plasmid pDLF*mgaP* (chromosome-encoded and plasmid-encoded MgaP) and R6 harbouring plasmid pDLF*mgaPi* (chromosome-encoded MgaP) (Methods, Section 4.6.3). As expected, the expression level of the *mgaP* gene was found to be higher in R6 harbouring pDLF*mgaP*. The fold change (log_2FC) in gene expression due to the presence of pDLF*mgaP* was ~1.9. Thus, in the next sections, we will refer to R6/pDLF*mgaP* as a strain that produces high levels of MgaP, and to R6/pDLF*mgaPi* as a strain that produces low levels of MgaP. Both strains have been used in this Thesis to analyse the effect of MgaP on the expression of the *pclA* gene (Results, Chapter 3.4). In addition, plasmids pDLF*mgaP* and pDLF*mgaPi* were introduced into the R6Δ*mga* mutant strain, which lacks the regulatory *mgaSpn* gene (Solano-Collado *et al.*, 2012). Both strains, R6Δ*mga*/ pDLF*mgaP* (high levels of MgaP) and R6Δ*mga*/ pDLF*mgaPi* (low levels of MgaP), allowed us to analyse the effect of MgaP on the expression of *pclA* in the absence of Mga*Spn* (homologue of MgaP) (Results, Chapter 3.4).

3.2. Identification of the *PmgaP* promoter

In the R6 pneumococcal genome, whose sequence was published in 2001 (Hoskins *et al.*, 2001) (GenBank AE007317.1), the *spr1404* gene (named *mgaP* in this work) encodes a putative Mga-like regulatory protein (Gene ID: 934617). The *mgaP* gene is flanked by the genes *spr1403* (*pclA*, pneumococcal collagen-like protein A) and *spr1405* (peroxide stress protein YaaA) (Figure 29). According to the NCBI Entrez Genome Database, translation of the *mgaP* gene would start at coordinate 1388112 (ATG codon). However, we think that the ATG codon at coordinate 1388136 is likely the translation start site of the *mgaP* gene because it is preceded by a potential ribosome binding site sequence (T**GGAGG**) (Figure 29). Translation from this ATG codon would result in a protein of 494 residues (MgaP), as there is a stop codon (TAA) at coordinate 1389620. Moreover, the BPROM program (*Softberry, Inc.*; Table 6) predicted a promoter sequence upstream of the *mgaP* gene (promoter *PmgaP*). It has a canonical -10 element (**TATAAT**) and a possible -35 element (**TT**T**A**T**A**) at the suboptimal spacer length of 19 nucleotides (Figure 29).

To analyse whether the *PmgaP* promoter was functional *in vivo*, we constructed several transcriptional fusions based on the *gfp* reporter gene (Figure 31). First, a 185-bp DNA fragment (coordinates 1387937 to 1388121) was inserted into the pASTT promoter-probe vector (Methods, Section 4.6.4). The recombinant plasmid (pASTT-*PmgaP*) was then introduced into R6Δ*mga*/pDLF*mgaPi* (strain that produces low levels of MgaP) and R6Δ*mga*/pDLF*mgaP* (strain that produces high levels of MgaP) (see Results, Chapter 3.1). In both strains harbouring pASTT-*PmgaP*, *gfp* expression was ~2.5-fold higher than the basal level (strains harbouring pASTT, 0.08 ± 0.02 units) (Figure 31). This result indicated that (i) the 185-bp DNA fragment has promoter activity, and (ii) this promoter activity is not affected by higher levels of MgaP. Similar results were obtained with plasmid pASTT-*PmgaPΔ105*, which allowed us to conclude that the 80-bp region between coordinates 1388042 and 1388121 contains the *PmgaP* promoter (see also Figure 30). Finally, deletion of the region that spans coordinates 1388121 and 1388099 (pASTT-*PmgaPΔ-10*) removes the -10 element of the *PmgaP* promoter and, consequently, reduces the expression of *gfp* to basal levels (Figure 31). Hence, we conclude that the *PmgaP* promoter is functional under our bacterial growth conditions. Furthermore, it does not seem to be autoregulated.

Figure 31. Transcriptional fusions based on the *PmgaP* promoter region and the *gfp* reporter gene. The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The T1T2 box represents the tandem transcriptional terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Three regions from the R6 chromosome were inserted independently into the *Sac*I site of the promoter-probe vector pASTT. The coordinates of such regions are indicated. The grey arrow represents the -10 element of the *PmgaP* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} of 0.4). In each case, three independent cultures were analysed.

Next, to identify the *in vivo* transcription initiation site of the *mgaP* gene, we performed primer extension assays (Methods, Section 7.3). To this aim, we used total RNA isolated from R6 cells and the oligonucleotide Dw*1404*-2 (Table 4), which anneals to *mgaP* transcripts. As shown in Figure 32 (lane 2), a cDNA product of 114 nucleotides was detected, which could correspond to a transcription initiation event at coordinate 1388119. This coordinate is located 6 nucleotides downstream of the -10 element of the *PmgaP* promoter. Additionally, we performed primer extension assays using total

RNA isolated from R6 cells harbouring the pASTT-*PmgaP* plasmid, which carries the *gfp* gene under the control of the *PmgaP* promoter (see Figure 31). As primer, we used the oligonucleotide Int-*gfp* (Table 4), which anneals to *gfp* transcripts. As shown in Figure 32 (lane 1), a cDNA product of 105 nucleotides was detected, which could correspond to a transcription initiation event at coordinate 1388120. This coordinate is located 7 nucleotides downstream of the -10 element of the *PmgaP* promoter. In addition to the mentioned cDNA products, a possible non-specific product of 121 nucleotides was detected in both primer extension reactions (lanes 1 and 2). From these results we conclude that the pneumococcal RNA polymerase recognizes the *PmgaP* promoter and initiates transcription at coordinate 1388119/1388120.

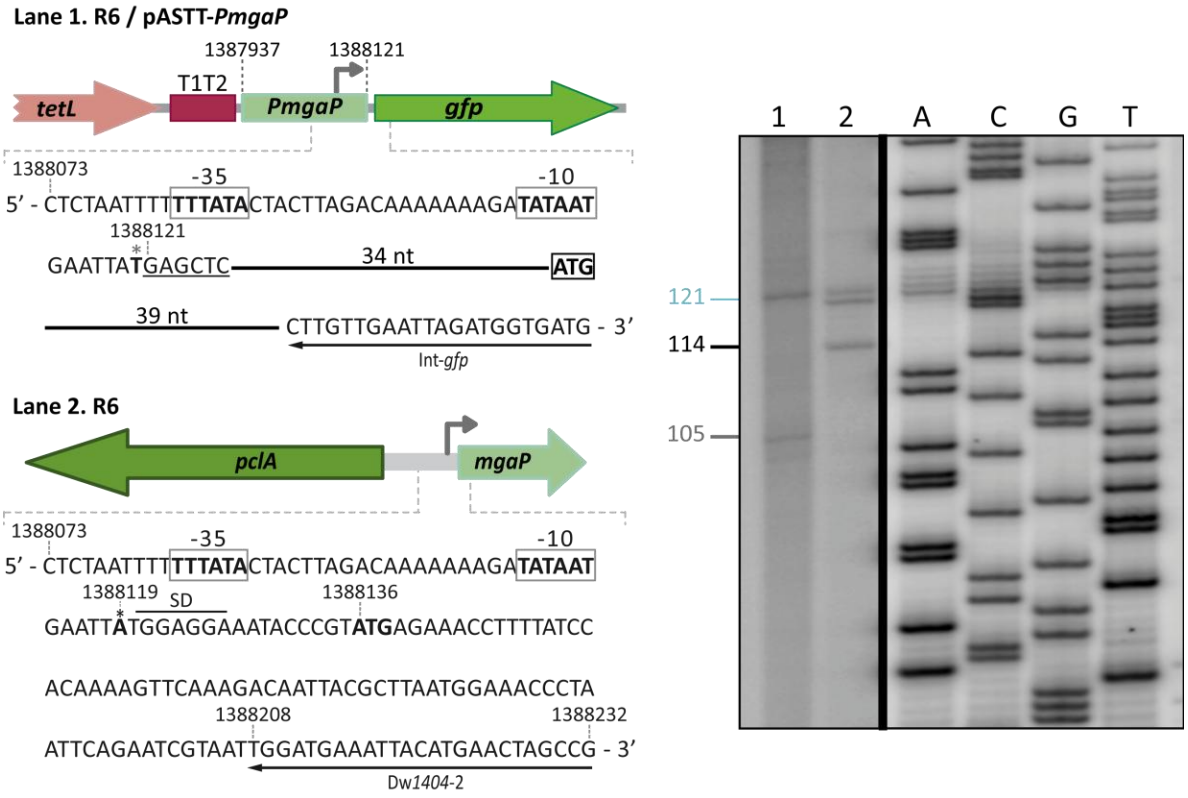


Figure 32. The *mgaP* gene is transcribed from the *PmgaP* promoter. Primer extension reactions were carried out using total RNA from R6 cells without plasmid (lane 2) and from R6 cells harbouring pASTT-*PmgaP* (lane 1). The size (in nucleotides) of the cDNA products (lanes 1 and 2) is indicated on the left of the gel. Dideoxy-mediated chain termination sequencing reactions (M13mp18 DNA and primer -40 M13) were used as DNA size markers (lanes A, C, G, T). The main sequence elements (-35 box and -10 box) of the *PmgaP* promoter and the translation start codon (ATG) of the *mgaP* and *gfp* genes are indicated in boldface letters in both sequences. SD: Shine-Dalgarno sequence. The asterisks indicate the 3'-end of the cDNA products synthesized using either oligonucleotide Dw*1404*-2 (black asterisk) or Int-*gfp* (grey asterisk). The *Sac*I site (GAGCTC) is underlined.

3.3. Prediction of functional domains in MgaP

The pneumococcal Mga*Spn* protein is a member of the Mga/AtxA family of global transcriptional regulators. This family includes Mga (*S. pyogenes*), AtxA (*B. anthracis*) and MafR (*E. faecalis*) (see Introduction and Figures 3, 4, 5 and 6). In addition to Mga*Spn* (493 residues), the R6 genome encodes MgaP (494 residues), which is likely a new member of the Mga/AtxA family. In fact, and according to predictions, the organization of functional domains in MgaP (Figure 33) resembles that found in the members of the Mga/AtxA family. Specifically, the Conserved Domain Database (CDD) and the Pfam Protein Families Database (El-Gebali *et al.*, 2018; Lu *et al.*, 2020) predicted that MgaP has two DNA-binding domains within the N-terminal region: the so-called HTH_Mga (Family PF08280, residues 6 to 65) and Mga (Family PF05043, residues 72 to 158) domains. They also predicted that the central region of MgaP contains a PTS Regulation Domain (PRD) (Family PRD_Mga PF08270, residues 174 to 391). Furthermore, analysis of the MgaP protein with the protein structure prediction server Phyre2 (Kelley *et al.*, 2015) revealed that its C-terminal region (residues 398 to 488) has structural homology to an EIIB-like component of the PTS (Figure 33). Two helix-turn-helix DNA-binding domains, one or two PRDs and an EIIB-like domain have been also identified in AtxA, Mga, Mga*Spn* and MafR (Figures 3, 4, 5 and 6) (Hammerstrom *et al.*, 2015; Hondorp *et al.*, 2013; Ruiz-Cruz *et al.*, 2016; Solano-Collado, 2014; Solano-Collado *et al.*, 2013). In Mga (*S. pyogenes*), the two DNA-binding domains were shown to be required for DNA binding and transcriptional activation (McIver and Myles, 2002; Vahling and McIver, 2006). Moreover, it has been shown that (i) the activity of AtxA (*B. anthracis*) is modulated by phosphorylation of histidine residues within the PRDs (Hammerstrom *et al.*, 2015; Tsvetanova *et al.*, 2007), and (ii) Mga is phosphorylated *in vivo* (Sanson *et al.*, 2015) and can be phosphorylated *in vitro* by components of the PTS (Hondorp *et al.*, 2013). These findings suggest that the PTS could control the activity of the MgaP protein.

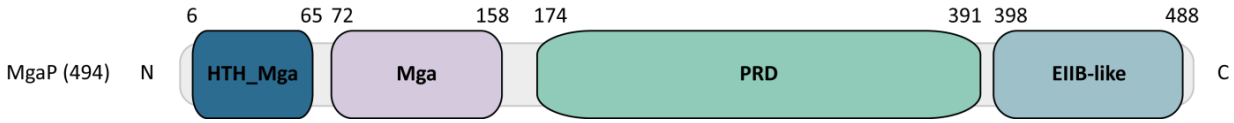


Figure 33. Predicted functional domains in MgaP. Predicted domains according to *in silico* analyses using Conserved Domain Database (CDD) (Lu *et al.*, 2020), Pfam (El-Gebali *et al.*, 2018) and Phyre2 (Kelley *et al.*, 2015) programs. MgaP contains two N-terminal DNA-binding domains (HTH_Mga and Mga domains). A central domain has structural homology to a PTS regulatory domain (PRD). The C-terminal region shows structural homology to an EIIB-like motif.

3.4. MgaP activates directly the transcription of the *plcA* gene

Studies performed in the pneumococcal R6 strain demonstrated that the Mga*Spn* protein acts directly as a positive transcriptional regulator (Solano-Collado *et al.*, 2012). Mga*Spn* enhances the activity of the *P1623B* promoter *in vivo* and, consequently, the expression of the *spr1623-spr1626* operon, which is adjacent to the regulatory *mgaSpn* gene. The *P1623B* promoter is divergent from the promoter where transcription of *mgaSpn* is initiated (promoter *Pmga*). Furthermore, *in vitro* DNA binding experiments showed that Mga*Spn* recognizes a DNA site located upstream of the *P1623B* promoter (positions -60 to -99) (Solano-Collado *et al.*, 2013). As mentioned above, the MgaP protein (494 residues) exhibits sequence similarity (60.3%) to the Mga*Spn* regulator (493 residues), and both proteins are predicted to have the same organization of functional domains (Results, Chapter 3.3). Based on these observations, and since the *pclA* (collagen-like protein) and *mgaP* genes, which constitute a specific cluster, are divergently transcribed (Paterson *et al.*, 2008) (Figure 34), we hypothesized that transcription of the *pclA* gene could be regulated by the MgaP protein. In this Chapter, we present evidence that supports this hypothesis.

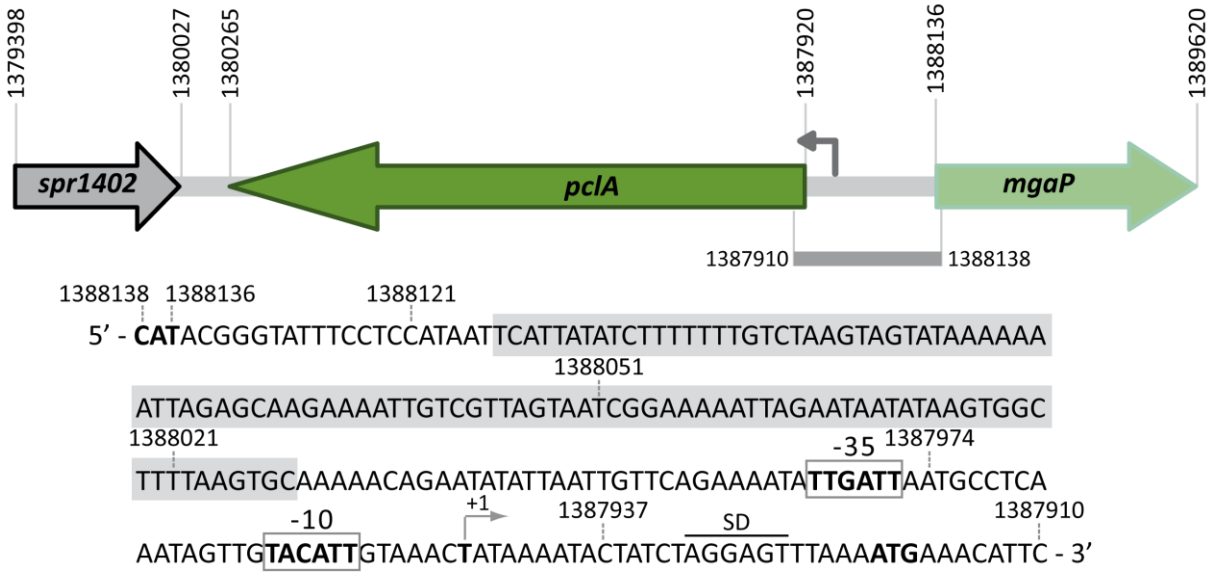


Figure 34. Genetic organization of the *pclA* region in the pneumococcal R6 chromosome. The nucleotide sequence of the region spanning coordinates 1388138 to 1387910 is shown. The grey arrow upstream of the *pclA* (*spr1403*) gene represents its promoter (*PpclA*). The translation start codon (ATG) of the *pclA* gene and the translation start codon (complementary to CAT) of the *mgaP* gene are indicated in boldface letters. SD: Shine-Dalgarno sequence. The transcription start site (+1 position) of the *pclA* gene and the main sequence elements (-35 box and -10 box) of the *PpclA* promoter are indicated. The MgaP-His binding site defined in this work is shown (grey shadowed box).

EFFECT OF MgaP ON THE TRANSCRIPTION LEVELS OF *pclA*

Paterson *et al.* reported that the pneumococcal gene *spr1403* (NC_003098.1) encodes a putative cell wall anchored protein, which contains large regions of collagen-like repeats, the number of which varies between strains. These authors named this protein PclA for pneumococcal collagen-like protein A. Moreover, they reported that a *pclA* mutant was defective in adherence and invasion of host cells (Paterson *et al.*, 2008). Some years later, Imai *et al.* investigated the distribution of *pclA* among 156 pneumococcal isolates (by PCR) and 11 referent pneumococcal strains (genomes totally sequenced) (Imai *et al.*, 2011). They found that the presence of *pclA* (33.3% of the isolates), was significantly associated with Pneumococcal Molecular Epidemiology Network clones (PMEN) (McGee *et al.*, 2001). This correlation suggested that *pclA* might contribute to the selection of prevalent clones (Imai *et al.*, 2011).

To analyse whether MgaP regulates the expression of the *pclA* gene, we performed qRT-PCR assays (Methods, Section 7.5). To this aim, we isolated total RNA from plasmid-harbouring strains: R6/pDLF*mgaP* (high levels of MgaP) and R6/pDLF*mgaPi* (low levels of MgaP) (see characteristics of both strains in Results, Chapter 3.1). The results obtained indicated that the expression of the *pclA* gene was up-regulated in the presence of high levels of MgaP. The fold change (log_2FC) in *pclA* expression due to high levels of MgaP was ~2.4. Additionally, we analysed the effect of MgaP on the transcription levels of *pclA* in the absence of the Mga*Spn* regulator using total RNA from R6Δ*mga*/pDLF*mgaP* (high levels of MgaP) and R6Δ*mga*/pDLF*mgaPi* (low levels of MgaP). The R6Δ*mga* strain lacks the *mgaSpn* gene (Solano-Collado *et al.*, 2012). Again, the log_2FC in *pclA* expression due to high levels of MgaP was ~3. Thus, we conclude that MgaP activates the transcription of *pclA in vivo*, both in the presence and in the absence of the Mga*Spn* regulator.

EFFECT OF MgaP ON THE ACTIVITY OF THE *PpclA* PROMOTER

To investigate whether MgaP activates the promoter of the *pclA* gene (promoter *PpclA*), we analysed the nucleotide sequence upstream of the *pclA* gene (Figure 34). The ATG codon at coordinate 1387920 (possible translation start site of *plcA*) is preceded by a putative ribosome binding site sequence (AGGAG). Translation from this ATG codon would result in a protein of 2,551 residues (PclA), as there is a stop codon (TAA) at coordinate 1380265. Furthermore, sequence analysis of the region spanning coordinates 1388224 to 1387910 revealed the existence of a putative promoter (*PpclA*). The -35 (**TTGA**TT) and -10 (**TA**C**A**T**T**) elements of this promoter are separated by 17 nucleotides (optimal length) (Figure 34).

To characterize the *PpclA* promoter, a 288-bp DNA fragment (coordinates 1388224 to 1387937) (Figure 35) was inserted into the pASTT promoter-probe vector, which is based on the *gfp* reporter gene (Methods, Section 4.6.4). The recombinant plasmid (pASTT-*PpclA*) (Tables 3 and 7) was first introduced into the R6 strain, and the expression of *gfp* was similar to the basal level (R6 harbouring pASTT: 0.07 ± 0.01 units). However, different results were obtained when pASTT-*PpclA* was introduced into R6Δ*mga*/pDLF*mgaPi* (strain that produces low levels of MgaP) and R6Δ*mga*/pDLF*mgaP* (strain that produces high levels of MgaP) (Figure 35). Compared to the strain that produces low levels of MgaP, the expression of *gfp* (plasmid pASTT-*PpclA*) was ~2.5-fold higher in the strain that produces high levels of MgaP. Moreover, the expression of *gfp* in the strain that produces low levels of MgaP was similar to the basal level (R6Δ*mga* harbouring both pDLF*mgaPi* and pASTT: 0.09 ± 0.03 units). Therefore, the 288-bp DNA fragment contains an MgaP-dependent promoter activity. Similar results were obtained with plasmids pASTT-*PpclAΔ103* and pASTT-*PpclAΔ173* (Table 7), which allowed us to conclude that the 115-bp region between coordinates 1388051 and 1387937 contains both the *PpclA* promoter and sequences required for its activation by MgaP (Figure 35). A further deletion analysis showed that (i) removal of the -10 element of the *PpclA* promoter resulted in loss of the MgaP-dependent promoter activity (plasmid pASTT-*PpclAΔ-10*; deletion between coordinates 1387974 and 1387937), and (ii) sequences located between coordinates 1388051 and 1388021 are needed for MgaP-mediated activation of the *PpclA* promoter (plasmid pASTT-*PpclAΔ203* and pASTT-*PpclAΔ224*) (see also Figure 34).

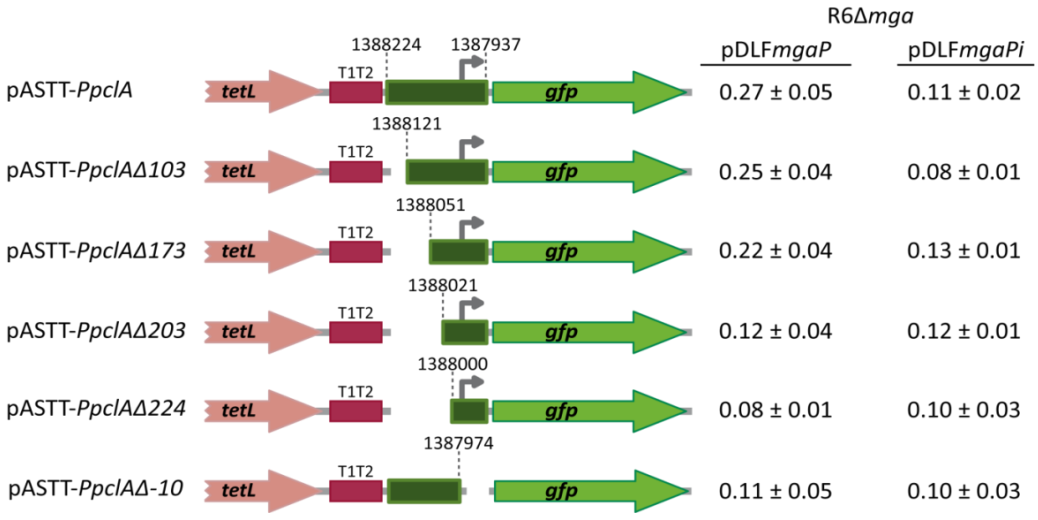


Figure 35. Deletion analysis of the *pclA* promoter region. The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The T1T2 box represents the tandem transcriptional terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Six regions from the R6 chromosome were inserted independently into the *Sac*I site of promoter-probe vector pASTT. The coordinates of such regions are indicated. The grey arrow represents the -10 element of the *PpclA* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} of 0.4). In each case, three independent cultures were analysed.

Finally, the functionality of the *PpclA* promoter located on the pASTT-*PpclAΔ103* plasmid (Figure 35) was also analysed by primer extension (Methods, Section 7.3) (Figure 36). As mentioned above, activity of the *PpclA* promoter (measured by *gfp* expression) was only detected when pASTT-*PpclAΔ103* was introduced into R6Δ*mga*/pDLF*mgaP* (strain that produces high levels of MgaP). Thus, we isolated total RNA from R6Δ*mga* cells harbouring both plasmids pDLF*mgaP* and pASTT-*PpclAΔ103*. When the oligonucleotide Int-*gfp* (it anneals to *gfp* transcripts) was used as a primer (Table 4), a cDNA product of 114 nucleotides was detected, which could correspond to a transcription initiation event at coordinate 1387946 (Figure 36). This coordinate is located 7 nucleotides downstream of the -10 element of the *PpclA* promoter. This result confirmed that the *PpclA* promoter is functional *in vivo* under our experimental conditions.

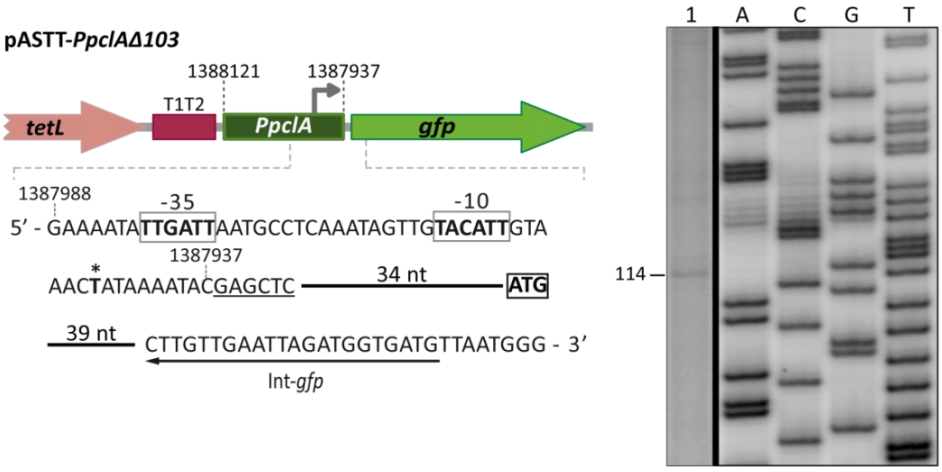


Figure 36. Identification of the *PpclA* promoter by primer extension. Primer extension reactions were carried out using total RNA from R6Δ*mga* cells harbouring both plasmids pDLF*mgaP* and pASTT-*PpclAΔ103*. The size (in nucleotides) of the cDNA product is indicated on the left of the gel. Dideoxy-mediated chain termination sequencing reactions (M13mp18 DNA and primer -40 M13) were used as DNA size markers (lanes A, C, G, T). The translation start codon (ATG) of the *gfp* gene and the main sequence elements (-35 box and -10 box) of the *PpclA* promoter are indicated in boldface letters. The black asterisk indicates the 3'-end of the cDNA product synthesized using the oligonucleotide Int-*gfp* (it anneals to *gfp* transcripts). The *Sac*I site (GAGCTC) is underlined.

PURIFICATION OF THE MgaP-HIS PROTEIN

The procedure used to overproduce and purify a His-tagged version of the MgaP protein (MgaP-His) was essentially based on the one described for $MafR_{OG1RF}$-His (Results, Chapter 1.1) (see also Methods, Section 8.2). Basically, the promoterless *mgaP* gene of the pneumococcal R6 strain was cloned into the *E. coli* inducible expression vector pET24b, which is based on the Φ10 promoter of the T7 phage (Figure 8). The recombinant plasmid, pET24b-*mgaP*-His, was introduced into the *E. coli*

BL21 (DE3) strain, which carries the gene that encodes the T7 RNAP under the control of the *lac*UV5 promoter (IPTG-inducible promoter). Thus, we used IPTG to induce the expression of *mgaP*-His. MgaP-His carries the Leu-Glu-6xHis peptide fused to its C-terminus. The protocol used to purify MgaP-His included the following steps: (i) precipitation of nucleic acids and MgaP-His with PEI at a low ionic strength (0.3 M NaCl); (ii) elution of MgaP-His from the PEI pellet with higher ionic strength (0.7 M NaCl), and (iii) fast-pressure liquid chromatography on a nickel affinity column. Protein fractions were analysed by SDS-polyacrylamide (12%) gel electrophoresis (Figure 37). MgaP-His migrates between the 45 and 66 kDa bands of the molecular weight marker (Figure 37), which is consistent with the molecular weight of the MgaP-His monomer calculated from the predicted amino acid sequence (59.94 kDa; 502 residues).

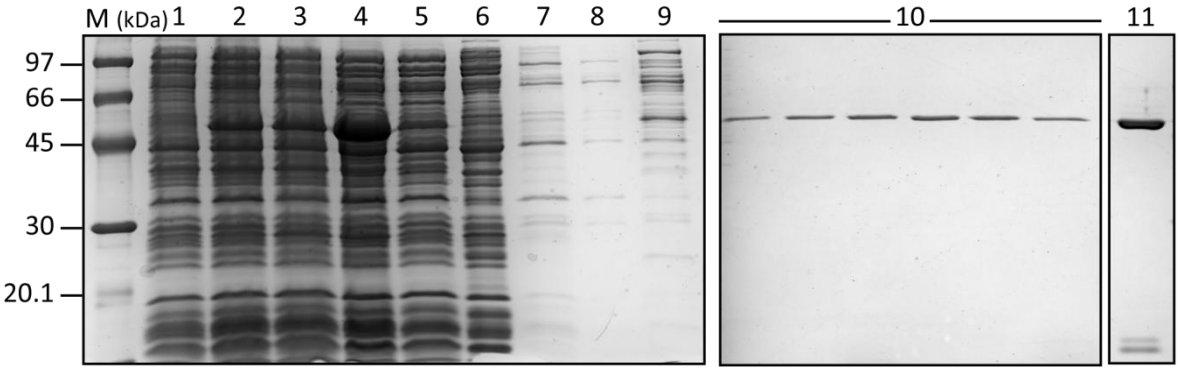


Figure 37. Purification of MgaP-His. Protein fractions were analysed by SDS-polyacrylamide (12%) gel electrophoresis. Gels were stained with Coomassie Blue. Lanes 1 to 3: Induction of *mgaP*-His gene expression: (1) uninduced culture; (2) induction with IPTG for 25 min; (3) after treatment with rifampicin for 60 min. Lanes 4 to 11: purification steps: (4) cleared cell lysate; (5) supernatant after PEI precipitation at low ionic strength; (6-8) proteins eluted from the PEI pellet using the same low ionic strength buffer; (9) proteins eluted from the PEI pellet using a higher ionic strength buffer; (lanes group 10) nickel affinity chromatography: proteins retained in the column were eluted with a 10 mM to 250 mM imidazole gradient; (11) protein preparation after concentration. M indicates the molecular weight standards (in kDa) (LMW Marker, GE Healthcare).

BINDING OF MgaP TO THE *PpclA* PROMOTER REGION

To investigate whether MgaP binds to the *PpclA* promoter region, we carried out EMSA experiments using MgaP-His and a 270-bp DNA fragment (coordinates 1388196 to 1387927) that contains both the *PpclA* promoter and the site required for its activation by MgaP *in vivo* (Figure 35). The 270-bp DNA fragment was radioactively labelled at the 5'-end of the coding strand. First, the labelled DNA (2 nM) was incubated with increasing concentrations of MgaP-His (200 to 800 nM) in the absence of competitor DNA (Methods, Section 10.1). Free and bound DNAs were separated by electrophoresis on a native polyacrylamide (6%) gel. As shown in Figure 38A, four DNA-protein complexes (C1 to C4)

were observed at low concentration of MgaP-His (200 nM). However, at 400 nM such complexes disappeared and higher-order complexes appeared. Additionally, the labelled DNA (2 nM) was incubated with increasing concentrations of MgaP-His (200 to 800 nM) in the presence of non-labelled competitor calf thymus DNA (2 µg/ml) (Figure 38B). A similar pattern of DNA-protein complexes was visualized, but the disappearance of faster moving complexes required higher concentrations of MgaP (500 mM). These results indicated that MgaP-His generates multimeric complexes on linear dsDNA, a feature reported for its homologue, the Mga*Spn* regulator (Solano-Collado *et al.*, 2013).

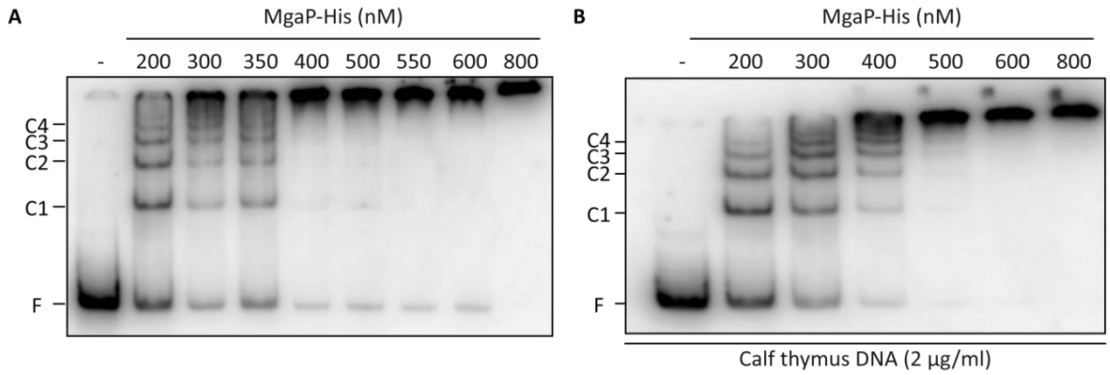


Figure 38. Binding of MgaP-His to the *PpclA* promoter region. The 270-bp DNA fragment (*PpclA* promoter, coordinates 1388196 to 1387927) was radiolabelled at the 5'-end of the coding strand. Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. Labelled-DNA was visualized using a Fujifilm Image Analyzer (FLA-3000). **A. Formation of multimeric MgaP-His-DNA complexes in the absence of competitor DNA.** The labelled DNA fragment (2 nM) was incubated with increasing concentrations of MgaP-His. Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated. **B. Formation of multimeric MgaP-His-DNA complexes in the presence of competitor DNA.** The labelled DNA fragment (2 nM) was incubated with increasing concentrations of MgaP-His in the presence of non-labelled competitor calf thymus DNA. Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C4) are indicated.

The interaction of MgaP-His with the *PpclA* promoter region was further analysed by DNase I footprinting experiments (Methods, Section 10.2) (Figure 39). First, the aforementioned 270-bp DNA fragment (coordinates 1388196 to 1387927) was radioactively labelled at the 5'-end of the coding strand, and the labelled DNA (2 nM) was incubated with increasing concentrations of MgaP-His. On the coding strand and at 400 nM of MgaP-His, protections against DNase I digestion were observed at a particular region (from positions -169 to -68 relative to the transcription start site of the *PpclA* promoter). Moreover, positions -139 and -103 were more sensitive to DNase I cleavage. Next, a 281-bp DNA fragment (coordinates 1388232 to 1387952) was radioactively labelled at the 5'-end of the non-coding strand, and the labelled DNA (2 nM) was incubated with increasing concentrations of MgaP-His (Figure 39). On the non-coding strand and at 400 nM of MgaP-His, major changes in

DNase I sensitivity (diminished cleavages) were observed from positions -152 to -83. On both strands and at 800 nM of MgaP-His, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the ability of MgaP-His to generate multimeric complexes (see Figure 38). Thus, MgaP-His recognizes a specific DNA site (primary site) upstream of the *PpclA* core promoter. Such a DNA site is located between the positions -169 and -68 (Figure 39). Since sequences located between the positions -105 and -75 (coordinates 1388051 and 1388021) are needed for MgaP-mediated activation of the *PpclA* promoter *in vivo* (see Figure 35), we conclude that MgaP activates directly the transcription of the *pclA* gene.

3.5. MgaP and Mga*Spn* recognize different DNA sites on the R6 chromosome

BINDING OF MgaP TO THE *P1623B* PROMOTER REGION

Footprinting experiments showed that the Mga*Sn* regulator recognizes a site located upstream of the *P1623B* promoter (positions -60 to -99) (Solano-Collado *et al.*, 2013) (see nucleotide sequence in Figure 40). Binding of Mga*Spn* to this site (known as the *PB* activation region) is required to activate the *P1623B* promoter *in vivo*, and consequently to activate the transcription of the *spr1623-spr1626* operon (Solano-Collado *et al.*, 2012). Due to the high level of sequence similarity between Mga*Spn* and MgaP (60.3%), we analysed whether MgaP recognized the *PB* activation region. To this aim, we performed DNase I footprinting assays using MgaP-His and a 222-bp DNA fragment (coordinates 1598298 to 1598519 of R6) that has the *PB* activation region at internal position. Solano-Collado *et al.* used this DNA fragment to demonstrate that Mga*Spn* recognizes the *PB* activation region as a primary binding site (Solano-Collado *et al.*, 2013). The 222-bp DNA fragment was radioactively labelled at the 5'-end of the coding strand (relative to the *P1623B* promoter), and the labelled DNA (2 nM) was incubated with increasing concentrations of MgaP-His. As shown in Figure 40, specific regions protected against DNase I digestion were not observed, indicating that MgaP-His does not recognize the *PB* activation region. Nevertheless, by EMSA experiments, we found that MgaP-His is able to generate multimeric complexes on the 222-bp DNA fragment (Figure 41). The pattern of complexes generated by MgaP-His was similar to that shown in Figure 38 (binding of MgaP-His to the *PpclA* promoter region). This DNA-binding behaviour is similar to the one described for Mga*Spn* (Solano-Collado *et al.*, 2013) and MafR (see Results, Chapter 1). Like these regulators, MgaP can generate multimeric complexes on linear dsDNAs without showing a preference for a specific DNA site.

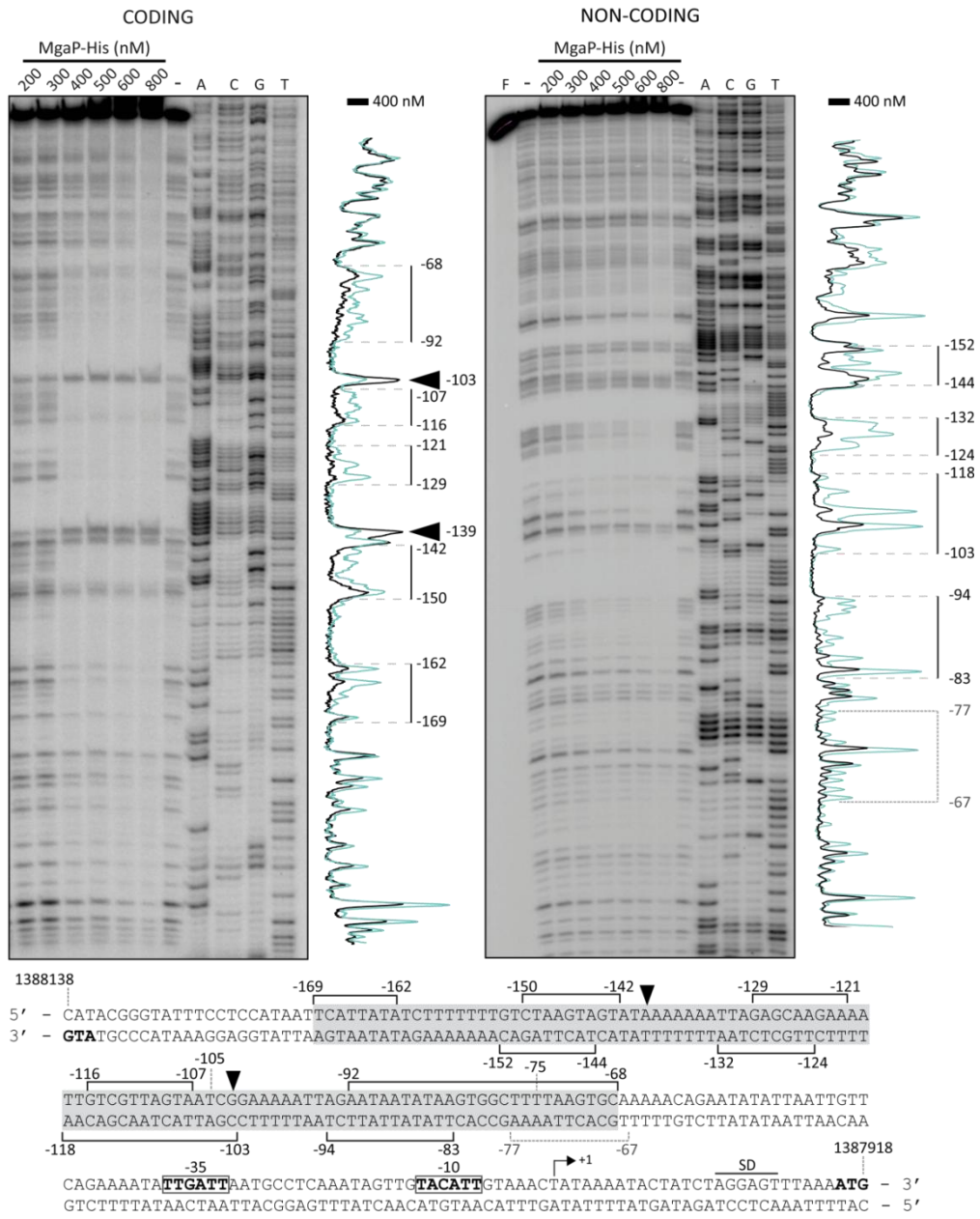


Figure 39. DNase I footprints of complexes formed by MgaP-His on the *PpclA* promoter region. The 270-bp DNA fragment (coordinates 1388196 to 1387927) was radioactively labelled at the 5'-end of the coding strand. The 281-bp DNA fragment (coordinates 1388232 to 1387952) was radioactively labelled at the 5'-end of the non-coding strand. The labelled DNAs (2 nM) were incubated with the indicated concentrations of MgaP-His and then they were digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 823-bp PCR-amplified DNA fragment (R*mgaP*-q and R*pclA*-S primers) and the ^{32}P-labelled Dw*1404* oligonucleotide (coding) or the ^{32}P-labelled Up*1404*-2 oligonucleotide (non-coding). Densitometer scans corresponding to free DNA (turquoise line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 1388138 to 1387918 is shown. The transcription initiation site (+1 position) of the *pclA* gene and the -35 and -10 boxes of the *PpclA* promoter are indicated. SD: Shine-Dalgarno sequence. The start codons (ATG) of the *mgaP* (1388136) and *pclA* (1387920) genes are indicated in boldface letters. Black brackets indicate the main regions protected against DNase I digestion. Sites more sensitive to DNase I cleavage are indicated with arrowheads. The region (from -169 to -68) that contains the site recognized by MgaP-His is indicated with a grey box.

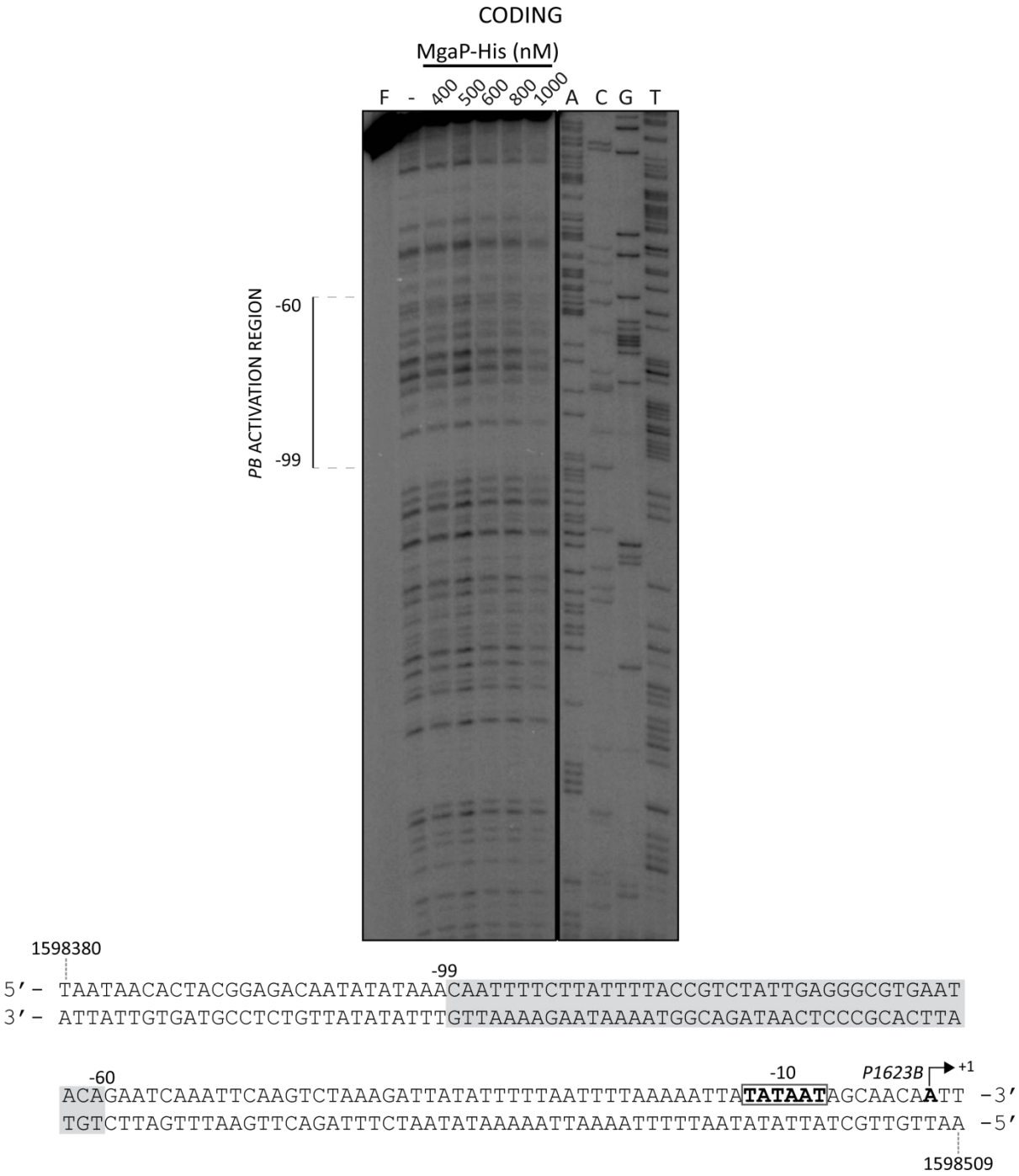


Figure 40. DNase I footprinting assay using MgaP-His and a DNA fragment that contains the *PB* activation region. The 222-bp DNA fragment (coordinates 1598298 to 1598519) was radioactively labelled at the 5'-end of the coding strand (relative to the *P1623B* promoter). The labelled DNA (2 nM) was incubated with the indicated concentrations of MgaP-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 1,418-bp PCR-amplified DNA fragment (*1622*A and *1622*B primers) and the ^{32}P-labelled *1622*H oligonucleotide (coding). All the lanes came from the same gel (delineation with dividing lines). The nucleotide sequence of the region spanning coordinates 1598380 to 1598509 is shown. The transcription initiation site (+1 position) of the *spr1623* gene and the -10 element of the *P1623B* promoter are indicated. The *PB* activation region (from -99 to -60) that contains the site recognized by the Mga*Spn* regulator is indicated with a grey box in the nucleotide sequence and with a bracket on the left of the gel.

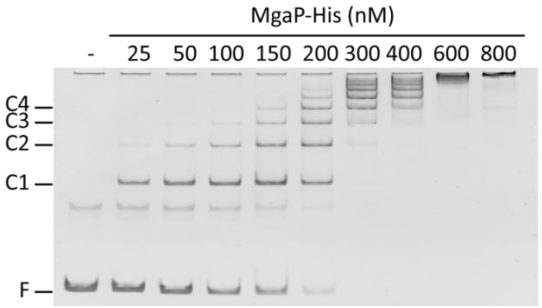


Figure 41. Formation of multimeric MgaP-His-DNA complexes on the *P1623B* promoter region. 10 nM of the 222-bp DNA fragment (*P1623B* promoter, coordinates 1598298 to 1598519) was mixed with increasing concentrations (25 to 800 nM) of MgaP-His. Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 and C4) are indicated.

The inability of MgaP-His to recognize the *PB* activation region *in vitro* (Figure 40) correlated with the inability of MgaP to influence the transcription of the *spr1623* gene *in vivo* (first gene of the *spr1623-spr1626* operon). Firstly, by qRT-PCR assays (Methods, Section 7.5), we compared the relative expression of the *spr1623* gene, as well as its adjacent *mgaSpn* regulatory gene, in two strains: R6/pDLF*mgaP* (high levels of MgaP) and R6/pDLF*mgaPi* (low levels of MgaP). No differences in *spr1623* and *mgaSpn* expression were observed. In addition, we analysed the effect of MgaP on the transcription of *spr1623* in the absence of the Mga*Spn* regulator using total RNA from R6Δ*mga*/pDLF*mgaP* cells (high levels of MgaP) and from R6Δ*mga*/pDLF*mgaPi* cells (low levels of MgaP). The transcription levels of *spr1623* were similar in both genetic backgrounds.

BINDING OF Mga*Spn* TO THE *PpclA* PROMOTER REGION

By DNase I footprinting experiments, we found that MgaP-His recognizes a specific DNA site upstream of the *PpclA* core promoter (positions -68 to -169). Such a region contains sequences that are essential for MgaP-mediated activation of the *PpclA* promoter (Results, Chapter 3.4). Next, we studied the interaction of the Mga*Spn* regulator (MgaP homologue) with the *PpclA* promoter region using a His-tagged version of Mga*Spn* (Mga*Spn*-His) and the 270-bp DNA fragment (coordinates 1388196 to 1387927), which contains both the *PpclA* promoter and the MgaP-His binding site (see Figure 39). The presence of a His-tag at the C-terminal end of Mga*Spn* does not affect its DNA-binding properties (Solano-Collado *et al.*, 2016). The 270-bp DNA fragment was radioactively labelled at the 5'-end of the coding strand, and the labelled DNA (2 nM) was incubated with different concentrations of Mga*Spn*-His protein. First, EMSA experiments showed that Mga*Spn*-His generates multimeric complexes on the 270-bp DNA (Figure 42), in agreement with previous results (Solano-Collado *et al.*, 2016; Solano-Collado *et al.*, 2013). Furthermore, DNase I footprinting assays revealed that Mga*Spn*-His is able to recognize particular regions on the 270-bp DNA fragment (Figure 43). On

the coding strand and at 75 nM of Mga*Spn*-His, major changes in DNase I sensitivity (diminished cleavages) were observed from positions -173 to -196, and from positions -102 to -115. In addition, positions -47, -69, -87 and -131 were slightly more sensitive to DNase I digestion. This result was further confirmed in shorter electrophoretic runs (Figure 44). The interaction of Mga*Spn*-His with the *PpclA* promoter region was also analysed using the 281-bp DNA fragment (coordinates 1388232 to 1387952), which was radioactively labelled at the 5'-end of the non-coding strand (Figure 43). At 100 nM of Mga*Spn*-His, diminished DNase I cleavages were mainly observed from positions -172 to -213, and from -103 to -110. Moreover, positions -87, -88, -126, -145, -160, -173, -245 and -251 were more sensitive to DNase I digestion. On both strands and at higher concentrations of Mga*Spn*-His (Figure 43), protections against DNase I digestion were observed along the DNA fragments, which is consistent with the pattern of protein-DNA complexes observed by EMSA (Figure 42). Taken together, these results indicated that Mga*Spn*-His recognizes preferentially two regions (Figure 45). One of them (Site A; positions -173 to -213) is adjacent to the MgaP-His binding site (positions -68 to -169), whereas the other one (Site B; positions -102 to -115) is included within the MgaP-His binding site. Thus, Mga*Spn* and MgaP have different DNA-binding specificities on the *PpclA* promoter region.

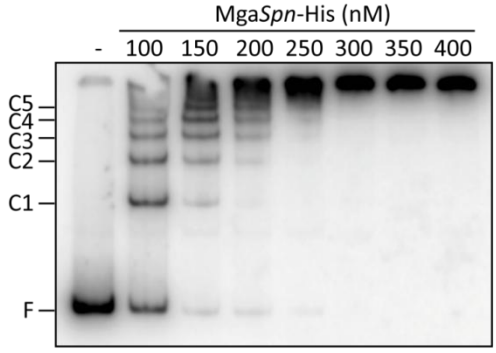


Figure 42. Binding of Mga*Spn*-His to the *PpclA* promoter region. EMSA was carried out incubating Mga*Spn*-His (100 to 400 nM) and 2 nM of the ^{32}P-labelled 270-bp DNA fragment (*PpclA* promoter, coordinates 1388196 to 1387927). Free and bound DNAs were separated by native polyacrylamide (6%) gel electrophoresis and labelled DNA was visualized using a Fujifilm Image Analyzer (FLA-3000). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1 to C5) are indicated.

By qRT-PCR assays, we found that the Mga*Spn* regulator does not influence the expression of the *pclA* and *mgaP* genes. Specifically, we determined the relative expression of both genes in two strains: R6Δ*mga*/pDLF287 (absence of Mga*Spn*) and R6Δ*mga*/pDL*PsulA*::*mga* (high levels of plasmid-encoded Mga*Spn*), which were previously used to analyse the effect of Mga*Spn* on the activity of particular promoters (Solano-Collado *et al.*, 2012). The transcription levels of *pclA* and *mgaP* were similar in both strains, suggesting that the *in vitro* binding of Mga*Spn* to the *PpclA* promoter region (Figure 45) is not functional *in vivo*.

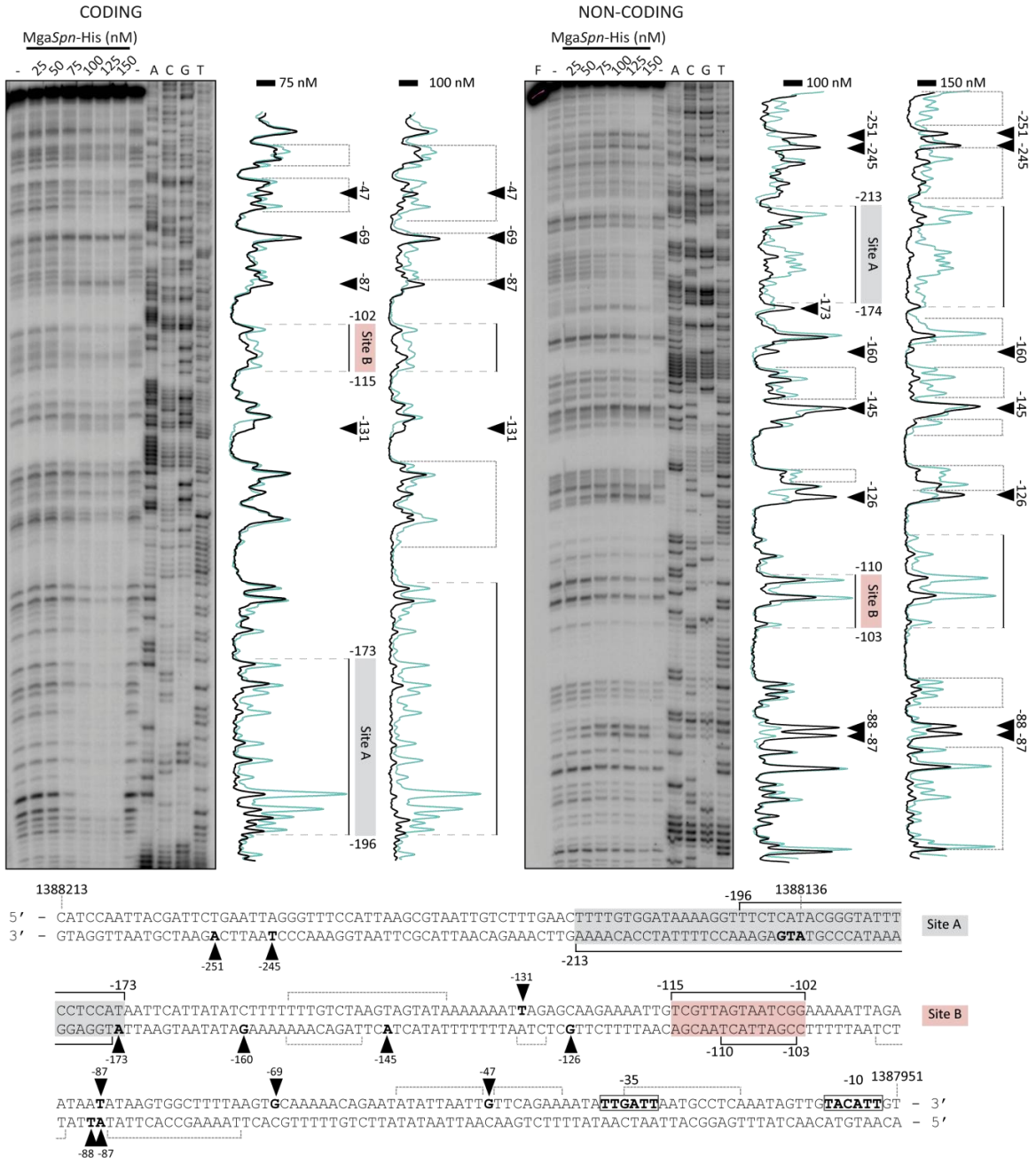


Figure 43. DNase I footprints of complexes formed by Mga*Spn*-His on the *PpclA* promoter region. DNA fragments containing the *PpclA* promoter were radioactively labelled at the 5'-end of the coding strand (270-bp DNA fragment, coordinates 1388196 to 1387927) or at the 5'-end of the non-coding strand (281-bp DNA fragment, coordinates 1388232 to 1387952). The labelled DNAs (2 nM) were incubated with the indicated concentrations of Mga*Spn*-His and then they were digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 823-bp PCR-amplified DNA fragment (R*mgaP*-q and R*pclA*-S primers) and the ^{32}P-labelled Dw*1404* oligonucleotide (coding) or the ^{32}P-labelled Up*1404*-2 oligonucleotide (non-coding). Densitometer scans corresponding to free DNA (turquoise line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 1388213 to 1387951 is shown. The translation start codon (ATG) of *mgaP* (1388136) and the -35 and -10 boxes of the *PpclA* promoter are indicated in boldface letters. Black brackets indicate the main regions protected against DNase I digestion. Sites more sensitive to DNase I cleavage are indicated with arrowheads. The sites recognized by Mga*Spn*-His are indicated with a shadowed box: site A (grey box; from -173 to -213) and site B (red box; from -102 to -115).

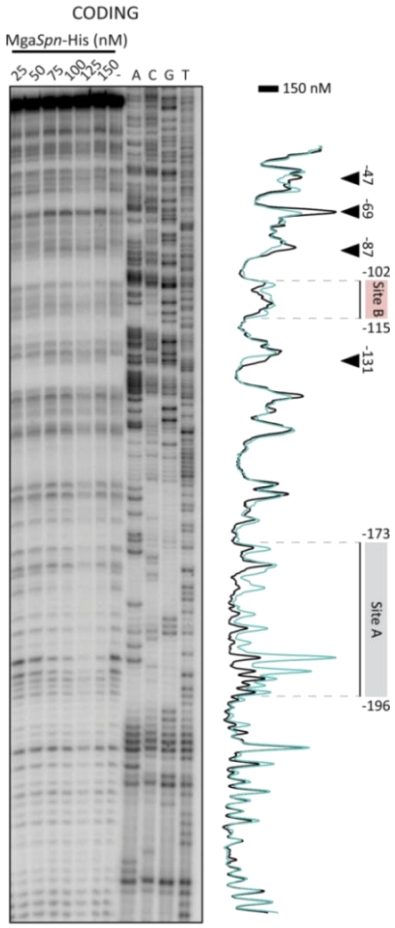


Figure 44. Sites recognized by Mga*Spn*-His on the *PpclA* promoter region. The 270-bp DNA fragment (*PpclA* promoter, coordinates 1388196 to 1387927) was radioactively labelled at the 5'-end of the coding strand. The labelled DNA (2 nM) was incubated with the indicated concentrations of Mga*Spn*-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (lanes A, C, G, and T) were run in the same gel. Sequencing reactions were prepared using a 823-bp PCR-amplified DNA fragment (R*mgaP*-q and R*pclA*-S primers) and the ^{32}P-labelled Dw*1404* oligonucleotide (coding). Black brackets indicate the main regions protected against DNase I digestion. Sites more sensitive to DNase I cleavage are indicated with arrowheads. The sites recognized by Mga*Spn*-His are indicated with a shadowed box: site A (grey box; from -173 to -196) and site B (red box; from -102 to -115).

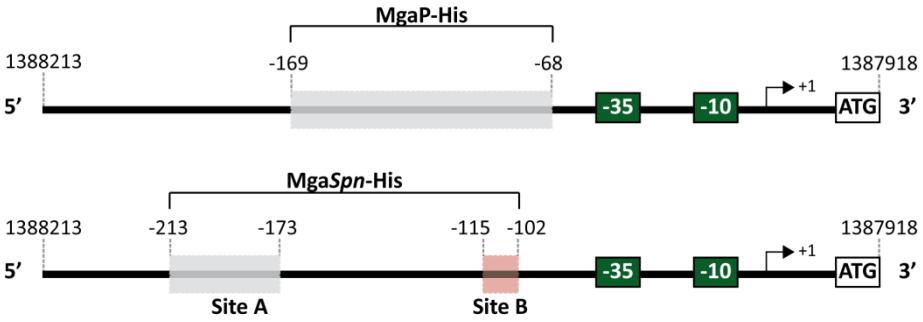


Figure 45. Scheme of the binding sites of Mga*Spn*-His and MgaP-His on the *PpclA* promoter region. The region spanning coordinates 1387918 to 1388213 of the pneumococcal R6 genome is represented. This region includes the *PpclA* core promoter and the transcription initiation site (+1 position) of the *pclA* gene identified in this work. The shadowed boxes indicate the Mga*Spn*-His and the MgaP-His binding sites defined by DNase I footprinting assays (see Figure 43 and Figure 39, respectively).

DISCUSSION

E. faecalis and *S. pneumoniae* are able to proliferate in different niches of the human host, either as commensals or as leading causes of serious infections. The adaptive capacity of both bacteria is closely associated with their opportunistic lifestyle. Although some regulatory circuits involved in the adaptive responses of both bacteria have been studied in detail, many of them remain poorly defined. This Thesis has been focused on the functional characterization of MafR (*E. faecalis*) and MgaP (*S. pneumoniae*), two new members of the Mga/AtxA family of global transcriptional regulators. We have identified three genes regulated directly by MafR and one gene regulated directly by MgaP. The known or predicted function of these target genes suggests a regulatory role of both proteins in adaptation/colonization processes. In addition, we have studied the interaction of MafR and MgaP with DNA. Our results support that a characteristic of the regulators that belong to the Mga/AtxA family is their ability to recognize particular DNA structures rather than specific nucleotide sequences.

MafR is a transcriptional activator

Bacterial adaptation to new niches and to environmental fluctuations requires global changes in gene expression. Proteins that function as global transcriptional regulators are key elements in such processes due to their ability to activate and/or repress the expression of multiple genes in response to specific signals. *E. faecalis* is a well-adapted bacterium that carries genes necessary to survive and to colonize its habitats, as well as to persist in the hospital environment (Ruiz-Garbajosa *et al.*, 2006). Various transcriptome analyses support that its adaptation is associated with global changes in gene expression. For instance, growth of *E. faecalis* in human blood, human urine, and intestine of living mice was shown to have a great impact on the transcription of numerous genes (Lindenstrauß *et al.*, 2014; Vebø *et al.*, 2009; Vebø *et al.*, 2010).

MafR is a member of the Mga/AtxA family of global regulators. Previous studies of our laboratory demonstrated that MafR is involved in global regulation of gene expression (Ruiz-Cruz, 2015; Ruiz-Cruz *et al.*, 2016). MafR activates, directly or indirectly, the transcription of numerous genes on a genome-wide scale (at least 87 genes). Many of such genes encode components of PTS-type membrane transporters (15 genes), components of ABC-type membrane transporters (9 genes), and proteins involved in the metabolism of carbon sources (18 genes) (Ruiz-Cruz *et al.*, 2016). Interestingly, some of these genes were found to be up-regulated during the growth of *E. faecalis* in blood and/or in human urine (Vebø *et al.*, 2009; Vebø *et al.*, 2010). For example, MafR influences positively the expression of the putative operon *OG1RF_11135-33* (ABC transporter, sugar substrate), which was shown to be up-regulated during the growth of *E. faecalis* in blood (Vebø *et al.*, 2009), and

highly induced during peritoneum infections (Muller *et al.*, 2015). Besides, genes *malB* and *malP* of the *malPBMR* operon (maltose metabolism) and its divergent-adjacent gene *malT* (PTS transporter) were shown to be up-regulated during the growth of *E. faecalis* in blood (Vebø *et al.*, 2009) and down-regulated in MafR-lacking cells (Ruiz-Cruz *et al.*, 2016). Also, the *malPBMR* operon was found to be slightly up-regulated during the growth of *E. faecalis* in human urine (Vebø *et al.*, 2010) and the growth of a MafR-lacking strain was shown to be impaired in media containing maltose (Ruiz-Cruz *et al.*, 2016).

In this Thesis, by qRT-PCR, transcriptional fusions and DNase I footprinting experiments, we have demonstrated that MafR acts as a transcriptional activator. It activates directly the transcription of three genes: *OG1RF_12294*, *OG1RF_11486* and *OG1RF_10478*. Gene *OG1RF_12294* encodes a putative P-type ATPase cation transporter that has sequence similarity to several eukaryotic and prokaryotic proteins characterized as calcium P-type ATPases (see Table S4). Thus, MafR could have a regulatory role in maintaining cellular calcium homeostasis (Figure 46). Although the role of calcium ions in prokaryotes remains elusive, some studies have shown that Ca^{2+} affects several physiological processes, such as division, secretion, transport, and stress response (Domínguez *et al.*, 2015).

Gene *OG1RF_11486* encodes a putative ECF transporter S-component, likely involved in the uptake of a queuosine precursor. Hence, MafR could have an additional regulatory role in the biosynthesis of queuosine, a modified nucleoside found at the wobble position of particular transfer RNAs (Hutinet *et al.*, 2017). There is evidence that the queuosine contributes to the efficiency of protein synthesis (Figure 46). Furthermore, a study using *Shigella flexneri* has revealed that the intracellular concentration of the virulence-related transcriptional regulator VirF is reduced in the absence of queuosine (Durand *et al.*, 2000). Although the lack of queuosine does not seem to be critical in exponential growth, it has been reported that it affects the growth of some bacteria under stress conditions (Noguchi *et al.*, 1982; Thibessard *et al.*, 2004).

Gene *OG1RF_10478* encodes a protein of unknown function. However, according to our bioinformatic analyses, OG1RF_10478 might be a putative PTS EIIB cellobiose-specific component. This prediction suggests a link between MafR and the transport of cellobiose (Figure 46). In general, the EIIB component is one of the three domains that constitute the carbohydrate-specific EII complex of the PTS, which uses phosphoenolpyruvate (PEP) rather than ATP as an energy source to drive translocation. The PTS not only transports and phosphorylates carbohydrates but also regulates numerous cellular processes (Galinier and Deutscher, 2017). A study performed with the EIIB cellobiose-specific component of *S. agalactiae* has shown that this component not only contributes to biofilm formation and virulence regulation, but also plays an important role in the invasion and

colonization of *S. agalactiae* (Xu *et al.*, 2019). In the case of enterococci, the ability to metabolize numerous carbohydrates enables them to colonize diverse environments (Ramsey *et al.*, 2014). *E. faecalis* V583 has 35 probable PTS-type sugar transporters (Paulsen *et al.*, 2003) and, as mentioned above, microarrays studies designed for *E. faecalis* OG1RF showed that MafR influences positively the transcription of at least 15 genes that encode components of PTS transporters (Ruiz-Cruz *et al.*, 2016). The MafR regulator has integrated a putative PTS-recognized phosphorylation domain (known as the PRD domain). PRD domains have been also found in the other members of the Mga/AtxA family (see Figure 49), which suggests that they might sense changes in carbon source availability. A transcriptome analysis revealed that Mga activates or represses the expression of various genes involved in the utilization of sugars (Ribardo and McIver, 2006). Moreover, it has been shown that Mga is phosphorylated *in vivo* (Sanson *et al.*, 2015) and can be phosphorylated *in vitro* by components of the PTS (Hondorp *et al.*, 2013). Also, the activity of AtxA is modulated by phosphorylation of histidine residues within PRD domains (Tsvetanova *et al.*, 2007).

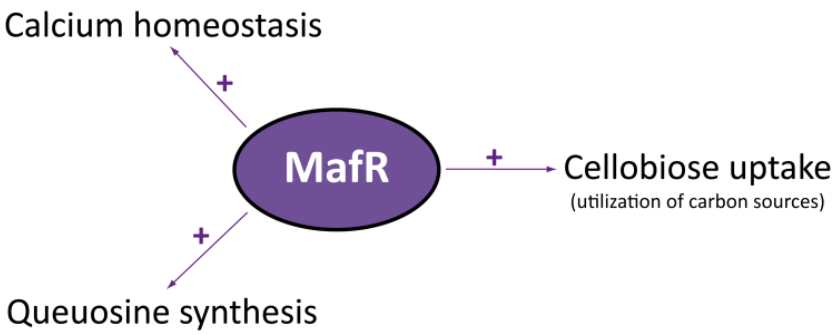


Figure 46. Regulatory role of MafR. A compilation of the possible regulatory role of MafR based on the predicted function of its directly-regulated target genes in *E. faecalis*. The arrows indicate that MafR could have a positive effect on the mentioned processes.

Bacteria use a variety of mechanisms to activate transcription from specific promoters. Genetic and biochemical studies have shown that some proteins stimulate transcription by binding to a specific DNA site either upstream of or overlapping the core promoter (Browning and Busby, 2016). In this Thesis, we have performed DNase I footprinting experiments to demonstrate that MafR recognizes a specific site on linear DNA fragments that contain the promoter under study (*P12294*, *P11486* and *P10478*). By this approach, we have established that MafR activates directly the transcription of the three target genes by binding to a specific DNA site overlapping the core promoter (Figure 47). These results suggest that MafR might enhance the efficiency of the target promoters by recruitment of RNA polymerase through direct interactions with the sigma factor. Besides, MafR might induce conformational changes in the target promoters, as it has been described for some transcriptional

activators (Browning and Busby, 2016). Transcriptional activation from specific promoters has also been reported for other members of the Mga/AtxA family. The pneumococcal Mga*Spn* regulator stimulates transcription of a four-gene operon (*spr1623-spr1626*) by binding to a specific DNA site upstream of the promoter (position -60 to -99) (Solano-Collado *et al.*, 2013). Regarding the Mga regulator from *S. pyogenes*, the position of its DNA-binding site with respect to the start of transcription varies among the promoters tested. Nevertheless, the majority of the promoters contain a Mga binding site located around the position -54, thereby overlapping the -35 element of the promoter (Hondorp and McIver, 2007).

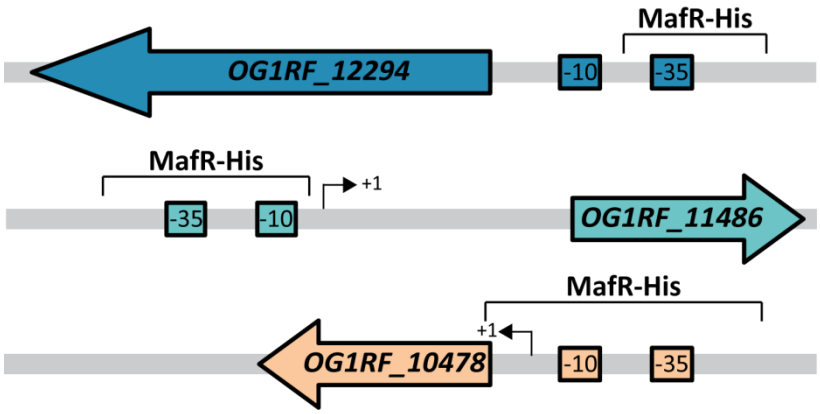


Figure 47. MafR-binding sites. MafR activates directly the transcription of the three target genes *OG1RF_12294*, *OG1RF_11486* and *OG1RF_10478* by binding to a specific DNA site overlapping the core promoter (*P12294*, *P11486* and *P10478*).

DNA-binding properties of MafR

The interaction of MafR with DNA was investigated previously in our laboratory using MafR (untagged form) encoded by the *E. faecalis* V583 genome. This protein forms dimers in solution (Ruiz-Cruz, 2015). Gel retardation assays showed that $MafR_{V583}$ binds to linear dsDNAs with low sequence specificity generating multimeric complexes. The pattern of such complexes suggested that multiple units of $MafR_{V583}$ (likely dimers) bind sequentially to the same DNA molecule (Ruiz-Cruz, 2015). In this Thesis, we have analysed the DNA-binding properties of MafR encoded by the *E. faecalis* OG1RF genome using a His-tagged version of the protein ($MafR_{OG1RF}$-His) and various linear dsDNAs. Compared to $MafR_{V583}$, $MafR_{OG1RF}$ has three amino acid changes (Ala37Thr, Gln131Leu, Met145Thr) within the predicted DNA-binding domain (residues 11 to 164). By gel retardation assays, we found that $MafR_{OG1RF}$-His was able to generate multimeric complexes on all the tested DNAs. This result indicated that neither the presence of the His-tag at the C-terminus nor the three amino acid substitutions affect the DNA-binding behaviour of MafR. In addition, we found that removal of the

first three amino acids residues results in a variant that is unable to generate multimeric complexes. In the Mga/AtxA family of transcriptional regulators, two additional proteins are known to generate multimeric complexes on linear dsDNAs, Mga*Spn* (Solano-Collado *et al.*, 2013) and MgaP (this Thesis). Both pneumococcal proteins bind to DNA in a non-sequence-specific manner. Formation of multimeric protein-DNA complexes has not been reported for the Mga and AtxA regulators. However, protein-protein interaction studies have shown that Mga oligomerizes in solution (Hondorp *et al.*, 2012) and AtxA exists in a homo-oligomeric state (Hammerstrom *et al.*, 2011). Moreover, both proteins do not seem to recognize specific DNA sequences in their target DNAs (see below).

Simple protein-DNA recognition mechanisms do not exist (Siggers and Gordân, 2014). Based on the structures of various protein-DNA complexes, Rohs *et al*. proposed that particular proteins use likely a combination of readout mechanisms to achieve DNA-binding specificity: base readout and shape readout (Rohs *et al.*, 2010). In the base readout mechanism (or direct readout) the protein recognizes the unique chemical signatures of the DNA bases, whereas in the shape readout mechanism (or indirect readout) the protein recognizes a sequence-dependent DNA shape (DNA structural characteristics) (Rohs *et al.*, 2010).

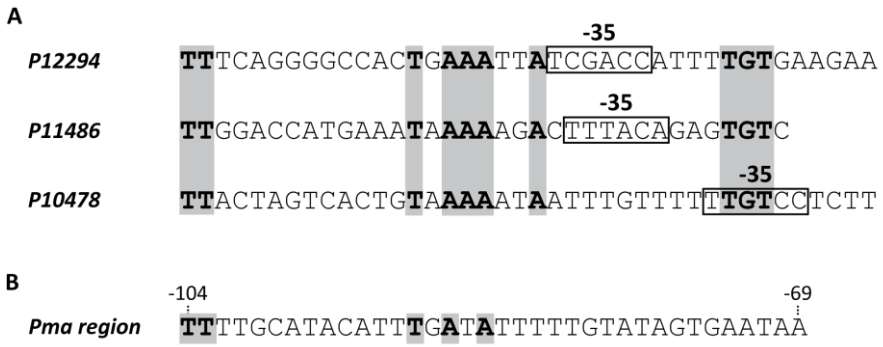


Figure 48. DNA sites recognized by MafR. A. Nucleotide sequence alignment of the DNA sites recognized by MafR on the *P12294*, *P11486* and *P10478* promoter regions. Identical nucleotides are highlighted in grey boxes. **B.** Nucleotide sequence alignment of the DNA site recognized by MafR on the *Pma* promoter region (position -69 to -104). Nucleotides shared with the MafR binding sites shown in (**A**) are highlighted in grey boxes.

Previous DNase I footprinting experiments using a DNA fragment that contains the promoter of the *mafR* gene (*Pma* promoter) showed that MafR binds preferentially to a site located upstream of the promoter (positions -69 to -104). Such a MafR binding site is adjacent to the peak of a potential intrinsic curvature. Nevertheless, the function of this interaction remains unknown (Ruiz-Cruz, 2015). Now, using DNA fragments that contain a particular promoter (*P12294*, *P11486* or *P10478*), we have shown that MafR recognizes a specific DNA site that overlaps the -35 element of the target

promoter. This interaction enhances the efficiency of the promoter and consequently increases the transcription of the corresponding gene. As shown in Figure 48A, the three MafR binding sites identified in this work have a low sequence identity: they share the **T**(G/A)**AAA**(T/A)(T/G)**A** sequence element. Furthermore, according to the bend.it program (Vlahoviček *et al.*, 2003), such sites are located at or flanked by regions of potential bendability (Figures 14B, 18B and 24B). The MafR binding site located upstream of the *Pma* promoter shares a short DNA sequence motif (**T**G**A**T**A**) with the three MafR binding sites (Figure 48B). Therefore, MafR does not seem to recognize a specific nucleotide sequence. Although potential intrinsic curvatures have been identified in the promoter regions of several MafR-regulated operons (Ruiz-Cruz *et al.*, 2016), further work is required to determine whether MafR recognizes a specific site on such promoter regions.

A common feature of the regulators that belong to the Mga/AtxA family seems to be the recognition of particular DNA structures rather than a specific nucleotide sequence, as it is the case of many bacterial nucleoid-associated proteins that influence transcription in a positive or negative manner (Dillon and Dorman, 2010; Fang and Rimsky, 2008). By hydroxyl radical footprinting experiments, Mga*Spn* was shown to bind to two sites of the *S. pneumoniae* R6 chromosome, one of them is located upstream of the *P1623B* promoter (*PB* activation region; positions -60 to -99) and the other one overlaps the *Pmga* promoter (positions -23 to +21) (Solano-Collado *et al.*, 2013). The former interaction enhances the efficiency of the promoter (Solano-Collado *et al.*, 2012), whereas the function of the latter remains unknown. Such Mga*Spn* binding sites have a low sequence identity (see Figure 53) and, according to predictions, they contain an intrinsic curvature flanked by regions of bendability (Solano-Collado *et al.*, 2013). Additional findings support that Mga*Spn* recognizes structural features in its target DNAs: (i) EMSA experiments using curved and non-curved DNAs showed that Mga*Spn* has a higher affinity for a naturally occurring curved DNA (Solano-Collado *et al.*, 2013), and (ii) Mga*Spn* was shown to have a preference for AT-rich DNA regions (Solano-Collado *et al.*, 2016). Mga*Spn* shares some DNA-binding properties with the *E. coli* H-NS protein (Solano-Collado *et al.*, 2016), which is involved in both organization of the bacterial chromosome and regulation of gene expression (Dillon and Dorman, 2010). Concerning the Mga regulator from *S. pyogenes*, sequence alignments of all established Mga binding sites revealed that they exhibit a low sequence identity (13.4%) (Hause and McIver, 2012), although a consensus Mga binding sequence was initially proposed (McIver *et al.*, 1995). Moreover, a mutational analysis in some target promoters indicated that Mga binds to DNA in a promoter-specific manner (Hause and McIver, 2012). In the case of AtxA from *B. anthracis*, sequence similarities are not apparent in its target promoters, and *in silico* and *in vitro* studies revealed that the promoter regions of several target genes are intrinsically curved (Hadjifrangiskou and Koehler, 2008). Examples of DNA-binding proteins that recognize particular

DNA shapes have been reported (Abe *et al.*, 2015; Rohs *et al.*, 2010), in addition to examples where both DNA base sequence and shape recognition are required for protein binding (Al-Zyoud *et al.*, 2016; Deng *et al.*, 2015; Ding *et al.*, 2015).

MgaP is a transcriptional activator

The genetic variability of *S. pneumoniae* plays a central role in the adaptive responses of this human pathogen. The genome sequences of the pneumococcal strains TIGR4 (serotype 4) and R6 (a derivative of the serotype 2 D39 strain) were published in 2001 (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001). As pointed out by Brückner *et al*., the two genomes are not only different in size (2.16 versus 2 Mb) but differ also by approximately 10% of their genes, many of which are organized in clusters. The TIGR4 genome has 12 gene clusters that are absent from the R6 genome, which has six gene clusters that are not present in TIGR4 (Brückner *et al.*, 2004). One of the R6-specific clusters contains two genes that are divergently transcribed, *spr1403* (*pclA*) and *spr1404* (*mgaP*). This gene cluster has been the focus of our investigation.

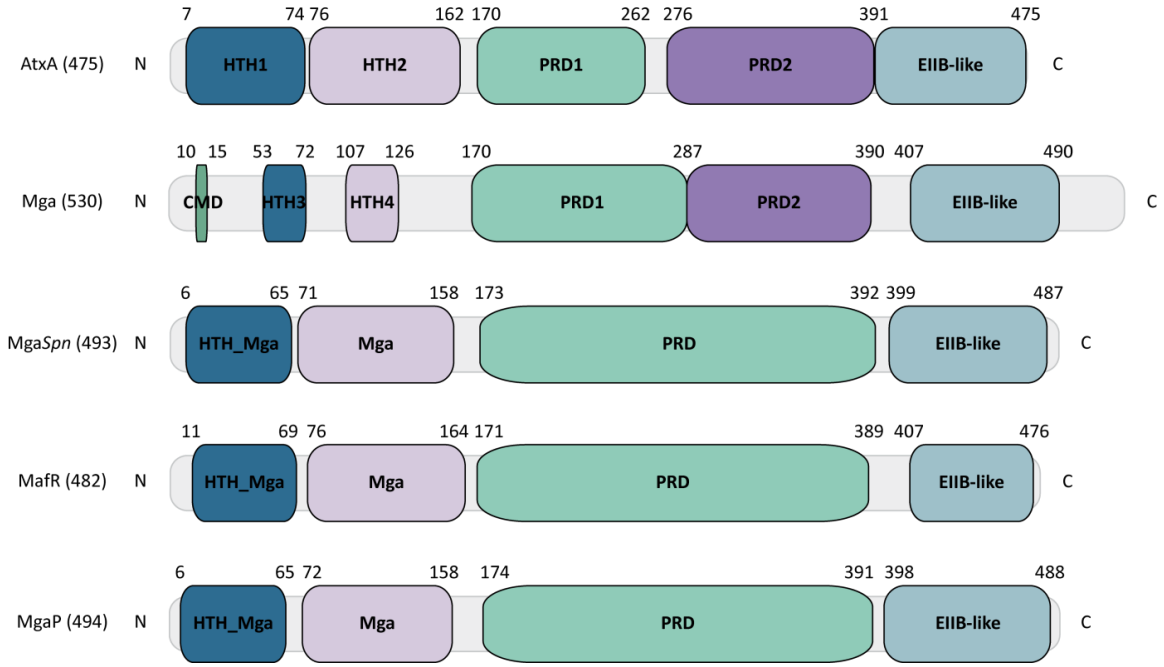


Figure 49. Organization of functional domains in AtxA, Mga, Mga*Spn*, MafR and MgaP. All members of the Mga/AtxA family share a similar domain organization. At the N-terminal region (DNA-binding domain), there are two helix-turn-helix motifs: HTH1 and HTH2 in AtxA; HTH3 and HTH4 in Mga; HTH_Mga and Mga in Mga*Spn*, MafR and MgaP. The central domains are PRDs. Whereas in AtxA and Mga there are two PRDs (PRD1 and PRD2), in Mga*Spn*, MafR and MgaP there is one PRD. At the C-terminal region, all the members contain a PTS EIIB-like domain (Hammerstrom *et al.*, 2015; Hondorp *et al.*, 2013; Ruiz-Cruz, 2015; Solano-Collado, 2014).

Mga (*S. pyogenes*), AtxA (*B. anthracis*), Mga*Spn* (*S. pneumoniae*) and MafR (*E. faecalis*) belong to the Mga/AtxA family of transcriptional regulators. Our present study on the pneumococcal MgaP protein supports that it is an additional member of this family: (i) MgaP shares significant sequence similarity with the proteins of the family (between 36% and 60%), (ii) MgaP is predicted to have a similar organization of functional domains, including two N-terminal helix-turn-helix DNA-binding motifs and a central PTS regulation domain (PRD) (Figure 49), (iii) like Mga*Spn* (Solano-Collado *et al.*, 2013) and MafR (Ruiz-Cruz, 2015 and this work), MgaP is able to generate multimeric complexes on linear dsDNAs, and (iv) MgaP functions as a transcriptional activator.

For most bacterial genes, promoter recognition by RNA polymerase is a key regulatory step in gene expression. In general, all bacteria have various sigma factors, and each of them confers promoter specificity to the RNA polymerase. Most transcription in exponentially growing bacterial cells is initiated by RNA polymerase carrying the housekeeping sigma factor (Browning and Busby, 2016). By qRT-PCR assays, we have shown that the *mgaP* gene of the pneumococcal R6 strain is transcribed under standard laboratory conditions. The transcription levels were higher during the exponential phase of the bacterial growth curve. In addition, we have identified the promoter of the *mgaP* gene using promoter-reporter fusions and primer extension assays. The *PmgaP* promoter has the consensus -10 element (5'-**TATAAT**-3') of the promoters recognized by the housekeeping sigma factor, a feature also present in the promoter of the regulatory *mgaSpn* gene, which was reported to be expressed in R6 cells under standard laboratory conditions (Solano-Collado *et al.*, 2012). We have now determined the relative expression of *mgaSpn* in R6 cells grown under such conditions (qRT-PCR assays). Compared to stationary phase, transcription of *mgaSpn* was found to be higher at logarithmic phase (log_2FC ~2.1). Hence, both genes, *mgaP* and *mgaSpn*, are expressed in exponentially growing R6 cells. Unlike *mgaP, mgaSpn* was reported to be present in 10 referent pneumococcal strains whose genomes have been totally sequenced (Solano-Collado *et al.*, 2012).

Paterson *et al.* reported that deletion of the *spr1404* gene (here *mgaP*) had no effect on the transcription levels of the *pclA* gene (Paterson *et al.*, 2008). However, using a plasmid-harbouring R6 strain that synthesizes high levels of MgaP (chromosome- and plasmid-encoded MgaP), we have demonstrated that MgaP activates the transcription of the *pclA* gene *in vivo*. This activation requires a region located upstream of the *PpclA* promoter (positions -76 to -106). Moreover, such a region is included within the MgaP binding site (positions -68 to -169) defined by DNase I footprinting experiments. Therefore, a combination of *in vivo* and *in vitro* analyses has established, for the first time, the role of MgaP as a transcriptional activator. The *pclA* gene (MgaP target gene) encodes a putative cell wall anchored protein (known as pneumococcal collagen-like protein A), which contains large regions of collagen-like repeats (Paterson *et al.*, 2008). It has been reported that PclA does not

contribute significantly to nasopharyngeal colonization or invasive pneumonia in murine infection models. Nevertheless, a *pclA* deletion mutant strain was shown to be defective in adherence and invasion of host cells *in vitro*, suggesting that PclA could play a significant role in pneumococcal adherence to human cells (Paterson *et al.*, 2008). In the Gram-positive bacterium *S. pyogenes*, the *sclA* and *sclB* genes encode collagen-like proteins and their transcription is regulated by the Mga protein: *sclA* is up-regulated, whereas *sclB* is down-regulated (Nobbs *et al.*, 2009). It has been described that SclA is able to bind cellular fibronectin, laminin and integrins, contributing to GAS adherence and colonization (Caswell *et al.*, 2007; Caswell *et al.*, 2010; Humtsoe *et al.*, 2005).

Analyses of the distribution of the *mgaP* and *pclA* genes among numerous clinical isolates have revealed that these genes are present only in some clinical isolates (about 33-39% of the strains examined depending on the study) (Imai *et al.*, 2011; Paterson *et al.*, 2008). This strain-specific gene cluster is always present in the same location of the pneumococcal genome and is not flanked by insertion sequences or phage elements (Paterson *et al.*, 2008). Furthermore, the presence of the *pclA* gene has been associated with Pneumococcal Molecular Epidemiology Network clones (antimicrobial-resistant clones), suggesting that PclA might contribute to the selection of prevalent clones (Imai *et al.*, 2011).

The opportunistic human pathogen *S. pneumoniae* colonizes the nasopharynx of healthy individuals but can cause serious diseases when invades other host niches (*e.g.* lung or meninges). A study based on RNA-seq has quantified the relative abundance of the transcriptome of the D39V strain under 22 different infection-relevant conditions. The expression data, as well as the co-expression matrix, were published in an interactive data centre named PneumoExpress (https://veeninglab.com/pneumoexpress) (Aprianto et al., 2018). The D39V genome contains the *plcA-mgaP* gene cluster (*SPV_1376* and *SPV_1377* in D39V), as well as the *mgaSpn* gene (*SPV_1587*) and its target operon of unknown function (*SPV_1588-1591*). Searching in such a database, we have found that the highest expression level of *plcA* and *mgaP* corresponds to bacteria grown in nose mimicking conditions (NMC), which simulate colonization. Both genes were also highly expressed in bacteria grown in lung mimicking conditions (LMC), which simulate pneumonia, and in cerebrospinal fluid-mimicking conditions (CSFMC) from 37°C to 40°C, which simulate meningeal fever. In the case of *mgaSpn* and its target operon, the highest expression level corresponds also to bacteria grown in conditions that simulate colonization.

In this Thesis, we have demonstrated that MgaP activates directly the transcription of the *pclA* gene by binding to a specific site upstream of the *PpclA* promoter. The MgaP binding site defined by DNase I footprinting experiments spans positions -68 and -169 (Figure 50). This region contains two peaks of

potential intrinsic curvature (~10-11 degrees per helical turn) (Figure 51). Moreover, the MgaP binding site includes sequences (between -76 and -106) that are required for MgaP-mediated activation of the *PpclA* promoter *in vivo*. These results suggest that MgaP might recruit RNA polymerase to the promoter by contacting the C-terminal domain of the α-subunit, as described for other transcription activators (Browning and Busby, 2016). This activation mechanism often occurs when one or more of the promoter elements has a sequence that is suboptimal for the binding of RNA polymerase (Browning and Busby, 2016). The *PpclA* promoter shows a 4/6 match at both the -35 (5'-**TTGA**TT-3') and -10 (5'-**TA**C**A**T**T**-3') elements. The pneumococcal Mga*Spn* regulator is thought to stimulate transcription of the *spr1623-spr1626* operon by a similar mechanism. In this case, Mga*Spn* enhances the efficiency of the *P1623B* promoter by binding to a site located between the positions -60 and -99 (the *PB* activation region) (Solano-Collado *et al.*, 2013). The *P1623B* promoter has a consensus -10 element (5'-**TATAAT**-3') but lacks a -35 element. As discussed above, transcriptional activation from specific promoters has also been shown for the MafR regulator (this work) and for the Mga regulator (Hondorp and McIver, 2007). However, in both cases, the binding site of the regulator overlaps the -35 element of the target promoter, suggesting that these regulators might activate transcription by a mechanism that involves direct interactions with the sigma factor.

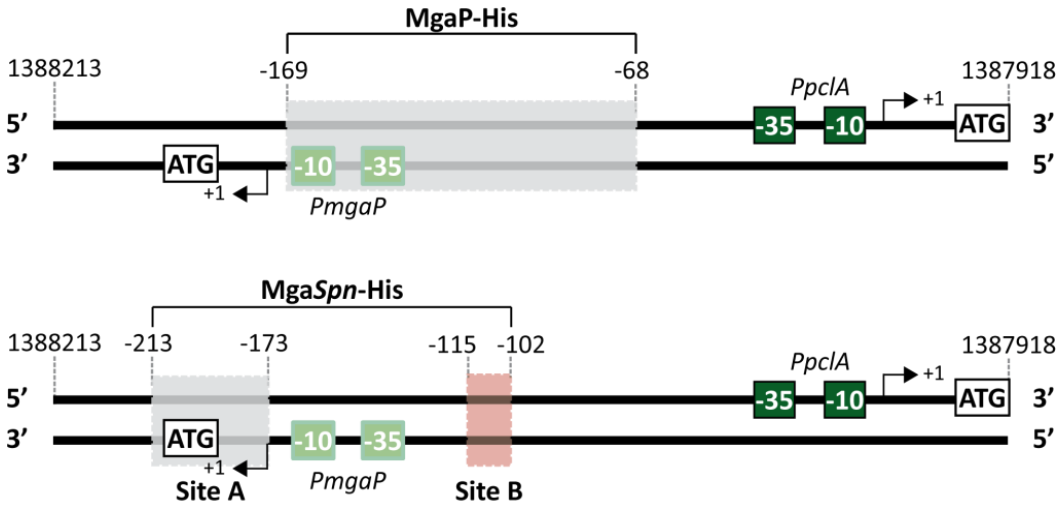


Figure 50. Scheme of the binding sites of MgaP-His and Mga*Spn*-His on the DNA region that contains the *PpclA* promoter. The region spanning coordinates 1387918 to 1388213 of the pneumococcal R6 genome is represented. This region includes the *PpclA* core promoter, the transcription initiation site (+1 position) of the *pclA* gene, the *PmgaP* core promoter and the transcription initiation site (+1 position) of the *mgaP* gene. The shadowed boxes indicate the Mga*Spn*-His and MgaP-His binding sites defined by DNase I footprinting assays (see Figure 43 and Figure 39, respectively).

Since the *pclA* and *mgaP* genes are divergently transcribed, the MgaP binding site located upstream of the *PpclA* promoter overlaps the *PmgaP* promoter (Figure 50). Nevertheless, using promoter-reporter transcriptional fusions (plasmid pASTT-*PmgaP*) and pneumococcal strains that produce different levels of MgaP, we have found that MgaP does not affect the activity of the *PmgaP* promoter, which suggests that MgaP does not regulate the expression of its own gene *in vivo*.

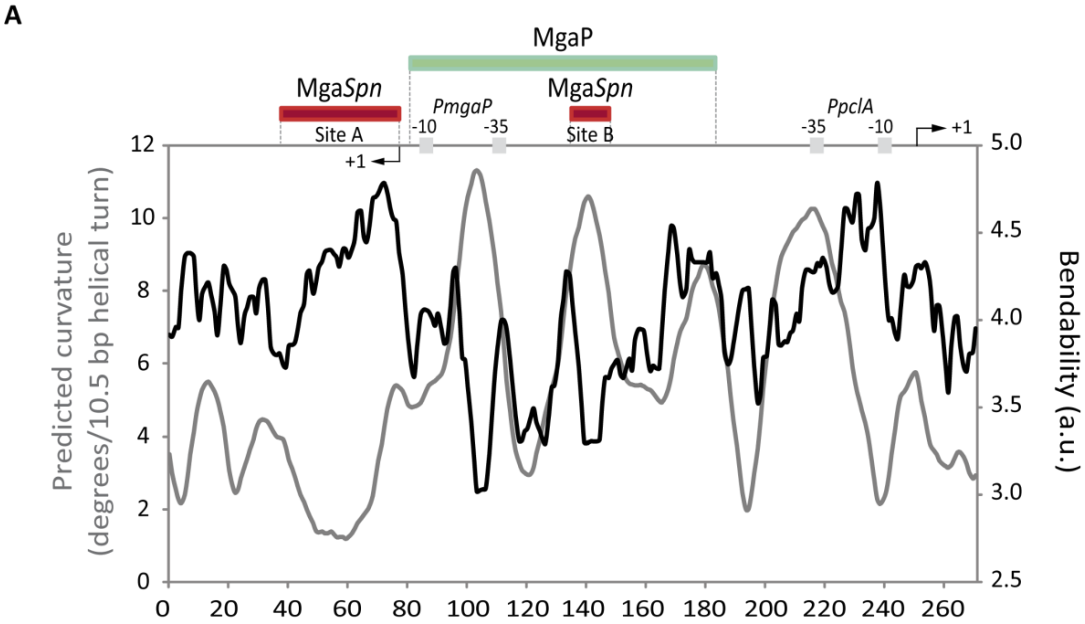


Figure 51. Bendability/curvature propensity plot of the *pclA* promoter region. The region spanning coordinates 1388196 to 1387927 is represented in the plot. The location of the *PpclA* core promoter, the transcription start site of the *pclA* gene, the *PmgaP* core promoter and the transcription start site of the *mgaP* gene are indicated. The MgaP-His (green) and Mga*Spn*-His (red) binding sites are indicated.

MgaP and Mga*Spn* have different DNA-binding specificities

The MgaP and Mga*Spn* transcriptional regulators have sequence similarity (60.3%) and are predicted to have two N-terminal helix-turn-helix DNA-binding motifs (Figure 52). In addition, both regulators bind to linear DNA fragments with low sequence specificity, generate multimeric complexes, and recognize specific DNA sites on the promoter regions of their target genes (Solano-Collado *et al*., 2013 and this work). Nevertheless, despite these similarities, we have demonstrated that these proteins have different DNA-binding specificities. Firstly, MgaP does not recognize the *PB* activation region (positions -60 to -99 of the *P1623B* promoter), which was characterized previously as an Mga*Spn* binding site essential for Mga*Spn*-mediated activation of the *P1623B* promoter *in vivo* (Solano-Collado *et al.*, 2012; Solano-Collado *et al.*, 2013). The inability of MgaP to bind to the *PB* activation region is consistent with its inability to activate the *P1623B* promoter *in vivo*: the

transcription levels of the *spr1623* gene (under the control of *P1623B*) were similar in strains that synthesize different levels of MgaP. Secondly, on DNA fragments that contain the *PpclA* promoter, we have demonstrated that Mga*Spn* does not recognize the MgaP binding site (positions -68 to -169). Mga*Spn* binds preferentially to Site A (positions -173 to -213), which is adjacent to the MgaP binding site (Figure 50). Site A contains regions of potential bendability (Figure 51). At present, we do not know whether the interaction of Mga*Spn* with Site A has a function *in vivo*, since the transcription levels of *pclA* were similar in strains that synthesize (high levels) or not Mga*Spn*. Previous studies showed that Mga*Spn* recognizes the *Pmga* promoter region (positions -23 to +21 of its own promoter), but the function of this interaction remains also unknown (Solano-Collado *et al.*, 2013). As shown in Figure 53, the three Mga*Spn* binding sites identified on the R6 chromosome have a low sequence identity, which supports its preference for particular DNA structures. Additional Mga*Spn* binding sites have been identified on the *hly* operon regulatory region of the *E. coli* plasmid pHly152. Analysis of such binding sites suggested that Mga*Spn*, like the nucleoid-associated protein H-NS, has a preference for AT-rich DNA regions rather than for specific DNA sequences (Solano-Collado *et al.*, 2016). In conclusion, as a member of the Mga/AtxA family of regulators, MgaP likely recognizes structural features in its target DNAs. Nevertheless, verification of this hypothesis requires the identification of additional MgaP biding sites.

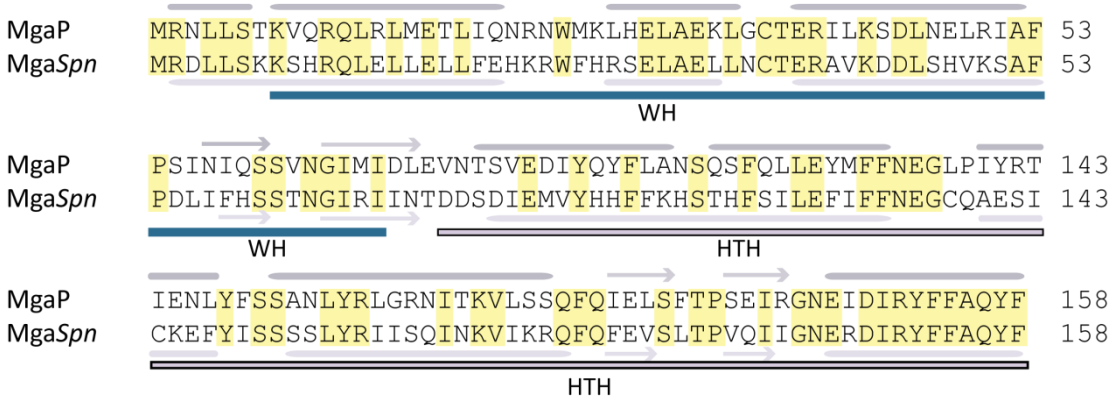


Figure 52. Sequence alignment of the N-terminal region of MgaP and Mga*Spn* using the Clustal Omega program. In Mga*Spn*, the two putative helix-turn-helix motifs, WH (winged helix-turn-helix) and HTH (helix-turn-helix), involved in DNA binding are indicated (Solano-Collado et al., 2012). Identical residues are shown in yellow. For both proteins, predicted secondary structure elements (SABLE program) are indicated (cylinders: α-helices; arrows: β-sheets) (Madeira et al., 2019; Porollo et al., 2004).

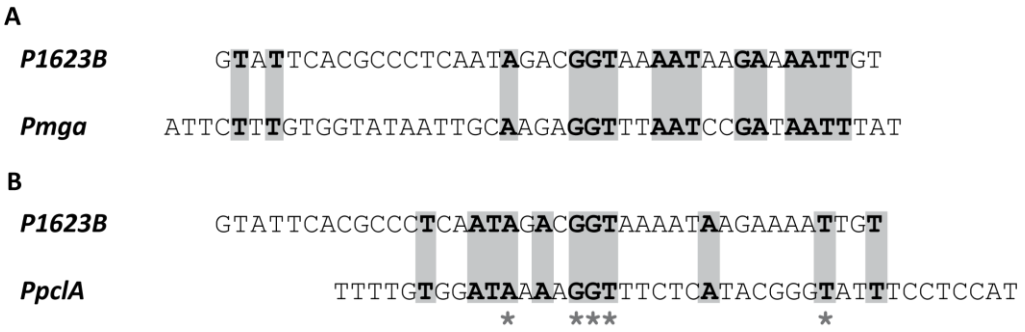


Figure 53. Sites recognized by Mga*Spn*. (A). Nucleotide sequence alignment of the DNA sites recognized by Mga*Spn* on the *P1623B* promoter region (*PB* activation region, positions -99 to -60) and on the *Pmga* promoter region (positions -23 to +21). Both Mga*Spn* binding sites were defined by hydroxyl radical footprinting assays (Solano-Collado *et al.*, 2013). Identical nucleotides are highlighted in grey boxes. **(B).** Nucleotide sequence alignment of the DNA sites recognized by Mga*Spn* on the *P1623B* promoter region (Solano-Collado *et al.*, 2013) and on the *PpclA* promoter region (Site A, positions -213 to -173; defined by DNase I footprinting assay in this work). Asterisks show common nucleotides in the three Mga*Spn* binding sites. By EMSA, the minimum DNA size required for Mga*Spn* binding was shown to be between 20 and 26-bp (Solano-Collado *et al.*, 2013).

CONCLUSIONS

This Thesis has contributed to a better knowledge of two members of the Mga/AtxA family of transcriptional regulators, the MafR protein of *E. faecalis* and the MgaP protein of *S. pneumoniae*. The main conclusions of this work are:

1. Compared to MafR from *E. faecalis* V583, MafR from strain OG1RF has three amino acid changes (Ala37Thr, Gln131Leu, Met145Thr) within the predicted DNA-binding domain. These substitutions do not affect its ability to generate multimeric complexes on linear DNA fragments. Furthermore, removal of the first three amino acid residues of MafR results in a variant that is unable to interact with DNA.

2. The amino acid sequence of MafR is highly conserved among the strains whose genomes have been sequenced.

3. Under laboratory conditions, the expression level of *mafR* is higher in exponentially growing bacterial populations compared to stationary-phase bacteria.

4. MafR is a transcriptional activator. It activates directly the transcription of the *OG1RF_12294*, *OG1RF_11486* and *OG1RF_10478* genes. The predicted function of these genes suggests that MafR could have a regulatory role in calcium homeostasis, queuosine synthesis and cellobiose uptake.

5. MafR enhances the activity of its target promoters by binding to a specific DNA site that overlaps the -35 element. The three MafR binding sites identified in this work exhibit a low sequence identity, which suggests that MafR does not recognize a specific nucleotide sequence.

6. The *mgaP* gene of *S. pneumoniae* R6 is expressed under laboratory conditions. Compared to the stationary phase of bacterial growth, the transcription level of *mgaP* is higher at the exponential phase.

7. The transcription start site of the *mgaP* gene is located 16-17 nucleotides upstream of the translation start codon (*PmgaP* promoter). MgaP does not seem to regulate the activity of its own promoter.

8. MgaP is a transcriptional activator. It stimulates transcription of the *pclA* gene (collagen-like protein) by binding to a specific DNA site upstream of the core promoter (position -68 to -169).

9. The DNA-binding behaviour of MgaP is similar to the one described for Mga*Spn*, a pneumococcal transcriptional activator of the Mga/AtxA family. Both proteins bind to linear DNA fragments in a non-sequence-specific manner and generate multimeric complexes. Also, both proteins recognize a specific DNA site upstream of their target promoters. Despite these similarities, MgaP and Mga*Spn* have different DNA-binding specificities: MgaP does not recognize the Mga*Spn* binding site and vice versa.

CONCLUSIONES

Esta Tesis doctoral ha contribuido a un mejor conocimiento de dos miembros de la familia de reguladores transcripcionales conocida como Mga/AtxA, la proteína MafR de *E. faecalis* y la proteína MgaP de *S. pneumoniae*. Las principales conclusiones de este trabajo son:

1. En comparación con MafR de *E. faecalis* V583, MafR de la estirpe OG1RF presenta tres cambios aminoacídicos (Ala37Thr, Gln131Leu, Met145Thr) en el supuesto dominio de unión a ADN. Estos cambios no afectan su capacidad de generar complejos multiméricos en fragmentos de ADN lineal. Por otra parte, la eliminación de los tres primeros aminoácidos de MafR resulta en una variante incapaz de interaccionar con el ADN.

2. La secuencia aminoacídica de MafR está muy conservada entre las estirpes cuyos genomas han sido secuenciados.

3. En condiciones de laboratorio, el nivel de expresión de *mafR* es mayor en la fase exponencial del crecimiento bacteriano que en la fase estacionaria.

4. MafR es un activador transcripcional. Esta proteína activa directamente la transcripción de los genes *OG1RF_12294*, *OG1RF_11486* y *OG1RF_10478*. La función predicha de estos genes sugiere que MafR podría tener un papel regulador en la homeostasis del calcio, en la síntesis de queuosina y en la incorporación de celobiosa.

5. MafR aumenta la actividad de los promotores de sus genes diana uniéndose a un sitio específico del ADN. Dicho sitio solapa con el elemento -35 del promotor. Los tres sitios de unión de MafR identificados en este trabajo muestran una identidad de secuencia muy baja. Esto sugiere que MafR no reconoce una secuencia nucleotídica específica.

6. El gen *mgaP* de *S. pneumoniae* R6 se expresa en condiciones de laboratorio. El nivel de transcripción de este gen es mayor en la fase exponencial del crecimiento bacteriano que en la fase estacionaria.

7. El inicio de transcripción del gen *mgaP* está a 16-17 nucleótidos del codón de inicio de la traducción (promotor *PmgaP*). MgaP no parece regular la actividad de su propio promotor.

8. MgaP es un activador transcripcional. Esta proteína estimula la transcripción del gen *pclA* (proteína de tipo colágeno) uniéndose a un sitio específico localizado aguas arriba del promotor (posiciones -68 a -169).

9. Las características de unión a ADN que presenta MgaP son similares a las descritas para Mga*Spn*, un activador transcripcional de *S. pneumoniae* que pertenece a la familia Mga/AtxA. Ambas proteínas se unen a fragmentos de ADN lineal sin reconocer una secuencia específica y generan complejos multiméricos. Además, ambas proteínas reconocen un sitio específico localizado aguas arriba de los promotores de sus genes diana. A pesar de estas similitudes, MgaP y Mga*Spn* presentan diferentes especificidades de unión a ADN: MgaP no reconoce el sitio de unión de Mga*Spn* y viceversa.

REFERENCES

1. Aasen, I.M., Moretro, T., Katla, T., Axelsson, L., and Storro, I. (2000). Influence of complex nutrients, temperature and pH on bacteriocin production by *Lactobacillus sakei* CCUG 42687. Applied Microbiology and Biotechnology. 53(2), 159-166.

2. Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, Harmen J., Rohs, R., and Mann, Richard S. (2015). Deconvolving the recognition of DNA shape from sequence. Cell. 161(2), 307-318.

3. Al-Zyoud, W.A., Hynson, R.M.G., Ganuelas, L.A., Coster, A.C.F., Duff, A.P., Baker, M.A.B., Stewart, A.G., Giannoulatou, E., Ho, J.W.K., Gaus, K., Liu, D., Lee, L.K., and Böcking, T. (2016). Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. Nucleic Acids Research. 44(3), 1411-1420.

4. Almengor, A.C., and McIver, K.S. (2004). Transcriptional activation of *sclA* by Mga requires a distal binding site in *Streptococcus pyogenes*. Journal of Bacteriology. 186(23), 7847-7857.

5. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25(17), 3389-3402.

6. Aprianto, R., Slager, J., Holsappel, S., and Veening, J.-W. (2018). High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Research. 46(19), 9990-10006.

7. Arias, C.A., and Murray, B.E. (2012). The rise of the *Enterococcus*: beyond vancomycin resistance. Nature Reviews Microbiology. 10(4), 266-278.

8. Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A.R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D.W., Normark, S., Rappuoli, R., and Henriques-Normark, B. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. Proceedings of the National Academy of Sciences (USA). 103(8), 2857-2862.

9. Berry, A.M., and Paton, J.C. (1996). Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. Infection and Immunity. 64(12), 5255-5262.

10. Bessen, D., Jones, K.F., and Fischetti, V.A. (1989). Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. Journal of Experimental Medicine. 169(1), 269-283.

11. Bidossi, A., Mulas, L., Decorosi, F., Colomba, L., Ricci, S., Pozzi, G., Deutscher, J., Viti, C., and Oggioni, M.R. (2012). A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. PLoS ONE. 7(3), e33320.

12. Bier, N., Hammerstrom, T.G., and Koehler, T.M. (2020). Influence of the phosphoenolpyruvate: carbohydrate phosphotransferase system on toxin gene expression and virulence in *Bacillus anthracis*. Molecular Microbiology. 113(1), 237-252.

13. Blancato, V.S., and Magni, C. (2010). A chimeric vector for efficient chromosomal modification in *Enterococcus faecalis* and other lactic acid bacteria. Letters in Applied Microbiology. 50(5), 542-546.

14. Bogaert, D., de Groot, R., and Hermans, P.W.M. (2004). *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. The Lancet Infectious Diseases. 4(3), 144-154.

15. Borgmann, S., Niklas, D.M., Klare, I., Zabel, L.T., Buchenau, P., Autenrieth, I.B., and Heeg, P. (2004). Two episodes of vancomycin-resistant *Enterococcus faecium* outbreaks caused by two genetically different clones in a newborn intensive care unit. International Journal of Hygiene and Environmental Health. 207(4), 386-389.

16. Bourgogne, A., Garsin, D.A., Qin, X., Singh, K.V., Sillanpaa, J., Yerrapragada, S., Ding, Y., Dugan-Rocha, S., Buhay, C., Shen, H., Chen, G., Williams, G., Muzny, D., Maadani, A., Fox, K.A., Gioia, J., Chen, L., Shang, Y., Arias, C.A., Nallapareddy, S.R., Zhao, M., Prakash, V.P., Chowdhury, S., Jiang, H., Gibbs, R.A., Murray, B.E., Highlander, S.K., and Weinstock, G.M. (2008). Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. Genome Biology. 9(7), 1-16.

17. Brestoff, J.R., and Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. Nature Immunology. 14(7), 676-684.

18. Bridy-Pappas, A.E., Margolis, M.B., Center, K.J., and Isaacman, D.J. (2005). *Streptococcus pneumoniae*: description of the pathogen, disease epidemiology, treatment, and prevention. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 25(9), 1193-1212.

19. Brooks, L.R.K., and Mias, G.I. (2018). *Streptococcus pneumoniae*'s virulence and host immunity: aging, diagnostics, and prevention. Frontiers in Immunology. 9(1366).

20. Brown, J.S., Gilliland, S.M., and Holden, D.W. (2001). A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. Molecular Microbiology. 40(3), 572-585.

21. Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nature Reviews Microbiology. 14(10), 638-650.

22. Brückner, R., Nuhn, M., Reichmann, P., Weber, B., and Hakenbeck, R. (2004). Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. International Journal of Medical Microbiology. 294(2-3), 157-168.

23. Budzik, J.M., and Schneewind, O. (2006). Pili prove pertinent to enterococcal endocarditis. The Journal of Clinical Investigation. 116(10), 2582-2584.

24. Butler, J.C., and Schuchat, A. (1999). Epidemiology of pneumococcal infections in the elderly. Drugs & Aging. 15(1), 11-19.

25. Caswell, C.C., Lukomska, E., Seo, N.-S., Höök, M., and Lukomski, S. (2007). Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the α2β1 integrin enhances pathogen survival and re-emergence. Molecular Microbiology. 64(5), 1319-1331.

26. Caswell, C.C., Oliver-Kozup, H., Han, R., Lukomska, E., and Lukomski, S. (2010). Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. FEMS Microbiology Letters. 303(1), 61-68.

27. Clewell, D.B., Weaver, K.E., Coque, T.M., Francia, M.V., and Hayes, F. (2014). Extrachromosomal and mobile elements in enterococci: transmission, maintenance, and epidemiology. In:

Enterococci: From Commensals to Leading Causes of Drug Resistant Infection, E.-i.-c. Michael S Gilmore, Don B Clewell, Yasuyoshi Ike, and Nathan Shankar ed. (Boston), pp. 225-303.

28. Coburn, P.S., Pillar, C.M., Jett, B.D., Haas, W., and Gilmore, M.S. (2004). *Enterococcus faecalis* senses target cells and in response expresses cytolysin. Science. 306(5705), 2270-2272.

29. Cook, L.C., and Federle, M.J. (2014). Peptide pheromone signaling in *Streptococcus* and *Enterococcus*. FEMS Microbiology Reviews. 38(3), 473-492.

30. Cunningham, M.W. (2000). Pathogenesis of group A streptococcal infections. Clinical Microbiology Reviews. 13(3), 470-511.

31. Chen, F.M., Breiman, R.F., Farley, M., Plikaytis, B., Deaver, K., and Cetron, M.S. (1998). Geocoding and linking data from population-based surveillance and the US census to evaluate the impact of median household income on the epidemiology of invasive *Streptococcus pneumoniae* infections. American Journal of Epidemiology. 148(12), 1212-1218.

32. Chow, J.W., Thal, L.A., Perri, M.B., Vazquez, J.A., Donabedian, S.M., Clewell, D.B., and Zervos, M.J. (1993). Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. Antimicrobial Agents and Chemotherapy. 37(11), 2474-2477.

33. Dai, Z., and Koehler, T.M. (1997). Regulation of anthrax toxin activator gene (*atxA*) expression in *Bacillus anthracis*: temperature, not CO_2/bicarbonate, affects AtxA synthesis. Infection and Immunity. 65(7), 2576-2582.

34. Dale, J.L., Raynor, M.J., Ty, M.C., Hadjifrangiskou, M., and Koehler, T.M. (2018). A dual role for the *Bacillus anthracis* master virulence regulator AtxA: control of sporulation and anthrax toxin production. Frontiers in Microbiology. 9, 482.

35. De Vuyst, L., Callewaert, R., and Crabbé, K. (1996). Primary metabolite kinetics of bacteriocin biosynthesis by *Lactobacillus amylovorus* and evidence for stimulation of bacteriocin production under unfavourable growth conditions. Microbiology. 142(4), 817-827.

36. DebRoy, S., Gao, P., Garsin, D.A., Harvey, B.R., Kos, V., Nes, I.F., and Solheim, M. (2014). Transcriptional and post transcriptional control of enterococcal gene regulation. In: Enterococci: From Commensals to Leading Causes of Drug Resistant Infection, E.-i.-c. Michael S Gilmore, Don B Clewell, Yasuyoshi Ike, and Nathan Shankar ed. (Boston), pp. 337-365.

37. Deng, Z., Wang, Q., Liu, Z., Zhang, M., Machado, A.C.D., Chiu, T.-P., Feng, C., Zhang, Q., Yu, L., Qi, L., Zheng, J., Wang, X., Huo, X., Qi, X., Li, X., Wu, W., Rohs, R., Li, Y., and Chen, Z. (2015). Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. Nature Communications. 6(1), 7642.

38. Dersch, P., Khan, M.A., Mühlen, S., and Görke, B. (2017). Roles of regulatory RNAs for antibiotic resistance in bacteria and their potential value as novel drug targets. Frontiers in Microbiology. 8, 803-803.

39. Desai, S.K., and Kenney, L.J. (2017). To ~P or Not to ~P? Non-canonical activation by two-component response regulators. Molecular Microbiology. 103(2), 203-213.

40. Deutscher, J., Ake, F.M., Derkaoui, M., Zebre, A.C., Cao, T.N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., and Joyet, P. (2014). The bacterial phosphoenolpyruvate: carbohydrate

phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. Microbiology and Molecular Biology Reviews. 78(2), 231-256.

41. Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nature Reviews Microbiology. 8(3), 185-195.

42. Ding, P., McFarland, K., Jin, S., Tong, G., Duan, B., Yang, A., Hughes, T., Liu, W., Dove, S., Navarre, W., and Xia, B. (2015). A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. PLOS Pathogens. 11, e1004967.

43. Dintilhac, A., Alloing, G., Granadel, C., and Claverys, J.-P. (1997). Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. Molecular Microbiology. 25(4), 727-739.

44. Domínguez, D.C., Guragain, M., and Patrauchan, M. (2015). Calcium binding proteins and calcium signaling in prokaryotes. Cell Calcium. 57(3), 151-165.

45. Dower, W., Miller, J., and Ragsdale, C. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Research. 16(13), 6127-6145.

46. Dubourg, G., Edouard, S., and Raoult, D. (2019). Relationship between nasopharyngeal microbiota and patient's susceptibility to viral infection. Expert Review of Anti-infective Therapy. 17(6), 437-447.

47. Dunny, G.M., Brown, B.L., and Clewell, D.B. (1978). Induced cell aggregation and mating in *Streptococcus faecalis*: evidence for a bacterial sex pheromone. Proceedings of the National Academy of Sciences (USA). 75(7), 3479-3483.

48. Dunny, G.M., and Clewell, D.B. (1975). Transmissible toxin (hemolysin) plasmid in *Streptococcus faecalis* and its mobilization of a noninfectious drug resistance plasmid. Journal of Bacteriology. 124(2), 784-790.

49. Durand, J.M.B., Dagberg, B., Uhlin, B.E., and Björk, G.R. (2000). Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. Molecular Microbiology. 35(4), 924-935.

50. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C E., and Finn, R.D. (2018). The Pfam protein families database in 2019. Nucleic Acids Research. 47(1), 427-432.

51. Fang, F.C., and Rimsky, S. (2008). New insights into transcriptional regulation by H-NS. Current Opinion in Microbiology. 11(2), 113-120.

52. Fine, D.P. (1975). Pneumococcal type-associated variability in alternate complement pathway activation. Infection and Immunity. 12(4), 772-778.

53. Fisher, K., and Phillips, C. (2009). The ecology, epidemiology and virulence of *Enterococcus*. Microbiology. 155(6), 1749-1757.

54. Forteza, R., Lieb, T., Aoki, T., Savani, R.C., Conner, G.E., and Salathe, M. (2001). Hyaluronan serves a novel role in airway mucosal host defense. The FASEB Journal. 15(12), 2179-2186.

55. Fouet, A. (2010). AtxA, a *Bacillus anthracis* global virulence regulator. Research in Microbiology. 161(9), 735-742.

56. Franz, C.M.A.P., Huch, M., Abriouel, H., Holzapfel, W., and Gálvez, A. (2011). Enterococci as probiotics and their implications in food safety. International Journal of Food Microbiology. 151(2), 125-140.

57. Gaca, A.O., and Lemos, J.A. (2019). Adaptation to adversity: the intermingling of stress tolerance and pathogenesis in enterococci. Microbiology and Molecular Biology Reviews. 83(3), e00008-00019.

58. Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Research. 5(9), 3157-3170.

59. Galinier, A., and Deutscher, J. (2017). Sophisticated regulation of transcriptional factors by the bacterial phosphoenolpyruvate: sugar phosphotransferase system. Journal of Molecular Biology. 429(6), 773-789.

60. Garsin, D.A., Sifri, C.D., Mylonakis, E., Qin, X., Singh, K.V., Murray, B.E., Calderwood, S.B., and Ausubel, F.M. (2001). A simple model host for identifying Gram-positive virulence factors. Proceedings of the National Academy of Sciences (USA). 98(19), 10892-10897.

61. Geno, K.A., Gilbert, G.L., Song, J.Y., Skovsted, I.C., Klugman, K.P., Jones, C., Konradsen, H.B., and Nahm, M.H. (2015). Pneumococcal capsules and their types: past, present, and future. Clinical Microbiology Reviews. 28(3), 871-899.

62. Gomez-Mejia, A., Gamez, G., and Hammerschmidt, S. (2018). *Streptococcus pneumoniae* two-component regulatory systems: The interplay of the pneumococcus with its environment. International Journal of Medical Microbiology. 308(6), 722-737.

63. Goodsell, D.S., and Dickerson, R.E. (1994). Bending and curvature calculations in B-DNA. Nucleic Acids Research. 22(24), 5497-5503.

64. Hadjifrangiskou, M., and Koehler, T.M. (2008). Intrinsic curvature associated with the coordinately regulated anthrax toxin gene promoters. Microbiology. 154(8), 2501-2512.

65. Hammerstrom, T.G., Horton, L.B., Swick, M.C., Joachimiak, A., Osipiuk, J., and Koehler, T.M. (2015). Crystal structure of *Bacillus anthracis* virulence regulator AtxA and effects of phosphorylated histidines on multimerization and activity. Molecular Microbiology. 95(3), 426-441.

66. Hammerstrom, T.G., Roh, J.H., Nikonowicz, E.P., and Koehler, T.M. (2011). *Bacillus anthracis* virulence regulator AtxA: oligomeric state, function and CO_2-signalling. Molecular Microbiology. 82(3), 634-647.

67. Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology. 166(4), 557-580.

68. Hancock, L., and Perego, M. (2002). Two-component signal transduction in *Enterococcus faecalis*. Journal of Bacteriology. 184(21), 5819-5825.

69. Hancock, L.E., and Gilmore, M.S. (2002). The capsular polysaccharide of *Enterococcus faecalis* and its relationship to other polysaccharides in the cell wall. Proceedings of the National Academy of Sciences (USA). 99(3), 1574-1579.

70. Hancock, L.E., and Perego, M. (2004). Systematic inactivation and phenotypic characterization of two-component signal transduction systems of *Enterococcus faecalis* V583. Journal of Bacteriology. 186(23), 7951-7958.

71. Hanchi, H., Mottawea, W., Sebei, K., and Hammami, R. (2018). The genus *Enterococcus*: between probiotic potential and safety concerns — An update. Frontiers in Microbiology. 9(1791).

72. Hause, L.L., and McIver, K.S. (2012). Nucleotides critical for the interaction of the *Streptococcus pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. Journal of Bacteriology. 194(18), 4904-4919.

73. Hava, D.L., and Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. Molecular Microbiology. 45(5), 1389-1406.

74. Heikens, E., Bonten, M.J.M., and Willems, R.J.L. (2007). Enterococcal surface protein Esp is important for biofilm formation of *Enterococcus faecium* E1162. Journal of Bacteriology. 189(22), 8233-8240.

75. Hemsley, C., Joyce, E., Hava, D.L., Kawale, A., and Camilli, A. (2003). MgrA, an orthologue of Mga, acts as a transcriptional repressor of the genes within the *rlrA* pathogenicity islet in *Streptococcus pneumoniae*. Journal of Bacteriology. 185(22), 6640-6647.

76. Hofmann, J., Cetron, M.S., Farley, M.M., Baughman, W.S., Facklam, R.R., Elliott, J.A., Deaver, K.A., and Breiman, R.F. (1995). The prevalence of drug-resistant *Streptococcus pneumoniae* in Atlanta. The New England Journal of Medicine. 333(8), 481-486.

77. Hollenbeck, B.L., and Rice, L.B. (2012). Intrinsic and acquired resistance mechanisms in *Enterococcus*. Virulence. 3(5), 421-569.

78. Hondorp, E.R., Hou, S.C., Hause, L.L., Gera, K., Lee, C.-E., and McIver, K.S. (2013). PTS phosphorylation of Mga modulates regulon expression and virulence in the group A *Streptococcus*. Molecular Microbiology. 88(6), 1176-1193.

79. Hondorp, E.R., Hou, S.C., Hempstead, A.D., Hause, L.L., Beckett, D.M., and McIver, K.S. (2012). Characterization of the group A *Streptococcus* Mga virulence regulator reveals a role for the C-terminal region in oligomerization and transcriptional activation. Molecular Microbiology. 83(5), 953-967.

80. Hondorp, E.R., and McIver, K.S. (2007). The Mga virulence regulon: infection where the grass is greener. Molecular Microbiology. 66(5), 1056-1065.

81. Honsa, E.S., Johnson, M.D.L., and Rosch, J.W. (2013). The roles of transition metals in the physiology and pathogenesis of *Streptococcus pneumoniae*. Frontiers in Cellular and Infection Microbiology. 3(92).

82. Horinouchi, S., and Weisblum, B. (1982). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. Journal of Bacteriology. 150(2), 815-825.

83. Hoskins, J., Alborn, W.E., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R., Skatrud, P.L., and Glass, J.I. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. Journal of Bacteriology. 183(19), 5709-5717.

84. Hostetter, M.K. (1986). Serotypic variations among virulent pneumococci in deposition and degradation of covalently bound C3b: implications for phagocytosis and antibody production. The Journal of Infectious Diseases. 153(4), 682-693.

85. Humtsoe, J.O., Kim, J.K., Xu, Y., Keene, D.R., Höök, M., Lukomski, S., and Wary, K.K. (2005). A streptococcal collagen-like protein interacts with the $\alpha 2\beta 1$ integrin and induces intracellular signaling. Journal of Biological Chemistry. 280(14), 13848-13857.

86. Hutinet, G., Swarjo, M.A., and de Crécy-Lagard, V. (2017). Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. RNA Biology. 14(9), 1175-1184.

87. Hyams, C., Camberlein, E., Cohen, J.M., Bax, K., and Brown, J.S. (2010). The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infection and Immunity. 78(2), 704-715.

88. Ike, Y., Hashimoto, H., and Clewell, D.B. (1984). Hemolysin of *Streptococcus faecalis* subspecies zymogenes contributes to virulence in mice. Infection and Immunity. 45(2), 528-530.

89. Imai, S., Ito, Y., Ishida, T., Hirai, T., Ito, I., Yoshimura, K., Maekawa, K., Takakura, S., Niimi, A., Iinuma, Y., Ichiyama, S., and Mishima, M. (2011). Distribution and clonal relationship of cell surface virulence genes among *Streptococcus pneumoniae* isolates in Japan. Clinical Microbiology and Infection. 17(9), 1409-1414.

90. Jacob, A.E., Douglas, G.J., and Hobbs, S.J. (1975). Self-transferable plasmids determining the hemolysin and bacteriocin of *Streptococcus faecalis* var. zymogenes. Journal of Bacteriology. 121(3), 863-872.

91. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research. 36 (Web server issue), W5–W9.

92. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 10(6), 845-858.

93. Khosravi, A., and Mazmanian, S.K. (2013). Disruption of the gut microbiome as a risk factor for microbial infections. Current Opinion in Microbiology. 16(2), 221-227.

94. Kim, J.O., Romero-Steiner, S., Sørensen, U.B.S., Blom, J., Carvalho, M., Barnard, S., Carlone, G., and Weiser, J.N. (1999). Relationship between cell surface carbohydrates and intrastrain variation on opsonophagocytosis of *Streptococcus pneumoniae*. Infection and Immunity. 67(5), 2327-2333.

95. Kim, J.O., and Weiser, J.N. (1998). Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. The Journal of Infectious Diseases. 177(2), 368-377.

96. Kreft, B., Marre, R., Schramm, U., and Wirth, R. (1992). Aggregation substance of *Enterococcus faecalis* mediates adhesion to cultured renal tubular cells. Infection and Immunity. 60(1), 25-30.

97. Kreikemeyer, B., McIver, K.S., and Podbielski, A. (2003). Virulence factor regulation and regulatory networks in *Streptococcus pyogene*s and their impact on pathogen-host interactions. Trends in Microbiology. 11(5), 224-232.

98. Lacks, S. (1966). Integration efficiency and genetic recombination in pneumococcal transformation. Genetics. 53(1), 207-235.

99. Lacks, S. (1968). Genetic regulation of maltosaccharide utilization in *pneumococcus*. Genetics. 60(4), 685-706.

100. Lanie, J.A., Ng, W.L., Kazmierczak, K.M., Andrzejewski, T.M., Davidsen, T.M., Wayne, K.J., Tettelin, H., Glass, J.I., and Winkler, M.E. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. Journal of Bacteriology. 189(1), 38-51.

101. LeBlanc, D.J., Chen, Y.-Y.M., and Lee, L.N. (1993). Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1. Plasmid. 30(3), 296-302.

102. Lebreton, F., Willems, R.J.L., and Gilmore, M.S. (2014). *Enterococcus* diversity, origins in nature, and gut colonization. In: *Enterococci*: From Commensals to Leading Causes of Drug Resistant Infection, E.-i.-c. Michael S Gilmore, Don B Clewell, Yasuyoshi Ike, and Nathan Shankar ed. (Boston), pp. 3-46.

103. Lindenstrauß, A.G., Ehrmann, M.A., Behr, J., Landstorfer, R., Haller, D., Sartor, R.B., and Vogel, R.F. (2014). Transcriptome analysis of *Enterococcus faecalis* toward its adaption to surviving in the mouse intestinal tract. Archives of Microbiology. 196(6), 423-433.

104. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 25(4), 402-408.

105. Loughran, A.J., Orihuela, C.J., and Tuomanen, E.I. (2019). *Streptococcus pneumoniae*: invasion and inflammation. Microbiology Spectrum. 7(2).

106. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., and Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research. 48(1), 265-268.

107. Madeira, F., Park, Y.m., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 47 (Web server issue 1), W636-W641.

108. Majsnerowska, M., ter Beek, J., Stanek, W.K., Duurkens, R.H., and Slotboom, D.J. (2015). Competition between different S-components for the shared energy coupling factor module in energy coupling factor transporters. Biochemistry. 54(31), 4763-4766.

109. Marks, L.R., Reddinger, R.M., and Hakansson, A.P. (2012). High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. mBio. 3(5), e00200-00212.

110. Marks, L.R., Reddinger, R.M., and Hakansson, A.P. (2014). Biofilm formation enhances fomite survival of *Streptococcus pneumoniae* and *Streptococcus pyogenes*. Infection and Immunity. 82(3), 1141-1146.

111. Marra, A., Lawson, S., Asundi, J.S., Brigham, D., and Hromockyj, A.E. (2002). *In vivo* characterization of the *psa* genes from *Streptococcus pneumoniae* in multiple models of infection. Microbiology. 148(5), 1483-1491.

112. Martin, J.D., and Mundt, J.O. (1972). Enterococci in insects. Journal of Applied Microbiology. 24(4), 575-580.

113. McCall, R.M., Sievers, M.E., Fattah, R., Ghirlando, R., Pomerantsev, A.P., and Leppla, S.H. (2019). *Bacillus anthracis* virulence regulator AtxA binds specifically to the *pagA* promoter region. Journal of Bacteriology. 201(23), e00569-00519.

114. McGee, L., McDougal, L., Zhou, J., Spratt, B.G., Tenover, F.C., George, R., Hakenbeck, R., Hryniewicz, W., Lefévre, J.C., Tomasz, A., and Klugman, K.P. (2001). Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. Journal of Clinical Microbiology. 39(7), 2565-2571.

115. McIver, K.S. (2009). Stand-alone response regulators controlling global virulence networks in *Streptococcus pyogenes*. Contributions to Microbiology. 16, 103-119.

116. McIver, K.S., Heath, A.S., Green, B.D., and Scott, J.R. (1995). Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A *Streptococcus*. Journal of Bacteriology. 177(22), 6619-6624.

117. McIver, K.S., and Myles, R.L. (2002). Two DNA-binding domains of Mga are required for virulence gene activation in the group A *Streptococcus*. Molecular Microbiology. 43(6), 1591-1601.

118. McIver, K.S., and Scott, J.R. (1997). Role of *mga* in growth phase regulation of virulence genes of the group A *Streptococcus*. Journal of Bacteriology. 179(16), 5178-5187.

119. McIver, K.S., Thurman, A.S., and Scott, J.R. (1999). Regulation of *mga* transcription in the group A *Streptococcus*: specific binding of Mga within its own promoter and evidence for a negative regulator. Journal of Bacteriology. 181(17), 5373-5383.

120. Mohamed, J.A., Huang, W., Nallapareddy, S.R., Teng, F., and Murray, B.E. (2004). Influence of origin of isolates, especially endocarditis isolates, and various genes on biofilm formation by *Enterococcus faecalis*. Infection and Immunity. 72(6), 3658-3663.

121. Moreno-Blanco, A. (2016). Purificación de dos variantes de la proteína MafR de *Enterococcus faecalis*. Supervised by S. Ruiz-Cruz and A. Bravo. Centro de Investigaciones Biológicas, CSIC. Máster Genética y Biología Celular, Universidad Autónoma de Madrid, Universidad de Alcalá de Henares y Universidad Complutense de Madrid.

122. Muller, C., Cacaci, M., Sauvageot, N., Sanguinetti, M., Rattei, T., Eder, T., Giard, J.-C., Kalinowski, J., Hain, T., and Hartke, A. (2015). The intraperitoneal transcriptome of the opportunistic pathogen *Enterococcus faecalis* in mice. PLoS ONE. 10(5), e0126143.

123. Mundt, J.O. (1961). Occurrence of enterococci: bud, blossom, and soil studies. Journal of Applied Microbiology. 9(6), 541-544.

124. Mundt, J.O. (1963). Occurrence of enterococci in animals in a wild environment. Journal of Applied Microbiology. 11(2), 136-140.

125. Murray, B.E. (1990). The life and times of the *Enterococcus*. Clinical Microbiology Reviews. 3(1), 46-65.

126. Mylonakis, E., Engelbert, M., Qin, X., Sifri, C.D., Murray, B.E., Ausubel, F.M., Gilmore, M.S., and Calderwood, S.B. (2002). The *Enterococcus faecalis fsrB* gene, a key component of the *fsr* quorum-sensing system, is associated with virulence in the rabbit endophthalmitis model. Infection and Immunity. 70(8), 4678-4681.

127. Nakayama, J., Cao, Y., Horii, T., Sakuda, S., Akkermans, A.D., de Vos, W.M., and Nagasawa, H. (2001). Gelatinase biosynthesis-activating pheromone: a peptide lactone that mediates a quorum sensing in *Enterococcus faecalis*. Molecular Microbiology. 41(1), 145-154.

128. Nakayama, J., Chen, S., Oyama, N., Nishiguchi, K., Azab, E.A., Tanaka, E., Kariyama, R., and Sonomoto, K. (2006). Revised model for *Enterococcus faecalis fsr* quorum-sensing system: the small open reading frame *fsrD* encodes the gelatinase biosynthesis-activating pheromone propeptide corresponding to staphylococcal AgrD. Journal of Bacteriology. 188(23), 8321-8326.

129. Nallapareddy, S.R., Singh, K.V., Sillanpaa, J., Garsin, D.A., Hook, M., Erlandsen, S.L., and Murray, B.E. (2006). Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. Journal of Clinical Investigation. 116(10), 2799-2807.

130. Nobbs, A.H., Lamont, R.J., and Jenkinson, H.F. (2009). *Streptococcus* adherence and colonization. Microbiology and Molecular Biology Reviews. 73(3), 407-450.

131. Noguchi, S., Nishimura, Y., Hirota, Y., and Nishimura, S. (1982). Isolation and characterization of an *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. Journal of Biological Chemistry. 257(11), 6544-6550.

132. Okada, N., Geist, R.T., and Caparon, M.G. (1993). Positive transcriptional control of *mry* regulates virulence in the group A *Streptococcus*. Molecular Microbiology. 7(6), 893-903.

133. Olmsted, S.B., Dunny, G.M., Erlandsen, S.L., and Wells, C.L. (1994). A plasmid-encoded surface protein on *Enterococcus faecalis* augments its internalization by cultured intestinal epithelial cells. The Journal of Infectious Diseases. 170(6), 1549-1556.

134. Orihuela, C.J., Gao, G., Francis, K.P., Yu, J., and Tuomanen, E.I. (2004). Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. The Journal of Infectious Diseases. 190(9), 1661-1669.

135. Ortuno-Camuñas, A. (2017). Clonaje del gen *mgaP* de *Streptococcus pneumoniae* en vectores de expresión para análisis funcionales *in vivo* e *in vitro*. Supervised by S. Ruiz-Cruz and A. Bravo. Centro de Investigaciones Biológicas, CSIC. Máster en Bioquímica, Biología Molecular y Biomedicina, Universidad Complutense de Madrid.

136. Osipiuk, J., Wu, R., Jedrzejczak, R., Moy, S., and Joachimiak, A. (unpublished results). Putative Mga family transcriptional regulator from *Enterococcus faecalis*. The three-dimensional structure of EF3013; PDB 3SQN.

137. Palmgren, M.G., and Nissen, P. (2011). P-type ATPases. Annual Review of Biophysics. 40, 243-266.

138. Parente, E., and Ricciardi, A. (1994). Influence of pH on the production of enterocin 1146 during batch fermentation. Letters in Applied Microbiology. 19(1), 12-15.

139. Park, S.Y., Shin, Y.P., Kim, C.H., Park, H.J., Seong, Y.S., Kim, B.S., Seo, S.J., and Lee, I.H. (2008). Immune evasion of *Enterococcus faecalis* by an extracellular gelatinase that cleaves C3 and iC3b. The Journal of Immunology. 181(9), 6328-6336.

140. Pastor, P., Medley, F., and Murphy, T.V. (1998). Invasive pneumococcal disease in Dallas County, Texas: results from population-based surveillance in 1995. Clinical Infectious Diseases. 26(3), 590-595.

141. Paterson, G.K., and Mitchell, T.J. (2006). The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. Microbes and Infection. 8(1), 145-153.

142. Paterson, G.K., Nieminen, L., Jefferies, J.M., and Mitchell, T.J. (2008). PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. FEMS Microbiology Letters. 285(2), 170-176.

143. Paulsen, I.T., Banerjei, L., Myers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., and Fraser, C.M. (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science. 299(5615), 2071-2074.

144. Perez-Casal, J., Okada, N., Caparon, M.G., and Scott, J.R. (1995). Role of the conserved C-repeat region of the M protein of *Streptococcus pyogenes*. Molecular Microbiology. 15(5), 907-916.

145. Pinkston, K.L., Gao, P., Diaz-Garcia, D., Sillanpää, J., Nallapareddy, S.R., Murray, B.E., and Harvey, B.R. (2011). The Fsr quorum-sensing system of *Enterococcus faecalis* modulates surface display of the collagen-binding MSCRAMM Ace through regulation of *gelE*. Journal of Bacteriology. 193(17), 4317-4325.

146. Porollo, A.A., Adamczak, R., and Meller, J. (2004). POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. Bioinformatics. 20(15), 2460-2462.

147. Qin, X., Singh, K.V., Weinstock, G.M., and Murray, B.E. (2000). Effects of *Enterococcus faecalis fsr* genes on production of gelatinase and a serine protease and virulence. Infection and Immunity. 68(5), 2579-2586.

148. Qin, X., Singh, K.V., Weinstock, G.M., and Murray, B.E. (2001). Characterization of *fsr*, a regulator controlling expression of gelatinase and serine protease in *Enterococcus faecalis* OG1RF. Journal of Bacteriology. 183(11), 3372-3382.

149. Ramsey, M., Hartke, A., and Huycke, M. (2014). The physiology and metabolism of enterococci. In: Enterococci: From Commensals to Leading Causes of Drug Resistant Infection E.-i.-c. Michael S Gilmore, Don B Clewell, Yasuyoshi Ike, and Nathan Shankar ed. (Boston), pp. 424-465.

150. Raynor, M.J., Roh, J.H., Widen, S.G., Wood, T.G., and Koehler, T.M. (2018). Regulons and protein-protein interactions of PRD-containing *Bacillus anthracis* virulence regulators reveal overlapping but distinct functions. Molecular Microbiology. 109(1), 1–22.

151. Redd, S.C., Rutherford, G.W., III, Sande, M.A., Lifson, A.R., Hadley, W.K., Facklam, R.R., and Spika, J.S. (1990). The role of human immunodeficiency virus infection in pneumococcal bacteremia in San Francisco residents. The Journal of Infectious Diseases. 162(5), 1012-1017.

152. Reid, S.D., Green, N.M., Buss, J.K., Lei, B., and Musser, J.M. (2001). Multilocus analysis of extracellular putative virulence proteins made by group A *Streptococcus*: Population genetics, human serologic response, and gene transcription. Proceedings of the National Academy of Sciences (USA). 98(13), 7552-7557.

153. Ribardo, D.A., and McIver, K.S. (2006). Defining the Mga regulon: Comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A *Streptococcus*. Molecular Microbiology. 62(2), 491-508.

154. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 16(6), 276-277.

155. Rich, R.L., Kreikemeyer, B., Owens, R.T., LaBrenz, S., Narayana, S.V., Weinstock, G.M., Murray, B.E., and Hook, M. (1999). Ace is a collagen-binding MSCRAMM from *Enterococcus faecalis*. Journal of Biological Chemistry. 274(38), 26939-26945.

156. Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, A.G., Gasbarrini, A., and Mele, C.M. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms. 7(1), 14.

157. Rodionov, D.A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I.A., Erkens, G.B., Slotboom, D.J., Gelfand, M.S., Osterman, A.L., Hanson, A.D., and Eitinger, T. (2009). A novel class of modular transporters for vitamins in prokaryotes. Journal of Bacteriology. 191(1), 42-51.

158. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annual Review of Biochemistry. 79(1), 233-269.

159. Rosch, J.W., Gao, G., Ridout, G., Wang, Y.-D., and Tuomanen, E.I. (2009). Role of the manganese efflux system *mntE* for signalling and pathogenesis in *Streptococcus pneumoniae*. Molecular Microbiology. 72(1), 12-25.

160. Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. Molecular Microbiology. 25(5), 819-829.

161. Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology. 132, 365-386.

162. Ruiz-Cruz, S. (2015). Caracterización molecular de un regulador transcripcional del tipo Mga/AtxA en *Enterococcus faecalis*. Supervised by A. Bravo. Centro de Investigaciones Biológicas, CSIC. Tesis Doctoral, Departamento de Bioquímica y Biología Molecular I, Universidad Complutense de Madrid.

163. Ruiz-Cruz, S., Espinosa, M., Goldmann, O., and Bravo, A. (2016). Global regulation of gene expression by the MafR protein of *Enterococcus faecalis*. Frontiers in Microbiology. 6(1521), 1-13.

164. Ruiz-Cruz, S., Moreno-Blanco, A., Espinosa, M., and Bravo, A. (2018). DNA-binding properties of MafR, a global regulator of *Enterococcus faecalis*. FEBS Letters. 592(8), 1412-1425.

165. Ruiz-Cruz, S., Moreno-Blanco, A., Espinosa, M., and Bravo, A. (2019). Transcriptional activation by MafR, a global regulator of *Enterococcus faecalis*. Scientific Reports. 9(6146), 1-12.

166. Ruiz-Cruz, S., Solano-Collado, V., Espinosa, M., and Bravo, A. (2010). Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. Journal of Microbiological Methods. 83(2), 156-163.

167. Ruiz-Garbajosa, P., Bonten, M.J., Robinson, D.A., Top, J., Nallapareddy, S.R., Torres, C., Coque, T.M., Canton, R., Baquero, F., Murray, B.E., del Campo, R., and Willems, R.J. (2006). Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. Journal of Clinical Microbiology. 44(6), 2220-2228.

168. Sahm, D.F., Kissinger, J., Gilmore, M.S., Murray, P.R., Mulder, R., Solliday, J., and Clarke, B. (1989). *In vitro* susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. Antimicrobial Agents and Chemotherapy. 33(9), 1588-1591.

169. Saile, E., and Koehler, T.M. (2002). Control of anthrax toxin gene expression by the transition state regulator *abrB*. Journal of Bacteriology. 184(2), 370-380.

170. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences (USA). 74(12), 5463-5467.

171. Sanson, M., Makthal, N., Gavagan, M., Cantu, C., Olsen, R.J., Musser, J.M., and Kumaraswami, M. (2015). Phosphorylation events in the multiple gene regulator of group A *Streptococcus* significantly influence global gene expression and virulence. Infection and Immunity. 83(6), 2382-2395.

172. Schleifer, K.H., and Kilpper-Bälz, R. (1984). Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the Genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. International Journal of Systematic and Evolutionary Microbiology. 34(1), 31-34.

173. Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C_T method. Nature Protocols. 3(6), 1101-1108.

174. Segarra, R.A., Booth, M.C., Morales, D.A., Huycke, M.M., and Gilmore, M.S. (1991). Molecular characterization of the *Enterococcus faecalis* cytolysin activator. Infection and Immunity. 59(4), 1239-1246.

175. Shankar, V., Baghdayan, A.S., Huycke, M.M., Lindahl, G., and Gilmore, M.S. (1999). Infection-derived *Enterococcus faecalis* strains Are enriched in *esp*, a gene encoding a novel surface protein. Infection and Immunity. 67(1), 193-200.

176. Shepard, B.D., and Gilmore, M.S. (1995). Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. In: Electroporation Protocols for Microorganisms, J.A. Nickoloff ed. (Humana Press, Totowa, NJ), pp. 217-226.

177. Sherman, J.M. (1937). The streptococci. Bacteriological Reviews. 1(1), 3-97.

178. Siggers, T., and Gordân, R. (2014). Protein-DNA binding: complexities and multi-protein codes. Nucleic Acids Research. 42(4), 2099-2111.

179. Simpson, W.J., LaPenta, D., Chen, C., and Cleary, P.P. (1990). Coregulation of type 12 M protein and streptococcal C5a peptidase genes in group A streptococci: evidence for a virulence regulon controlled by the *virR* locus. Journal of Bacteriology. 172(2), 696-700.

180. Singh, K.V., Lewis, R.J., and Murray, B.E. (2009). Importance of the *epa* locus of *Enterococcus faecalis* OG1RF in a mouse model of ascending urinary tract infection. The Journal of Infectious Diseases. 200(3), 417-420.

181. Singh, K.V., Nallapareddy, S.R., and Murray, B.E. (2007). Importance of the *ebp* (endocarditis- and biofilm-associated pilus) locus in the pathogenesis of *Enterococcus faecalis* ascending urinary tract infection. The Journal of Infectious Diseases. 195(11), 1671-1677.

182. Singh, K.V., Nallapareddy, S.R., Nannini, E.C., and Murray, B.E. (2005). Fsr-independent production of protease(s) may explain the lack of attenuation of an *Enterococcus faecalis fsr* mutant versus a *gelE-sprE* mutant in induction of endocarditis. Infection and Immunity. 73(8), 4888-4894.

183. Singh, K.V., Nallapareddy, S.R., Sillanpaa, J., and Murray, B.E. (2010). Importance of the collagen adhesin ace in pathogenesis and protection against *Enterococcus faecalis* experimental endocarditis. PLOS Pathogens. 6(1), e1000716.

184. Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research. 33 (Web server issue), W244-W248.

185. Solano-Collado, V. (2014). Caracterización molecular del regulador transcripcional Mga*Spn* de *Streptococcus pneumoniae*. Supervised by A. Bravo. Centro de Investigaciones Biológicas, CSIC. Tesis Doctoral, Departamento de Bioquímica y Biología Molecular I, Universidad Complutense de Madrid.

186. Solano-Collado, V., Espinosa, M., and Bravo, A. (2012). Activator role of the pneumococcal *Mga*-like virulence transcriptional regulator. Journal of Bacteriology. 194(16), 4197-4207.

187. Solano-Collado, V., Hüttener, M., Espinosa, M., Juárez, A., and Bravo, A. (2016). Mga*Spn* and H-NS: two unrelated global regulators with similar DNA-binding properties. Frontiers in Molecular Biosciences. 3(60), 1-12.

188. Solano-Collado, V., Lurz, R., Espinosa, M., and Bravo, A. (2013). The pneumococcal Mga*Spn* virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. Nucleic Acids Research. 41(14), 6975-6991.

189. Sorensen, R.U., and Edgar, J.D.M. (2018). Overview of antibody-mediated immunity to *S. pneumoniae*: pneumococcal infections, pneumococcal immunity assessment, and recommendations for IG product evaluation. Transfusion. 58, 3106-3113.

190. Sorin, A., Rosas, G., and Rao, R. (1997). PMR1, a Ca^{2+}-ATPase in yeast Golgi, has properties distinct from sarco/endoplasmic reticulum and plasma membrane calcium pumps. Journal of Biological Chemistry. 272(15), 9895-9901.

191. Steel, H.C., Cockeran, R., Anderson, R., and Feldman, C. (2013). Overview of community-acquired pneumonia and the role of inflammatory mechanisms in the immunopathogenesis of severe pneumococcal disease. Mediators of Inflammation. 2013(490346), 1-18.

192. Stevens, D.L., and Bryant, A.E. (2016). Impetigo, erysipelas and cellulitis. In: *Streptococcus pyogenes*: Basic Biology to Clinical Manifestations F. JJ, S. DL, and F. VA eds. (University of Oklahoma Health Sciences Center, Oklahoma City (OK)), pp. 1-18.

193. Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. Annual Review of Biochemistry. 69, 183-215.

194. Strauch, M.A., Ballar, P., Rowshan, A.J., and Zoller, K.L. (2005). The DNA-binding specificity of the *Bacillus anthracis* AbrB protein. Microbiology. 151(6), 1751-1759.

195. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology. 189(1), 113-130.

196. Sundar, G.S., Islam, E., Gera, K., Le Breton, Y., and McIver, K.S. (2017). A PTS EII mutant library in group A *Streptococcus* identifies a promiscuous man-family PTS transporter influencing SLS-mediated hemolysis. Molecular Microbiology. 103(3), 518-533.

197. Tendolkar, P.M., Baghdayan, A.S., Gilmore, M.S., and Shankar, N. (2004). Enterococcal surface protein, Esp, enhances biofilm formation by *Enterococcus faecalis*. Infection and Immunity. 72(10), 6032-6039.

198. Teng, F., Jacques-Palaz, K.D., Weinstock, G.M., and Murray, B.E. (2002). Evidence that the enterococcal polysaccharide antigen gene (*epa*) cluster is widespread in *Enterococcus faecalis* and influences resistance to phagocytic killing of *E. faecali*s. Infection and Immunity. 70(4), 2010-2015.

199. Teng, F., Singh, K.V., Bourgogne, A., Zeng, J., and Murray, B.E. (2009). Further characterization of the *epa* gene cluster and Epa polysaccharides of *Enterococcus faecalis*. Infection and Immunity. 77(9), 3759-3767.

200. Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., and Fraser, C.M. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science. 293(5529), 498-506.

201. Thibessard, A., Borges, F., Fernandez, A., Gintz, B., Decaris, B., and Leblond-Bourget, N. (2004). Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. Applied and Environmental Microbiology. 70(4), 2220-2229.

202. Thomas, V.C., Hiromasa, Y., Harms, N., Thurlow, L., Tomich, J., and Hancock, L.E. (2009). A fratricidal mechanism is responsible for eDNA release and contributes to biofilm development of *Enterococcus faecalis*. Molecular Microbiology. 72(4), 1022-1036.

203. Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M., and Burnham, M.K.R. (2000). A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. Molecular Microbiology. 35(3), 566-576.

204. Thurlow, L.R., Thomas, V.C., Fleming, S.D., and Hancock, L.E. (2009). *Enterococcus faecalis* capsular polysaccharide serotypes C and D and their contributions to host innate immune evasion. Infection and Immunity. 77(12), 5551-5557.

205. Toledo-Arana, A., Valle, J., Solano, C., Arrizubieta, M.a.J., Cucarella, C., Lamata, M., Amorena, B., Leiva, J., Penadés, J.R., and Lasa, I. (2001). The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. Applied and Environmental Microbiology. 67(10), 4538-4545.

206. Tsvetanova, B., Wilson, A.C., Bongiorni, C., Chiang, C., Hoch, J.A., and Perego, M. (2007). Opposing effects of histidine phosphorylation regulate the AtxA virulence transcription factor in *Bacillus anthracis*. Molecular Microbiology. 63(3), 644-655.

207. Uchida, I., Hornung, J.M., Thorne, C.B., Klimpel, K.R., and Leppla, S.H. (1993). Cloning and characterization of a gene whose product is a trans-activator of anthrax toxin synthesis. Journal of Bacteriology. 175(17), 5329-5338.

208. Vahling, C.M., and McIver, K.S. (2006). Domains required for transcriptional activation show conservation in the Mga family of virulence gene regulators. Journal of Bacteriology. 188(3), 863-873.

209. Valdes, K.M., Sundar, G.S., Belew, A.T., Islam, E., El-Sayed, N.M., Le Breton, Y., and McIver, K.S. (2018). Glucose levels alter the Mga virulence regulon in the group A *Streptococcus*. Scientific Reports. 8(1), 4971-4971.

210. Van Baelen, K., Vanoevelen, J., Missiaen, L., Raeymaekers, L., and Wuytack, F. (2001). The golgi PMR1 P-type ATPase of *Caenorhabditis elegans*. Identification of the gene and demonstration of calcium and manganese transport. Journal of Biological Chemistry. 276(14), 10683-10691.

211. van Montfort, R.L.M., Pijning, T., Kalk, K.H., Reizer, J., Saier, M.H., Jr., Thunnissen, M.M.G.M., Robillard, G.T., and Dijkstra, B.W. (1997). The structure of an energy-coupling protein from bacteria, $IIB^{cellobiose}$, reveals similarity to eukaryotic protein tyrosine phosphatases. Structure. 5(2), 217-225.

212. van Schaik, W., Château, A., Dillies, M.-A., Coppée, J.-Y., Sonenshein, A.L., and Fouet, A. (2009). The global regulator CodY regulates toxin gene expression in *Bacillus anthracis* and is required for full virulence. Infection and Immunity. 77(10), 4437-4445.

213. Van Tyne, D., and Gilmore, M.S. (2014). Friend turned foe: evolution of enterococcal virulence and antibiotic resistance. Annual Review of Microbiology. 68(1), 337-356.

214. Vebø, H.C., Snipen, L., Nes, I.F., and Brede, D.A. (2009). The transcriptome of the nosocomial pathogen *Enterococcus faecalis* V583 reveals adaptive responses to growth in blood. PLoS ONE. 4(11), e7660.

215. Vebø, H.C., Solheim, M., Snipen, L., Nes, I.F., and Brede, D.A. (2010). Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. PLoS ONE. 5(8), e12489.

216. Vlahoviček, K., Kaján, L.s., and Pongor, S.n. (2003). DNA analysis servers: plot.it, bend.it, model.it and IS. Nucleic Acids Research. 31(13), 3686-3687.

217. Waar, K., Muscholl-Silberhorn, A.B., Willems, R.J., Slooff, M.J., Harmsen, H.J., and Degener, J.E. (2002). Genogrouping and incidence of virulence factors of *Enterococcus faecalis* in liver transplant patients differ from blood culture and fecal isolates. The Journal of Infectious Diseases. 185(8), 1121-1127.

218. Walsh, R.L., and Camilli, A. (2011). *Streptococcus pneumoniae* is desiccation tolerant and infectious upon rehydration. mBio. 2(3), e00092-00011.

219. Wang, J., Li, J.-W., Li, J., Huang, Y., Wang, S., and Zhang, J.-R. (2020). Regulation of pneumococcal epigenetic and colony phases by multiple two-component regulatory systems. PLOS Pathogens. 16(3), e1008417.

220. Waters, C.M., Antiporta, M.H., Murray, B.E., and Dunny, G.M. (2003). Role of the *Enterococcus faecalis* GelE protease in determination of cellular chain length, supernatant pheromone levels, and degradation of fibrin and misfolded surface proteins. Journal of Bacteriology. 185(12), 3613-3623.

221. Weiser, J.N., Austrian, R., Sreenivasan, P.K., and Masure, H.R. (1994). Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infection and Immunity. 62(6), 2582-2589.

222. Weiser, J.N., Ferreira, D.M., and Paton, J.C. (2018). *Streptococcus pneumoniae*: transmission, colonization and invasion. Nature Reviews Microbiology. 16(6), 355-367.

223. Xu, J., Xie, Y.-D., Liu, L., Guo, S., Su, Y.-L., and Li, A.-X. (2019). Virulence regulation of cel-EIIB protein mediated PTS system in *Streptococcus agalactiae* in Nile tilapia. Journal of Fish Diseases. 42(1), 11-19.

224. Xu, Y., Singh, K.V., Qin, X., Murray, B.E., and Weinstock, G.M. (2000). Analysis of a gene cluster of *Enterococcus faecalis* involved in polysaccharide biosynthesis. Infection and Immunity. 68(2), 815-823.

225. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. Gene. 33(1), 103-119.

226. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 13(134), 1-11.

ANNEXES

Table S1. Identical MafR groups. The name of the *E. faecalis* strain that represents the group is indicated. Data from the National Center for Biotechnology Information (05/04/2020)

CDS Region in Nucleotide	Strain	Assembly
Identical proteins to MafR from GA2		
NZ_AYKL01000045.138879-40327 (+)	**GA2**	**GCF_000648035.1**
NZ_WMFX01000003.139053-40501 (+)	**C34**	GCF_009735255.1
NZ_PTUY01000016.112087-13535 (+)	**CVM N60027F**	GCF_002943975.1
NZ_PTTX01000055.118245-19693 (-)	**CVM N60271F**	GCF_002944015.1
NZ_KB944720.1137859-139307 (+)	**RMC1**	GCF_000393135.1
Identical proteins to MafR from KS19		
NZ_NGMZ01000003.1118141-119589 (+)	**11D3_DIV0676**	GCF_002140335.1
NZ_NGMY01000001.12612413-2613861 (+)	**11D3_DIV0676_a**	GCF_002140345.1
NZ_QNGQ01000003.148500-49948 (+)	**15407**	GCF_003962565.1
NZ_NGMP01000004.1247157-248605 (-)	**2A11_DIV0231**	GCF_002141205.1
NZ_JVAI01000024.113251-14699 (+)	**606_EFLS**	GCF_001055435.1
NZ_PJXE01000002.1248003-249451 (-)	**EN359**	GCF_004126035.1
NZ_PJWZ01000010.155172-56620 (-)	**EN391b**	GCF_004126235.1
NZ_AYND01000016.129507-30955 (+)	**KS19**	**GCF_000648075.1**
NZ_JH804790.147968-49416 (+)	**R508**	GCF_000294325.2
Identical proteins to MafR from MTmid8		
NZ_VWNO01000006.142051-43499 (-)	**11-7**	GCF_009830615.1

CDS Region in Nucleotide	Strain	Assembly
NZ_VWNM01000010.141973-43421 (-)	**11-8**	GCF_009830565.1
NZ_CABGWV010000003.198596-100044 (+)	**4928STDY7071224**	GCF_902162785.1
NZ_CABHAZ010000003.198571-100019 (+)	**4928STDY7071744**	GCF_902163855.1
NZ_PUAJ01000028.121566-23014 (-)	**CVM N55113**	GCF_002947695.1
NZ_PUAR01000026.130195-31643 (+)	**CVM N55268**	GCF_002947395.1
NZ_GG692673.1293317-294765 (-)	**E1Sol**	GCF_000157395.1
NZ_QDDM01000003.1236354-237802 (-)	**EF294**	GCF_003075115.1
NZ_QDDN01000003.1236369-237817 (-)	**EF346**	GCF_003075155.1
NZ_RKMU01000004.1257252-258700 (-)	**LIT10 A36**	GCF_003796585.1
NZ_RKMO01000003.1257865-259313 (-)	**LIT15-A36'**	GCF_003796545.1
NZ_RKMP01000003.1257865-259313 (-)	**LIT15-A36**	GCF_003796565.1
NZ_RKMV01000003.1257865-259313 (-)	**LIT9-A36**	GCF_003795725.1
NZ_AYKU01000025.133859-35307 (+)	**MTmid8**	**GCF_000648135.1**
NZ_RKOM01000003.1257865-259313 (-)	**P10 CL A21**	GCF_003796965.1
NZ_RKPT01000003.1257865-259313 (-)	**P12 C A7**	GCF_003797105.1
NZ_RKOG01000003.1257865-259313 (-)	**P17 CL A21**	GCF_003796195.1
NZ_RKPJ01000003.1257865-259313 (-)	**P3 CL A14**	GCF_003797595.1
NZ_RKPZ01000003.1257865-259313 (-)	**P5 CL A7**	GCF_003797645.1

CDS Region in Nucleotide	Strain	Assembly
Identical proteins to MafR from B653		
NZ_KB944834.1156908-158356 (+)	**B653**	**GCF_000393335.1**
NZ_WMFW01000004.1246800-248248 (-)	**W314**	GCF_009735275.1
Identical proteins to MafR from X98		
NZ_KE350816.131306-32754 (+)	**02-MB-P-10**	GCF_000415005.1
NZ_LMBS01000007.1419406-420854 (+)	**16A-pre**	GCF_003056005.1
NZ_PGCH01000006.1149472-150920 (-)	**24FS**	GCF_002812965.1
NZ_FFHQ01000002.1500778-502226 (+)	**2842STDY5753961**	GCF_900017475.1
NZ_CAACXT010000001.1606972-608420 (+)	**3012STDY6244128**	GCF_900683495.1
NZ_CABGKJ010000004.1250924-252372 (-)	**4928STDY7071313**	GCF_902159565.1
NZ_CABGOA010000004.1250927-252375 (-)	**4928STDY7071439**	GCF_902160505.1
NZ_CABGNZ010000004.1250699-252147 (-)	**4928STDY7071440**	GCF_902160455.1
NZ_CABGOI010000004.1250699-252147 (-)	**4928STDY7071441**	GCF_902160595.1
NZ_CABGOE010000003.1251535-252983 (-)	**4928STDY7071442**	GCF_902160615.1
NZ_CABGON010000003.1251234-252682 (-)	**4928STDY7071443**	GCF_902160535.1
NZ_CABGOJ010000003.1251242-252690 (-)	**4928STDY7071444**	GCF_902160545.1
NZ_CABGPD010000003.137708-39156 (+)	**4928STDY7071480**	GCF_902160815.1
NZ_CABGPB010000003.137708-39156 (+)	**4928STDY7071481**	GCF_902160805.1
NZ_CABGSA010000005.137711-39159 (+)	**4928STDY7071563**	GCF_902161525.1
NZ_CABGRX010000004.1188052-189500 (-)	**4928STDY7071564**	GCF_902161515.1
NZ_CABGKB010000012.137714-39162 (+)	**4928STDY7071637**	GCF_902159465.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABHAG010000003.137160-38608 (+)	**4928STDY7071727**	GCF_902163675.1
NZ_CABHAC010000003.1100412-101860 (+)	**4928STDY7071728**	GCF_902163635.1
NZ_CABHAM010000004.1250700-252148 (-)	**4928STDY7071737**	GCF_902163685.1
NZ_LR607359.12980819-2982267 (+)	**4928STDY7071765**	GCF_902166575.1
NZ_LR607356.12962863-2964311 (+)	**4928STDY7071766**	GCF_902166605.1
NZ_LR607378.12278340-2279788 (+)	**4928STDY7387896**	GCF_902166815.1
NZ_PUBK01000064.115322-16770 (-)	**CVM N52648**	GCF_002946875.1
NZ_PTYG01000086.17896-9344 (+)	**CVM N52789**	GCF_002948635.1
NZ_PUAF01000044.115321-16769 (-)	**CVM N54676**	GCF_002948815.1
NZ_PUAL01000072.115322-16770 (-)	**CVM N55181**	GCF_002947595.1
NZ_PTWM01000080.120143-21591 (+)	**CVM N59544F**	GCF_002946275.1
NZ_PTWF01000048.138897-40345 (+)	**CVM N59584F**	GCF_002945595.1
NZ_PTVQ01000016.125881-27329 (-)	**CVM N59847F**	GCF_002945315.1
NZ_PTVE01000030.17896-9344 (+)	**CVM N59936F**	GCF_002945015.1
NZ_PTVB01000052.17010-8458 (+)	**CVM N59959F**	GCF_002944275.1
NZ_KI913094.1331766-333214 (-)	**EnGen0426**	GCF_000519945.1
NZ_SZVH01000008.137770-39218 (+)	**ES-13**	GCF_009659905.1
NZ_CP022059.1909711-911159 (-)	**FDAARGOS_338**	GCF_002208945.1
NZ_AYKK01000002.15551-6999 (-)	**FL2**	GCF_000648015.1
NZ_KB944767.194963-96411 (+)	**Fly 2**	GCF_000393215.1
NZ_VHRL01000004.1184651-186099 (-)	**H74**	GCF_006541115.1
NZ_KE350328.1209696-211144 (-)	**KI-6-1-110608-1**	GCF_000414985.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KE352795.110504-11952 (-)	**LA3B-2**	GCF_000415465.2
NZ_AYKM01000032.139187-40635 (+)	**MN16**	GCF_000648115.1
NZ_UGIR01000002.1476969-478417 (-)	**NCTC13705**	GCF_900447775.1
NZ_UGIJ01000002.1476969-478417 (-)	**NCTC13763**	GCF_900447725.1
NZ_CABEIO010000002.12561873-2563321 (+)	**NCTC8729**	GCF_901543445.1
NZ_LR594051.12561793-2563241 (+)	**NCTC8732**	GCF_901553725.1
NZ_QSSZ01000004.1268669-270117 (-)	**OM08-2AT**	GCF_003438055.1
NZ_KB944604.1170424-171872 (+)	**T12**	GCF_000392895.1
NZ_KB947301.1101283-102731 (+)	**T16**	GCF_000394975.1
NZ_KB944679.1141914-143362 (+)	**T18**	GCF_000393075.1
NZ_GL454526.136848-38296 (+)	**TX0043**	GCF_000147515.1
NZ_GL457161.147095-48543 (+)	**TX1341**	GCF_000148225.1
NZ_GL454567.114152-15600 (-)	**TX1342**	GCF_000147475.1
NZ_GG688424.1247954-249402 (-)	**X98**	**GCF_000157515.1**
Identical proteins to MafR from AZ19		
NZ_VWNZ01000003.1248799-250247 (-)	**207AE**	GCF_009830855.1
NZ_CABHBH010000008.19989-11437 (-)	**4928STDY7071764**	GCF_902163965.1
NZ_JVBD01000053.112348-13796 (-)	**584_EFLS**	GCF_001057435.1
NZ_NGMW01000001.12169774-2171222 (+)	**9A2_4861**	GCF_002140925.1
NZ_NGMX01000001.1238509-239957 (-)	**9F2_4866**	GCF_002140985.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KB944643.1107667-109115 (+)	**ATCC 35038**	GCF_000392955.1
NZ_AYLU01000053.112354-13802 (-)	**AZ19**	**GCF_000647995.1**
NZ_RJJO01000011.142668-44116 (+)	**Ca-2**	GCF_003996795.1
NZ_PTXJ01000058.115820-17268 (+)	**CVM N59409F**	GCF_002946215.1
NZ_PKMO01000003.1177675-179123 (-)	**EN741**	GCF_004120285.1
NZ_PJWP01000010.146818-48266 (-)	**EN768**	GCF_004125915.1
NZ_CP043724.12574486-2575934 (+)	**L14**	GCF_009914685.1
NZ_GL454530.112322-13770 (-)	**TX1302**	GCF_000147495.1
NZ_GL455430.112318-13766 (-)	**TX0109**	GCF_000148105.1
NZ_KB932564.1254045-255493 (-)	**UAA769**	GCF_000391565.1
Identical proteins to MafR from Com1		
NZ_KE350943.141859-43307 (+)	**02-MB-BW-10**	GCF_000415045.1
NZ_QFYJ01000003.140023-41471 (+)	**100-2016**	GCF_003144675.1
NZ_QFYI01000001.11905920-1907368 (-)	**124-2017**	GCF_003144635.1
NZ_NHNF01000004.1264537-265985 (-)	**12_Tun**	GCF_002220885.1
NZ_JTKU01000008.1263975-265423 (+)	**17**	GCF_000788185.1
NZ_CABGHQ010000011.143443-44891 (+)	**4928STDY7071220**	GCF_902158925.1
NZ_CABGWZ010000006.151919-53367 (-)	**4928STDY7071221**	GCF_902162855.1
NZ_CABGXS010000003.1257038-258486 (-)	**4928STDY7071258**	GCF_902162995.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABGLS010000003.1259391-260839 (-)	**4928STDY7071364**	GCF_902159915.1
NZ_CABGID010000003.1258962-260410 (-)	**4928STDY7071365**	GCF_902159085.1
NZ_CABGLP010000003.1258962-260410 (-)	**4928STDY7071366**	GCF_902159895.1
NZ_CABGLX010000003.1258962-260410 (-)	**4928STDY7071367**	GCF_902160035.1
NZ_CABGLW010000003.1258962-260410 (-)	**4928STDY7071368**	GCF_902159945.1
NZ_CABGJV010000007.1112285-113733 (-)	**4928STDY7071548**	GCF_902159445.1
NZ_CABGTJ010000009.112963-14411 (-)	**4928STDY7071609**	GCF_902161895.1
NZ_CABGTP010000011.112963-14411 (-)	**4928STDY7071611**	GCF_902161905.1
NZ_CABGTK010000004.1263883-265331 (+)	**4928STDY7071612**	GCF_902161915.1
NZ_QFYK01000001.12827401-2828849 (+)	**58-2013**	GCF_003144655.1
NZ_KB946667.1110751-112199 (+)	**Com 2**	GCF_000394375.1
NZ_KB947407.1109632-111080 (+)	**Com1**	**GCF_000395175.1**
NZ_PTZI01000026.119773-21221 (-)	**CVM N54118**	GCF_002948195.1
NZ_PTZJ01000076.119773-21221 (-)	**CVM N54121**	GCF_002948975.1
NZ_PTZV01000030.115081-16529 (+)	**CVM N54553**	GCF_002947985.1
NZ_PUAK01000045.119773-21221 (-)	**CVM N55119**	GCF_002947635.1
NZ_PTWS01000008.115074-16522 (+)	**CVM N59508F**	GCF_002945735.1
NZ_PTWA01000010.15452-6900 (-)	**CVM N59600F**	GCF_002943885.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KI912926.12545756-2547204 (+)	**EnGen0400**	GCF_000519425.1
NZ_KI913021.1105439-106887 (+)	**EnGen0414**	GCF_000519705.1
NZ_VHRS01000008.112206-13654 (+)	**H102**	GCF_006541755.1
NZ_AP018538.12534195-2535643 (+)	**KUB3006**	GCF_003966385.1
NZ_AP018543.12531127-2532575 (+)	**KUB3007**	GCF_003966405.1
NZ_QYMB01000013.142061-43509 (+)	**N015.C-18**	GCF_008082935.1

Table S2. *E. faecalis* genomes that encode a protein identical to MafR$_{V583}$. Data from the National Center for Biotechnology Information (03/04/2020)

CDS Region in Nucleotide	Strain	Assembly
NZ_CABEGK010000009.138965-40413 (+)		GCF_901521655.1
NZ_ASWX01000026.111596-13044 (-)	**10244**	GCF_000438765.1
NZ_JVZS01000108.19226-10674 (+)	**1187_EFLS**	GCF_001054015.1
NZ_JTKT01000044.157194-58642 (-)	**12**	GCF_000788155.1
NZ_JVXC01000028.138901-40349 (+)	**1248_EFLS**	GCF_001053145.1
NZ_NSJS01000009.160862-62310 (+)	**13.1**	GCF_002289025.1
NZ_JVTX01000054.111834-13282 (-)	**132_EFLS**	GCF_001053375.1
NZ_JVTG01000041.111653-13101 (-)	**1333_EFLS**	GCF_001053475.1
NZ_JTKW01000011.111517-12965 (+)	**19**	GCF_000788165.1
NZ_JVOC01000096.111622-13070 (-)	**257_EFLS**	GCF_001056115.1
NZ_JVJB01000075.145803-47251 (+)	**388_EFLS**	GCF_001056555.1
NZ_NSJR01000012.19305-10753 (+)	**44.1**	GCF_002288985.1
NZ_CABGHK010000005.1100279-101727 (-)	**4928STDY7071200**	GCF_902158875.1
NZ_CABGHW010000005.1100279-101727 (-)	**4928STDY7071218**	GCF_902158905.1
NZ_CABGYG010000004.1128975-130423 (+)	**4928STDY7071281**	GCF_902163155.1
NZ_CABGMA010000003.139612-41060 (+)	**4928STDY7071380**	GCF_902160005.1
NZ_CABGMN010000006.166444-67892 (+)	**4928STDY7071394**	GCF_902160125.1
NZ_CABGMW010000006.199827-101275 (-)	**4928STDY7071395**	GCF_902160185.1
NZ_CABGOB010000007.161365-62813 (+)	**4928STDY7071435**	GCF_902160465.1
NZ_CABGJG010000005.1100277-101725 (-)	**4928STDY7071437**	GCF_902159165.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABGJA010000005.198936-100384 (+)	**4928STDY7071438**	GCF_902159185.1
NZ_CABGQE010000006.1100260-101708 (-)	**4928STDY7071513**	GCF_902161085.1
NZ_CABGJT010000004.1128978-130426 (+)	**4928STDY7071574**	GCF_902159395.1
NZ_CABGSG010000004.1100279-101727 (-)	**4928STDY7071575**	GCF_902161665.1
NZ_CABGSN010000005.1129561-131009 (+)	**4928STDY7071576**	GCF_902161575.1
NZ_CABGUQ010000002.19303-10751 (+)	**4928STDY7071660**	GCF_902162275.1
NZ_CABGZY010000006.163625-65073 (+)	**4928STDY7071723**	GCF_902163585.1
NZ_CABHAI010000004.1128990-130438 (+)	**4928STDY7071736**	GCF_902163725.1
NZ_CABHAQ010000005.198103-99551 (+)	**4928STDY7071738**	GCF_902163805.1
NZ_CABHBB010000007.1100260-101708 (-)	**4928STDY7071753**	GCF_902163825.1
NZ_CABHAW010000012.1100259-101707 (-)	**4928STDY7071754**	GCF_902163915.1
NZ_CABHBQ010000007.1100279-101727 (-)	**4928STDY7071769**	GCF_902164075.1
NZ_CABHDL010000004.1129006-130454 (+)	**4928STDY7387716**	GCF_902164485.1
NZ_CABHDC010000010.145951-47399 (+)	**4928STDY7387721**	GCF_902164425.1
NZ_CABHEI010000004.1129033-130481 (+)	**4928STDY7387749**	GCF_902164855.1
NZ_CABHEG010000005.1100911-102359 (-)	**4928STDY7387750**	GCF_902164845.1
NZ_KB932369.1173931-175379 (+)	**7330082-2**	GCF_000390745.1
NZ_KK640503.174949-76397 (-)	**918**	GCF_000690925.1
NZ_KB944973.1206852-208300 (+)	**A-3-1**	GCF_000393495.1
NZ_KB944874.1156800-158248 (+)	**ATCC 10100**	GCF_000393395.1
NZ_KB947429.11691051-1692499 (+)	**ATCC 27275**	GCF_000395245.1
NZ_KB947435.11679303-1680751 (+)	**ATCC 27959**	GCF_000395265.1

CDS Region in Nucleotide	Strain	Assembly
NZ_JSES01000030.174792-76240 (-)	**ATCC 51299**	GCF_000772515.1
NZ_KB948285.162407-63855 (+)	**B1005**	GCF_000395985.1
NZ_KB948306.12395739-2397187 (+)	**B1138**	GCF_000396005.1
NZ_KB948323.162413-63861 (+)	**B1249**	GCF_000396025.1
NZ_KB932412.172356-73804 (+)	**B1290**	GCF_000390945.1
NZ_KB932415.12614098-2615546 (+)	**B1327**	GCF_000390965.1
NZ_KB932419.1166162-167610 (+)	**B1376**	GCF_000390985.1
NZ_PZZE01000001.12826211-2827659 (+)	**B1376**	GCF_003046865.1
NZ_KB932422.1274971-276419 (-)	**B1385**	GCF_000391005.1
NZ_KB932428.1161711-163159 (+)	**B1441**	GCF_000391025.1
NZ_KB932431.12347403-2348851 (+)	**B1505**	GCF_000391045.1
NZ_KB932436.1162878-164326 (+)	**B1532**	GCF_000391065.1
NZ_PZZF01000001.1801001-802449 (+)	**B1532**	GCF_003046985.1
NZ_KB932314.1203471-204919 (-)	**B15725**	GCF_000390505.1
NZ_KB932437.1800614-802062 (+)	**B1586**	GCF_000391085.1
NZ_KB932443.1163039-164487 (+)	**B1618**	GCF_000391105.1
NZ_KB932444.1165203-166651 (+)	**B1623**	GCF_000391125.1
NZ_KB932449.1161281-162729 (+)	**B1678**	GCF_000391145.1
NZ_KB932458.1164912-166360 (+)	**B1696**	GCF_000391165.1
NZ_KB932466.1166349-167797 (+)	**B1719**	GCF_000391185.1
NZ_KB932470.1163052-164500 (+)	**B1734**	GCF_000391205.1
NZ_PZZG01000002.12826523-2827971 (+)	**B1734**	GCF_003046945.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KB945564.172356-73804 (+)	**B1843**	GCF_000391225.1
NZ_KB948376.172356-73804 (+)	**B1851**	GCF_000396045.1
NZ_KB932480.172356-73804 (+)	**B1874**	GCF_000391245.1
NZ_KB948413.172356-73804 (+)	**B1933**	GCF_000396085.1
NZ_KB948433.172356-73804 (+)	**B2202**	GCF_000396105.1
NZ_KB948456.162425-63873 (+)	**B2211**	GCF_000396125.1
NZ_KB932491.159512-60960 (+)	**B2255**	GCF_000391285.1
NZ_KB948493.172356-73804 (+)	**B2277**	GCF_000396145.1
NZ_KB945593.172356-73804 (+)	**B2391**	GCF_000391305.1
NZ_KB932507.172356-73804 (+)	**B2488**	GCF_000391325.1
NZ_KB932514.172356-73804 (+)	**B2535**	GCF_000391345.1
NZ_PZZH01000001.12825180-2826628 (+)	**B2535**	GCF_003046965.1
NZ_KB932519.172356-73804 (+)	**B2557**	GCF_000391365.1
NZ_KB932520.189368-90816 (+)	**B2593**	GCF_000391385.1
NZ_KB948517.11834358-1835806 (+)	**B2670**	GCF_000396165.1
NZ_KB948592.1284748-286196 (+)	**B2685**	GCF_000396185.1
NZ_KB948595.165812-67260 (+)	**B2687**	GCF_000396205.1
NZ_KB948636.172356-73804 (+)	**B2802**	GCF_000396225.1
NZ_KB948683.172356-73804 (+)	**B2813**	GCF_000396245.1
NZ_KB948725.165632-67080 (+)	**B2864**	GCF_000396265.1
NZ_KB948745.162413-63861 (+)	**B2867**	GCF_000396285.1
NZ_KB948774.1166161-167609 (+)	**B2949**	GCF_000396305.1

CDS Region in Nucleotide	Strain	Assembly
NZ_PZZI01000002.12785569-2787017 (+)	**B2949**	GCF_003046905.1
NZ_KB932527.11819761-1821209 (+)	**B3031**	GCF_000391405.1
NZ_KB932529.159512-60960 (+)	**B3042**	GCF_000391425.1
NZ_KB932534.159512-60960 (+)	**B3053**	GCF_000391445.1
NZ_KB948805.172356-73804 (+)	**B3119**	GCF_000396325.1
NZ_KB932546.162419-63867 (+)	**B3126**	GCF_000391465.1
NZ_KB948848.172356-73804 (+)	**B3196**	GCF_000396345.1
NZ_KB948881.1166489-167937 (+)	**B3286**	GCF_000396365.1
NZ_PZZJ01000002.12827837-2829285 (+)	**B3286**	GCF_003046915.1
NZ_KB948900.1124817-126265 (+)	**B3336**	GCF_000396385.1
NZ_KB948925.1166224-167672 (+)	**B4008**	GCF_000396405.1
NZ_KB948964.172356-73804 (+)	**B4018**	GCF_000396425.1
NZ_KB948991.1166305-167753 (+)	**B4148**	GCF_000396445.1
NZ_KB949006.172356-73804 (+)	**B4163**	GCF_000396465.1
NZ_KB949061.172356-73804 (+)	**B4259**	GCF_000396485.1
NZ_KB949100.172356-73804 (+)	**B4267**	GCF_000396505.1
NZ_KB949122.162413-63861 (+)	**B4270**	GCF_000396525.1
NZ_KB949162.172356-73804 (+)	**B4411**	GCF_000396545.1
NZ_KB949233.162413-63861 (+)	**B4568**	GCF_000396565.1
NZ_KB949267.1166041-167489 (+)	**B4638**	GCF_000396585.1
NZ_KB949294.172356-73804 (+)	**B4672**	GCF_000396605.1
NZ_KB949324.12422921-2424369 (+)	**B4674**	GCF_000396625.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KB949330.172356-73804 (+)	**B4969**	GCF_000396645.1
NZ_PZZK01000002.12829227-2830675 (+)	**B4969**	GCF_003046875.1
NZ_KB946644.165587-67035 (+)	**B5035**	GCF_000394355.1
NZ_KB949383.160768-62216 (+)	**B5076**	GCF_000396665.1
NZ_CP041738.12825998-2827446 (+)	**B594**	GCF_000391485.2
NZ_WMGN01000036.18791-10239 (+)	**B6**	GCF_009734535.1
NZ_KB932327.1220811-222259 (-)	**B69486**	GCF_000390565.1
NZ_KB932554.174804-76252 (+)	**B878**	GCF_000391505.1
NZ_KB932556.172356-73804 (+)	**B939**	GCF_000391525.1
NZ_KI518231.1144185-145633 (+)	**BM4654**	GCF_000479065.1
NZ_WMGU01000005.199798-101246 (-)	**C379**	GCF_009734545.1
NZ_KB947187.1187033-188481 (+)	**CH570**	GCF_000394855.1
NZ_PTYC01000038.117687-19135 (-)	**CVM N52755**	GCF_002949275.1
NZ_PTXA01000087.118991-20439 (+)	**CVM N59466F**	GCF_002945915.1
NZ_KB945031.1173856-175304 (+)	**D1**	GCF_000393575.1
NC_018221.12655811-2657259 (+)	**D32**	GCF_000281195.1
NZ_KB947440.1154732-156180 (+)	**DS16**	GCF_000395285.1
NZ_KB947422.1208999-210447 (+)	**E1**	GCF_000395205.1
NZ_PJXL01000003.1286195-287643 (-)	**EN208**	GCF_004126365.1
NZ_KB949611.1168792-170240 (+)	**EnGen0253**	GCF_000396865.1
NZ_KI912956.11668739-1670187 (+)	**EnGen0402**	GCF_000519465.1
NZ_KI912960.12465563-2467011 (+)	**EnGen0403**	GCF_000519485.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KI912968.1172946-174394 (+)	**EnGen0404**	GCF_000519505.1
NZ_KI912974.11703587-1705035 (+)	**EnGen0405**	GCF_000519525.1
NZ_KI912990.11833094-1834542 (+)	**EnGen0408**	GCF_000519585.1
NZ_KI913001.11835480-1836928 (+)	**EnGen0410**	GCF_000519625.1
NZ_KI913012.1242353-243801 (+)	**EnGen0412**	GCF_000519665.1
NZ_KI913026.1172865-174313 (+)	**EnGen0415**	GCF_000519725.1
NZ_KI913052.1320649-322097 (-)	**EnGen0418**	GCF_000519785.1
NZ_KI913064.1154899-156347 (+)	**EnGen0421**	GCF_000519845.1
NZ_KI913084.1166753-168201 (+)	**EnGen0423**	GCF_000519885.1
NZ_KI913105.162425-63873 (+)	**EnGen0427**	GCF_000519965.1
NZ_VHRZ01000004.189869-91317 (-)	**G127E**	GCF_006541295.1
NZ_VHRX01000035.115103-16551 (-)	**G149**	GCF_006541345.1
NZ_PGCW01000103.112873-14321 (+)	**G701R2B0C1**	GCF_003933385.1
NZ_PGCV01000121.138965-40413 (+)	**G702R1B0**	GCF_003933445.1
NZ_FPEA01000010.174691-76139 (+)	**G884**	GCF_900117355.1
NZ_WMGT01000005.1100231-101679 (-)	**H114S2**	GCF_009734505.1
NZ_VHRN01000003.168292-69740 (+)	**H4**	GCF_006541805.1
NZ_GG668823.189903-91351 (+)	**HH22**	GCF_000160155.1
NZ_KB947080.198760-100208 (+)	**HH22**	GCF_000394775.1
NZ_LQAM01000012.1129090-130538 (+)	**ICU1-2c**	GCF_001594
NZ_BJTH01000032.171807-73255 (-)	**J107**	GCF_007034685.1
NZ_BJTJ01000030.157891-59339 (-)	**J138**	GCF_007034845.1

CDS Region in Nucleotide	Strain	Assembly
NZ_BJTS01000019.157949-59397 (-)	**J249**	GCF_007035845.1
NZ_GG692699.1268611-270059 (-)	**JH1**	GCF_000157375.1
NZ_KB945004.1172699-174147 (+)	**Merz192**	GCF_000393535.1
NZ_KB945026.1210616-212064 (+)	**Merz204**	GCF_000393555.1
NZ_KB944981.1153777-155225 (+)	**Merz89**	GCF_000393515.1
NZ_AOPW01000067.112092-13540 (+)	**MMH594**	GCF_000968735.2
NZ_KB947133.1161228-162676 (+)	**MMH594**	GCF_000394795.1
NZ_QYNC01000006.155240-56688 (-)	**N039.H-3**	GCF_008082415.1
NZ_BJMP01000014.19178-10626 (+)	**NBRC 12965**	GCF_006539225.1
NZ_UGIK01000003.1773817-775265 (-)	**NCTC12203**	GCF_900447755.1
NZ_UGIX01000001.1872045-873493 (-)	**NCTC13379**	GCF_900448045.1
NZ_UGJA01000003.13004079-3005527 (+)	**NCTC13779**	GCF_900447805.1
NZ_UGIW01000003.12638936-2640384 (+)	**NCTC2705**	GCF_900447885.1
NZ_UGIM01000001.12653195-2654643 (+)	**NCTC8175**	GCF_900447765.1
NZ_ADKN01000042.18681-10129 (+)	**PC1.1**	GCF_000178175.1
NZ_WMGP01000006.198105-99553 (+)	**R395**	GCF_009734935.1
NZ_KB947448.1205362-206810 (+)	**RC73**	GCF_000395305.1
NZ_KB947485.12396009-2397457 (+)	**RM3817**	GCF_000395385.1
NZ_NBDW01000007.140653-42101 (+)	**S14**	GCF_002110435.1
NZ_CP041877.11398929-1400377 (+)	**SCAID PHRX1-2018**	GCF_007632055.1
NZ_KB947150.1141694-143142 (+)	**SF100**	GCF_000394815.1
NZ_KB946585.1153259-154707 (+)	**SF1592**	GCF_000394315.1

CDS Region in Nucleotide	Strain	Assembly
NZ_KB946586.15031-6479 (+)	**SF1592**	GCF_000394315.1
NZ_KB946554.1132962-134410 (+)	**SF19**	GCF_000394295.1
NZ_KB944934.1144174-145622 (+)	**SF21520**	GCF_000393455.1
NZ_KB946404.1119270-120718 (+)	**SF26630**	GCF_000394115.1
NZ_KB947168.11764920-1766368 (+)	**SF370**	GCF_000394835.1
NZ_KB944940.1106377-107825 (+)	**TR161**	GCF_000393475.1
NZ_GL454322.1203289-204737 (-)	**TX0309A**	GCF_000147555.1
NZ_GL454242.138868-40316 (+)	**TX0309B**	GCF_000147575.1
NZ_GL454421.111530-12978 (-)	**TX4248**	GCF_000147925.1
NZ_KB946185.1144552-146000 (+)	**UAA1489**	GCF_000393795.1
NZ_KB932569.1155111-156559 (+)	**UAA823**	GCF_000391585.1
NZ_KB949671.1170375-171823 (+)	**UAA948**	GCF_000396905.1
NC_004668.12889087-2890535 (+)	**V583**	**GCF_000007785.1**
NZ_KE136402.1575409-576857 (-)	**V583**	GCF_000407045.1
NZ_KE136524.1115921-117369 (+)	**V583**	GCF_000407305.1
NZ_KB946433.11865663-1867111 (+)	**V587**	GCF_000394175.1
NZ_CP039296.12889370-2890818 (+)	**VE14089**	GCF_006494835.1
NZ_CP039548.12889370-2890818 (+)	**VE18379**	GCF_006494855.1
NZ_CP039549.12872556-2874004 (+)	**VE18395**	GCF_006494875.1
NZ_VHQT01000145.12304-3752 (+)	**W133**	GCF_006541445.1
NZ_WMGS01000005.1100231-101679 (-)	**W195**	GCF_009734895.1
NZ_VHQO01000003.1132274-133722 (-)	**W97**	GCF_006541435.1

Table S3. *E. faecalis* genomes that encode a protein identical to $MafR_{OF1RF}$. Data from the National Center for Biotechnology Information (03/04/2020)

CDS Region in Nucleotide	Strain	Assembly
NZ_VWNV01000003.139911-41359 (+)	**100AC**	GCF_009830765.1
NZ_QFYO01000001.12063531-2064979 (+)	**11-2011**	GCF_003144755.1
NZ_CP039752.1283658-285106 (-)	**110**	GCF_005154485.1
NZ_KB947327.11566104-1567552 (+)	**12107**	GCF_000395035.1
NZ_QNGN01000009.139059-40507 (+)	**13484**	GCF_003962585.1
NZ_NJIN01000002.121220-22668 (-)	**2070**	GCF_002763385.1
NZ_JVOF01000043.140668-42116 (+)	**254_EFLS**	GCF_001054775.1
NZ_JVOE01000045.16428-7876 (-)	**255_ESPC**	GCF_001056095.1
NZ_CAACXV010000001.11124475-1125923 (+)	**3012STDY6246399**	GCF_900683515.1
NZ_QFYN01000001.1770556-772004 (+)	**46-2014**	GCF_003144715.1
NZ_CABGHP010000003.147757-49205 (+)	**4928STDY7071217**	GCF_902158895.1
NZ_CABGHR010000002.1320769-322217 (-)	**4928STDY7071222**	GCF_902158825.1
NZ_CABGWW010000004.1264913-266361 (-)	**4928STDY7071227**	GCF_902162815.1
NZ_CABGXG010000004.1265053-266501 (-)	**4928STDY7071229**	GCF_902162865.1
NZ_CABGXX010000002.1320369-321817 (-)	**4928STDY7071270**	GCF_902163125.1
NZ_CABGKS010000002.114262-15710 (+)	**4928STDY7071308**	GCF_902159645.1
NZ_CABGHJ010000002.1102478-103926 (+)	**4928STDY7071317**	GCF_902158815.1
NZ_CABGHX010000002.1195112-196560 (-)	**4928STDY7071319**	GCF_902158765.1
NZ_CABGKY010000004.146090-47538 (-)	**4928STDY7071329**	GCF_902159705.1
NZ_CABGLE010000004.146090-47538 (-)	**4928STDY7071330**	GCF_902159755.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABGLO010000003.1271872-273320 (-)	**4928STDY7071336**	GCF_902159845.1
NZ_CABGLG010000003.1272521-273969 (-)	**4928STDY7071338**	GCF_902159825.1
NZ_CABGIK010000002.139065-40513 (+)	**4928STDY7071343**	GCF_902159055.1
NZ_CABGLC010000004.1281667-283115 (-)	**4928STDY7071346**	GCF_902159735.1
NZ_CABGLU010000002.149095-50543 (+)	**4928STDY7071350**	GCF_902159885.1
NZ_CABGIL010000008.148415-49863 (+)	**4928STDY7071351**	GCF_902158995.1
NZ_LR607334.12968082-2969530 (+)	**4928STDY7071355**	GCF_902164645.1
NZ_CABGMB010000003.1269300-270748 (-)	**4928STDY7071369**	GCF_902159935.1
NZ_CABGIV010000003.1271862-273310 (-)	**4928STDY7071372**	GCF_902159155.1
NZ_CABGME010000003.1272521-273969 (-)	**4928STDY7071381**	GCF_902159955.1
NZ_CABGMD010000003.1272628-274076 (-)	**4928STDY7071382**	GCF_902160045.1
NZ_CABGMM010000002.114273-15721 (+)	**4928STDY7071391**	GCF_902160105.1
NZ_CABGMP010000002.1320768-322216 (-)	**4928STDY7071392**	GCF_902160075.1
NZ_CABGMJ010000002.1320772-322220 (-)	**4928STDY7071393**	GCF_902160085.1
NZ_CABGNC010000003.1199555-201003 (-)	**4928STDY7071409**	GCF_902160285.1
NZ_CABGNI010000004.140098-41546 (+)	**4928STDY7071410**	GCF_902160295.1
NZ_CABGJF010000002.1320775-322223 (-)	**4928STDY7071471**	GCF_902159175.1
NZ_CABGPF010000002.114262-15710 (+)	**4928STDY7071475**	GCF_902160825.1
NZ_CABGPC010000002.1320765-322213 (-)	**4928STDY7071476**	GCF_902160795.1
NZ_CABGPZ010000003.1269521-270969 (-)	**4928STDY7071495**	GCF_902160985.1
NZ_CABGPV010000010.148986-50434 (+)	**4928STDY7071496**	GCF_902160965.1
NZ_CABGQH010000003.1268439-269887 (-)	**4928STDY7071516**	GCF_902161125.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABGQL010000003.139339-40787 (+)	**4928STDY7071517**	GCF_902161145.1
NZ_CABGQM010000003.138908-40356 (+)	**4928STDY7071518**	GCF_902161115.1
NZ_CABGJQ010000002.1267021-268469 (-)	**4928STDY7071559**	GCF_902159435.1
NZ_CABGTD010000002.1322340-323788 (-)	**4928STDY7071593**	GCF_902161845.1
NZ_CABGSX010000002.1136007-137455 (+)	**4928STDY7071602**	GCF_902161785.1
NZ_CABGTH010000006.139077-40525 (+)	**4928STDY7071605**	GCF_902161825.1
NZ_CABGTS010000004.139057-40505 (+)	**4928STDY7071616**	GCF_902162075.1
NZ_CABGJR010000004.139057-40505 (+)	**4928STDY7071617**	GCF_902159335.1
NZ_CABGTT010000002.1320576-322024 (-)	**4928STDY7071624**	GCF_902162025.1
NZ_CABGUB010000002.114203-15651 (+)	**4928STDY7071626**	GCF_902161975.1
NZ_CABHEC010000002.139065-40513 (+)	**4928STDY7071641**	GCF_902164755.1
NZ_CABHDZ010000002.140022-41470 (+)	**4928STDY7071642**	GCF_902164735.1
NZ_CABGZN010000003.149086-50534 (+)	**4928STDY7071712**	GCF_902163505.1
NZ_CABHAB010000003.1286888-288336 (-)	**4928STDY7071729**	GCF_902163655.1
NZ_CABHAH010000003.1285899-287347 (-)	**4928STDY7071730**	GCF_902163695.1
NZ_CABHAL010000003.1232241-233689 (-)	**4928STDY7071733**	GCF_902163705.1
NZ_CABHAN010000003.1232241-233689 (-)	**4928STDY7071734**	GCF_902163715.1
NZ_CABHAR010000002.114262-15710 (+)	**4928STDY7071743**	GCF_902163755.1
NZ_CABHAU010000003.1232241-233689 (-)	**4928STDY7071751**	GCF_902163885.1
NZ_CABHBD010000003.1272662-274110 (-)	**4928STDY7071752**	GCF_902163925.1
NZ_CABHBJ010000003.139613-41061 (+)	**4928STDY7071767**	GCF_902163945.1
NZ_CABHBT010000003.1269190-270638 (-)	**4928STDY7071768**	GCF_902164125.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CABHCJ010000002.1129867-131315 (+)	**4928STDY7387692**	GCF_902164245.1
NZ_CABHCQ010000003.1271758-273206 (-)	**4928STDY7387693**	GCF_902164315.1
NZ_CABHDT010000002.1320781-322229 (-)	**4928STDY7387732**	GCF_902164585.1
NZ_CABHEN010000008.148280-49728 (+)	**4928STDY7387767**	GCF_902164945.1
NZ_CABHKE010000003.1269292-270740 (-)	**4928STDY7387795**	GCF_902166505.1
NZ_CABHHZ010000004.1173045-174493 (-)	**4928STDY7387858**	GCF_902165815.1
NZ_CABHIT010000002.114385-15833 (+)	**4928STDY7387901**	GCF_902165985.1
NZ_KB947499.1116890-118338 (+)	**5952**	GCF_000395405.1
NZ_SEWV01000010.134361-35809 (-)	**612T**	GCF_004168115.1
NZ_JUVP01000106.148218-49666 (+)	**724_EFLS**	GCF_001058995.1
NZ_VWOG01000006.159759-61207 (+)	**8-2**	GCF_009830895.1
NZ_VWNL01000004.111851-13299 (-)	**8-3**	GCF_009830535.1
NZ_QFYM01000003.12136080-2137528 (-)	**90-2015**	GCF_003144735.1
NZ_KB945040.1129056-130504 (+)	**A-2-1**	GCF_000393615.1
NZ_PGCX01000057.139851-41299 (+)	**A113R1B0**	GCF_003933395.1
NZ_MTFY01000009.121237-22685 (+)	**ATCC 29212**	GCF_001999625.1
NZ_NAQY01000006.179683-81131 (+)	**ATCC BAA-2128**	GCF_002088065.1
NZ_WMGE01000010.139015-40463 (+)	**B168**	GCF_009735135.1
NZ_SDQZ01000008.171095-72543 (-)	**C34**	GCF_011009735.1
NZ_BDEN01000003.12589022-2590470 (+)	**CBA7120**	GCF_001662265.1
NZ_SIJP01000003.139904-41352 (+)	**CECT 7121**	GCF_004306085.1
NZ_KB946671.11644822-1646270 (+)	**Com 6**	GCF_000394395.1

CDS Region in Nucleotide	Strain	Assembly
NZ_PUBU01000003.123270-24718 (-)	**CVM N52454**	GCF_002947155.1
NZ_PUBQ01000041.122844-24292 (-)	**CVM N52564**	GCF_002947055.1
NZ_PTYM01000033.119107-20555 (-)	**CVM N53434**	GCF_002948495.1
NZ_PTZN01000060.112635-14083 (-)	**CVM N54156**	GCF_002948095.1
NZ_PUAO01000031.114020-15468 (+)	**CVM N55240**	GCF_002947555.1
NZ_PTWR01000039.111518-12966 (-)	**CVM N59509F**	GCF_002945715.1
NZ_PTWK01000026.119588-21036 (+)	**CVM N59550F**	GCF_002945625.1
NZ_PTWJ01000046.112092-13540 (+)	**CVM N59554F**	GCF_002945655.1
NZ_PTUR01000054.112827-14275 (-)	**CVM N59836F**	GCF_002944875.1
NZ_PTVP01000041.134260-35708 (-)	**CVM N59860F**	GCF_002945295.1
NZ_PTVF01000021.122924-24372 (-)	**CVM N59910F**	GCF_002944985.1
NZ_PTUX01000061.119652-21100 (+)	**CVM N60034F**	GCF_002943945.1
NZ_PTTZ01000058.125414-26862 (-)	**CVM N60268F**	GCF_002944055.1
NZ_GG692884.1167619-169067 (+)	**DS5**	GCF_000157235.1
NZ_JAAQVX010000001.12603116-2604564 (+)	**E.fA105**	GCF_011754415.1
NZ_PJWO01000009.154056-55504 (-)	**EN782**	GCF_004125695.1
NZ_KI912939.12338250-2339698 (+)	**EnGen0401**	GCF_000519445.1
NZ_KI913017.1104421-105869 (+)	**EnGen0413**	GCF_000519685.1
NZ_KI913040.11625825-1627273 (+)	**EnGen0417**	GCF_000519765.1
NZ_KI913075.1352522-353970 (+)	**EnGen0422**	GCF_000519865.1
NZ_KI913088.1340358-341806 (-)	**EnGen0425**	GCF_000519925.1

CDS Region in Nucleotide	Strain	Assembly
NZ_CP041012.1182450-183898 (+)	**FDAARGOS_611**	GCF_006364815.1
NZ_VHRY01000005.186063-87511 (-)	**G138E**	GCF_006541335.1
NZ_VHRW01000004.139521-40969 (+)	**G42**	GCF_006541405.1
NZ_FPDW01000010.149101-50549 (+)	**G824**	GCF_900117305.1
NZ_VHRP01000009.169481-70929 (-)	**H136**	GCF_006541265.1
NZ_VHRK01000002.155027-56475 (+)	**H96E**	GCF_006541255.1
NZ_APWU01000001.1111589-113037 (+)	**KACC 91532**	GCF_004802475.1
NZ_CP042216.12482658-2484106 (+)	**L8**	GCF_009498175.1
NZ_RKMR01000005.139059-40507 (+)	**LIT13 A36**	GCF_003795675.1
NZ_RKMQ01000005.1255415-256863 (-)	**LIT13 A37'**	GCF_003795665.1
NZ_RKMY01000004.1113404-114852 (+)	**LIT5-A37**	GCF_003796705.1
NZ_JPTY01000029.1172755-174203 (-)	**LRS29212**	GCF_000763355.1
NZ_ANMP01000023.139379-40827 (+)	**MA1**	GCF_000524625.1
NZ_JPTZ01000032.1196742-198190 (-)	**MA2**	GCF_000763435.1
NZ_QYLT01000019.148231-49679 (+)	**N010.J-3**	GCF_008083105.1
NZ_QYMR01000009.156971-58419 (-)	**N023.D-17**	GCF_008082705.1
NZ_UGII01000001.12603116-2604564 (+)	**NCTC12697**	GCF_900447845.1
NZ_UGIU01000002.12572418-2573866 (+)	**NCTC8131**	GCF_900448285.1
NZ_UGIG01000001.12473394-2474842 (+)	**NCTC8132**	GCF_900447855.1
NZ_CABEIP010000001.12493339-2494787 (+)	**NCTC8727**	GCF_901543465.1
NZ_CABEIN010000002.12620427-2621875 (+)	**NCTC8734**	GCF_901543485.1
NZ_LR134312.12405340-2406788 (+)	**NCTC8745**	GCF_900636775.1

CDS Region in Nucleotide	Strain	Assembly
NC_017316.12421605-2423053 (+)	**OG1RF**	**GCF_000172575.2**
NZ_CP025021.12421584-2423032 (+)	**OG1RF-SagA**	GCF_004006595.1
NZ_RKOH01000005.139059-40507 (+)	**P16 CL A21**	GCF_003796245.1
NZ_RKNN01000005.139059-40507 (+)	**P3 CL A35**	GCF_003795925.1
NZ_RKOR01000016.139058-40506 (+)	**P7 C A21**	GCF_003796375.1
NZ_RKPY01000016.139058-40506 (+)	**P7 C A7**	GCF_003797045.1
NZ_KB947339.1142262-143710 (+)	**Pan7**	GCF_000395075.1
NZ_VHRA01000019.139078-40526 (+)	**R50**	GCF_006541785.1
NZ_KB944830.175734-77182 (+)	**RMC65**	GCF_000393315.1
NZ_KE351838.146657-48105 (+)	**RP2S-4**	GCF_000415245.2
NZ_NBDR01000045.111517-12965 (-)	**S12**	GCF_002110425.1
NZ_KB947335.1102214-103662 (+)	**SF24396**	GCF_000395055.1
NZ_KB946410.1107387-108835 (+)	**SS-6**	GCF_000394135.1
NZ_MCFV01000003.1913511-914959 (+)	**ST4:12**	GCF_001878785.1
NZ_GG670346.1269378-270826 (-)	**T1**	GCF_000157135.1
NZ_KB944675.1126077-127525 (+)	**T10**	GCF_000393055.1
NZ_KB947308.11579131-1580579 (+)	**T13**	GCF_000394995.1
NZ_KB944664.1107780-109228 (+)	**T17**	GCF_000392995.1
NZ_KB944587.1106202-107650 (+)	**T20**	GCF_000392855.1
NZ_KB947401.1111919-113367 (+)	**T21**	GCF_000395135.1
NZ_KB945036.1127774-129222 (+)	**T4**	GCF_000393595.1
NZ_KB944671.11673320-1674768 (+)	**T9**	GCF_000393035.1

CDS Region in Nucleotide	Strain	Assembly
NZ_GL454360.147700-49148 (+)	**TX0027**	GCF_000147535.1
NZ_GL454747.140555-42003 (+)	**TX0031**	GCF_000147275.1
NZ_GL455107.112529-13977 (-)	**TX0102**	GCF_000147215.1
NZ_GL456756.112525-13973 (-)	**TX0312**	GCF_000148305.1
NZ_GL456550.1143149-144597 (-)	**TX0470**	GCF_000148265.1
NZ_GL455563.148022-49470 (+)	**TX2134**	GCF_000148065.1
NZ_LYBN01000013.1170855-172303 (-)	**UCD-PD3**	GCF_001692955.1
NZ_VHQP01000017.168602-70050 (-)	**W84**	GCF_006541885.1
NZ_JXMK01000014.154431-55879 (-)	**ZHOU_T10_3**	GCF_002105335.1
NZ_JXML01000011.154009-55457 (-)	**ZHOU_T60_1**	GCF_002105365.1

Table S4. Eukaryotic and prokaryotic proteins that have sequence homology to the enterococcal OG1RF_12294 protein (850 amino acids, GenBank AEA94981.1)

Protein (Accession number)	Organism	Number of amino acids	Identity	Similarity
PMR1 (NP_011348.1)	*Saccharomyces cerevisiae* S288C	950	32.5 %	52.5 %
PMR1 (CAC19896.1)	*Caenorhabditis elegans*	901	33.0 %	52.2 %
OG1RF_10600 (AEA93287.1)	*Enterococcus faecalis* OG1RF	881	36.0 %	53.1 %
OG1RF_11602 (AEA94289.1)	*Enterococcus faecalis* OG1RF	901	39.7 %	56.9 %
YloB (NP_389448.1)	*Bacillus subtils* 168	890	36.7 %	56.4 %
CaxP (WP_000032453.1)	*Streptococcus pneumoniae* D39	898	37.4 %	56.1 %
LMCA1 (CAC98919.1)	*Listeria monocytogenes* EGD-e	880	35.9 %	56.6 %
Lmo0818 (NP_464345.1)	*Listeria monocytogenes* EGD-e	876	34.8 %	53.4 %
PMA1 (WP_010872526.1)	*Synechocystis sp.* PCC 6803	905	36.3 %	53.5 %

Accesion numbers were obtained from NCBI (National Center for biotechnology Information).

Identity and similarity are given according to EMBOSS Needle (Pairwise Sequence Alignment) (Rice *et al.*, 2000).

Table S5. *S. pneumoniae* genomes that encode a protein identical to MgaP$_{R6}$. Data from the National Center for Biotechnology Information (25/04/2020)

CDS Region in Nucleotide	Strain	Assembly
NC_003098.11388136-1389620 (+)	**R6**	**GCF_000007045.1**
NC_008533.21395668-1397152 (+)	**D39**	GCF_000014365.2
NZ_CFDT01000031.118070-19554 (-)	**SMRU451**	GCF_001334855.1
NZ_CKNI01000007.136609-38093 (+)	**SMRU226**	GCF_001126545.1
NZ_CP027540.11396124-1397608 (+)	**D39V**	GCF_003003495.1
NZ_JVRJ01000055.136793-38277 (+)	**172_SPNE**	GCF_001071595.1
NZ_LR216015.11453682-1455166 (+)	**GPSC11**	GCF_900692585.1
NZ_LR536843.11286350-1287834 (+)	**GPSC55**	GCF_900795205.1
NZ_SJSZ01000010.132543-34027 (-)	**BC18042556**	GCF_004359405.1

RELATED PUBLICATIONS

FEBS Letters

FEBS PRESS science publishing by scientists

DNA-binding properties of MafR, a global regulator of *Enterococcus faecalis*

Sofía Ruiz-Cruz, Ana Moreno-Blanco, Manuel Espinosa and Alicia Bravo

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Correspondence
A. Bravo, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, E-28040 Madrid, Spain
Fax: +34 915360432
Tel: +34 918373112
E-mail: abravo@cib.csic.es

(Received 2 February 2018, revised 2 March 2018, accepted 5 March 2018, available online 26 March 2018)

doi:10.1002/1873-3468.13032

Edited by Renee Tsolis

Global transcriptional regulators play key roles during bacterial adaptation to environmental fluctuations. Protein MafR from *Enterococcus faecalis* was shown to activate the transcription of many genes on a genome-wide scale. We proposed that MafR is a global regulator of the Mga/AtxA family. Here, we purified an untagged form of the MafR protein and found that it binds to linear double-stranded DNAs in a nonsequence-specific manner. Moreover, multiple MafR units (likely dimers) bind sequentially to the DNA molecule generating multimeric complexes. On DNAs that contain the promoter of the *mafR* gene, MafR recognizes a potentially curved DNA region. We discuss that a characteristic of the Mga/AtxA regulators might be their ability to recognize particular DNA shapes across the bacterial genomes.

Keywords: *Enterococcus faecalis*; gene expression; global regulators; protein–DNA interactions

Bacterial adaptation to a new niche usually requires global changes in gene expression. Many of these changes are coordinated by proteins that function as global transcriptional regulators. The ability of such proteins to recognize multiple DNA sites across the bacterial genome makes possible to adjust the gene expression pattern in response to environmental fluctuations. Several findings from structural and biochemical studies have shown that simple protein–DNA recognition mechanisms do not exist [1]. Rohs *et al*. [2] classified two main readout mechanisms: base readout and shape readout. In the base readout mechanism, proteins recognize the unique chemical signatures of the DNA bases. In contrast, in the shape readout mechanism, proteins recognize a sequence-dependent DNA shape. Nevertheless, based on the structures of numerous protein–DNA complexes, it has been reasoned that particular proteins use likely a combination of readout mechanisms to achieve DNA-binding specificity [2].

The gram-positive bacterium *Enterococcus faecalis* is able to colonize different niches of the human host. It is generally found as a harmless commensal of the gastrointestinal tract. However, in immunocompromised hosts, *E. faecalis* can cause a variety of infections, such as urinary tract infections, endocarditis and bacteraemia [3–5]. Our knowledge of the regulatory elements involved in the adaptation of *E. faecalis* to particular host niches is still very limited. Genome-wide microarray assays designed for the *E. faecalis* strain OG1RF showed that protein MafR activates, directly or indirectly, the expression of numerous genes [6]. Some of them were found to be up-regulated during the growth of *E. faecalis* in blood and/or in human urine [7,8]. MafR (482 amino acids) has sequence similarity to three regulatory proteins: AtxA (40.7%; 475 amino acids) from *Bacillus anthracis*, Mga*Spn* (38.8%; 493 amino acids) from *Streptococcus pneumoniae*, and Mga (31.3%; 530 amino acids) from *S. pyogenes* [6]. The three proteins are members of an emerging class of

Abbreviations
dsDNA, double-stranded DNA; EMSA, electrophoretic mobility shift assay; OD, optical density.

FEBS Letters **592** (2018) 1412–1425 © 2018 The Authors. *FEBS Letters* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

global transcriptional regulators involved in virulence (the Mga/AtxA family) [9–11]. Furthermore, according to the Pfam database of protein families [12], MafR has two putative DNA-binding domains within the N-terminal region, the so-called HTH_Mga (Family PF08280, residues 11–69) and Mga (Family PF05043, residues 76–164) domains, which are also present in Mga, AtxA and Mga*Spn* [6,13,14]. Based on these findings, we proposed that MafR is a potential regulator of the Mga/AtxA family. However, the DNA-binding properties of MafR remain to be investigated.

In vitro protein–DNA interaction studies have been reported for Mga and Mga*Spn*, but not for AtxA. During exponential growth of *S. pyogenes*, Mga activates directly the transcription of several virulence genes. Some of them encode factors important for adherence to host tissues and for evasion of the host immune responses [15]. *In vitro* experiments showed that a His-tagged Mga protein binds to regions located upstream of the target promoters. The position of the Mga-binding sites with respect to the start of transcription differs among the promoters tested [13,16]. Although a consensus Mga-binding sequence was initially proposed [17], subsequent sequence alignments revealed that the sites recognized by Mga exhibit a low sequence identity (13.4%) [16]. Additionally, it has been shown that the His-tagged Mga protein is able to form higher order oligomers in solution [18].

The pneumococcal Mga*Spn* protein activates directly the transcription of a four-gene operon (*spr1623–spr1626*) of unknown function. This activation requires a region located upstream of the target promoter [14]. *In vitro* experiments showed that Mga*Spn* binds to linear double-stranded (ds) DNAs with little or no sequence specificity. Moreover, Mga*Spn* is able to generate multimeric complexes on linear DNAs [10]. Additional results supported that Mga*Spn* recognizes structural features in its target DNA as follows: (a) Mga*Spn* binds preferentially to DNA sites that contain a potential intrinsic curvature flanked by regions of bendability, (b) Mga*Spn* has a high affinity for a naturally occurring curved DNA, and (c) Mga*Spn* has a preference for AT-rich DNA sites [10,19]. Because of these results, we proposed that a preference for particular DNA structures rather than for specific DNA sequences might be a general feature of the global regulators that constitute the Mga/AtxA family [19]. In agreement with this hypothesis, sequence similarities in the promoter regions of the genes regulated by AtxA are not apparent, and some of those promoter regions are intrinsically curved [20].

In the present work, we purified an untagged form of the MafR protein and analyzed its DNA-binding properties. By gel retardation assays, we found that MafR binds to linear dsDNAs in a nonsequence-specific manner. Multiple units of MafR (likely dimers) bind orderly on the same DNA molecule generating multimeric complexes. Moreover, by footprinting experiments, we found that MafR binds to a potentially curved DNA region. Our results support that recognition of sequence-dependent DNA shapes might be a hallmark of the global regulators that belong to the Mga/AtxA family.

Materials and methods

Oligonucleotides, bacterial strains, and plasmids

Oligonucleotides used in this work are listed in Table 1. Chromosomal DNA was isolated from *E. faecalis* V583 [21], *E. faecalis* OG1RF [22], and *S. pneumoniae* R6 [23]. *E. faecalis* JH2-2 [24] was used as a host for plasmids based on pDLF [6] and pDLS. The expression vector pDLS is a pDL287 [25] derivative. It has a unique restriction site for *Sph*I downstream of the pneumococcal *PsulA* promoter [26]. For its construction, a 202-bp region (promoter *PsulA*) of the R6 genome was amplified by PCR using primers *pSul*F and *pSul*R. The PCR product was digested with *Cla*I, and the 181-bp restriction fragment was ligated to *Cla*I-linearized pDL287. Plasmids pDLF*mafR*$_{V583}$ and pDLS*mafR*$_{V583}$ carry the *P2493::mafR*$_{V583}$ and *PsulA::mafR*$_{V583}$ fusion genes, respectively. For their construction, a 1546-bp region of the V583 chromosome was amplified using the *mafSph*F and *mafSph*R primers. After *Sph*I digestion, the 1514-bp restriction fragment was inserted into the *Sph*I site of both expression vectors pDLF and pDLS.

For protein overproduction, an inducible expression system based on *Escherichia coli* BL21 (DE3) (a gift of F. W. Studier) and pET24b (Novagen) was used. To overproduce MafR$_{V583}$, a 1502-bp region of the V583 chromosome was amplified by PCR using the Up*mafR* and Dw*mafR* primers. These primers have a single restriction site for *Nde*I and *Xho*I, respectively. The amplified product was digested with both enzymes, and the 1470-bp digestion product was inserted into pET24b (plasmid pET24b-*mafR*$_{V583}$). To overproduce MafR$_{V583}$-His and MafR$_{OG1RF}$-His, a 1481-bp DNA region was amplified by PCR using the Up*mafR* and Dw*mafR*-His primers. The amplified DNA was digested with *Nde*I and *Xho*I, and the 1448-bp digestion product was inserted into pET24b (plasmids pET24b-*mafR*$_{V583}$-His and pET24b-*mafR*$_{OG1RF}$-His). To overproduce MafR$_{OG1RF}$Δ3N-His, a 1472-bp region of the OG1RF chromosome was amplified by PCR using the Up*mafR*-Δ3N and Dw*mafR*-His oligonucleotides. The 1439-bp digestion product was inserted into pET24b (plasmid pET24b-*mafR*$_{OG1RF}$Δ3N-His).

Table 1. Oligonucleotides used in this work.

Name	Sequence (5′ to 3′)[a]
*pSul*F	TGTTAATGGG**ATCGAT**TTCTGTTTG
*pSul*R	GACAT**ATCGAT**CACTCCC**GCATGC**ATTTTCATC
*mafSph*F	TTTTTATCCGTATTC**GCATGC**AAAAGGAGG
*mafSph*R	AACCAAACGAT**GCATGC**CGAAAGAAAGC
Up*mafR*	GCAAAAGGAGGTTTT**CATATG**TACTCCATG
Dw*mafR*	AGCCAAAAAA**CTCGAG**AATGTCCTCGCTAG
Dw*mafR*-His	CCTCGCTAGTT**CTCGAG**AAAATAAGAATGA
Up*mafR*-Δ3N	GGTTTTGCCATGTAC**CATATG**TTAAAACGT
3012A	AGGAATGGCTGTTGTAACCA
3012B	AGTGCGGCTCCTGTCGGTAA
3013A	AACAAACGAATTTGCCGAAGC
3013B	CAACTGTTCCAACAAACG
3013C	CCGTTATCACACGTTTTAACA
0091G2	GGCTATTTTGATGCACATATCTG
0092A2	CCCGCCTTCCTTCCCTTGCTC
EM1	AGTTGAATGTTTAAAGAAATGATGG
26A	TTCTTTGTGGTATAATTGCAAGAGGT
26B	ACCTCTTGCAATTATACCACAAAGAA
20A	TATATTGTCTCCGTAGTGTT
20B	AACACTACGGAGACAATATA
40A	TATATCATGCTATACCTATTCTTTGTGGTATAATTGCAAG
40B	CTTGCAATTATACCACAAAGAATAGGTATAGCATGATATA
32A	TTCTTTGTGGTATAATTGCAAGAGGTTTAATC
32B	GATTAAACCTCTTGCAATTATACCACAAAGAA

[a]Restriction sites are in bold, and the base changes that generate restriction sites are underlined.

DNA isolation

For small-scale preparations of plasmid DNA, the High Pure Plasmid Isolation Kit (Roche Applied Science, Penzberg, Germany) was used with the modifications reported for *Enterococcus* [6]. Chromosomal DNA was isolated from *E. faecalis* and *S. pneumoniae* as previously described [26].

Growth and transformation of bacteria

The *E. faecalis* was grown in Bacto™ Brain Heart Infusion (BHI) medium, which was supplemented with kanamycin (250 $\mu g \cdot mL^{-1}$) when the cells harbored a plasmid based on pDLF or pDLS. *E. coli* cells carrying a derivative of pET24b were grown in tryptone-yeast extract (TY) medium supplemented with kanamycin (30 $\mu g \cdot mL^{-1}$). Bacteria were grown at 37 °C. The protocols used to transform *E. faecalis* and *E. coli* by electroporation have been described [27,28].

Western blot

Plasmid-carrying enterococcal cells were grown as indicated above to an optical density at 650 nm (OD_{650}) of 0.3 (exponential phase). To prepare whole-cell extracts, cells were concentrated 40-fold in buffer LBW (25 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 0.2 $mg \cdot mL^{-1}$ lysozyme, 260 $units \cdot mL^{-1}$ mutanolysin), and incubated at 37 °C for 10 min. Total proteins were separated by SDS/PAGE (10%). Prestained proteins (Invitrogen, Waltham, MA, USA) were run in the same gel as molecular weight markers. Proteins were transferred electrophoretically to immunoblot polyvinylidene difluoride (PVDF) membranes (Bio-Rad) using a Mini Trans Blot (Bio-Rad, Hercules, CA, USA) as reported previously [26]. Membranes were probed with rabbit polyclonal antibodies against $MafR_{V583}$. Antigen–antibody complexes were detected using antirabbit horseradish peroxidase-conjugated antibodies, the Immun-Star™ HRP substrate kit (Bio-Rad), and the Luminescent Image Analyzer LAS-3000 (Fujifilm Life Science). Rabbit polyclonal antibodies against $MafR_{V583}$ were produced in the Animal Facility of the Centro de Investigaciones Biológicas, CSIC (Madrid, Spain). For the immunizations, purified protein ($MafR_{V583}$) and traditional protocols (Freund′s complete adjuvant, Freund′s incomplete adjuvant, subcutaneous administration) were used.

Polymerase chain reaction (PCR)

The Phusion High-Fidelity DNA polymerase (Thermo Scientific, Waltham, MA, USA) and the Phusion HF buffer were used. Reaction mixtures (50 μL) contained 5–30 ng of template DNA, 20 pmol of each primer, 200 μM each dNTP, and one unit of DNA polymerase. PCR conditions were reported previously [26]. PCR products were purified

with the QIAquick PCR purification kit (QIAGEN, Hilden, Germany).

PCR amplification of chromosomal DNA regions

By PCR, four regions of the V583 chromosome were amplified: (a) a 217-bp region (coordinates 2888932–2889148) using the 3012A and 3013A primers, (b) a 227-bp region (2888932–2889158) using the 3012A and 3013B primers, (c) a 260-bp region (coordinates 2888858–2889117) using the 3012B and 3013C primers, and (d) a 321-bp region (coordinates 94488–94808) using the 0091G2 and 0092A2 primers. From the *S. pneumoniae* R6 chromosome, a 322-bp region (coordinates 1598010–1598331) was amplified using the 26A and EM1 primers.

Annealing of complementary oligonucleotides

Oligonucleotides 20A, 26A, 32A, 40A and their complementary oligonucleotides (20B, 26B, 32B, 40B) (Table 1) were used to generate small dsDNA fragments. Equimolar amounts of complementary oligonucleotides were annealed in buffer TE (2 mM Tris-HCl, pH 8.0, 0.2 mM EDTA) containing 50 mM NaCl. Reaction mixtures (150 μL) were incubated at 95 °C for 10 min, cooled down slowly to 37 °C, kept at 37 °C for 10 min, and placed on ice for 10 min.

Overproduction and purification of proteins

E. coli BL21 (DE3) cells harboring a pET24b derivative were grown as indicated above to an OD_{600} of 0.45. Then, IPTG (1 mM) was added. After 25 min, cells were treated with rifampicin (200 $\mu g \cdot mL^{-1}$) for 60 min. Cells were then collected by centrifugation, washed twice with buffer VL (50 mM Tris-HCl, pH 7.6, 5% glycerol, 1 mM DTT, 1 mM EDTA) containing 200 mM NaCl, and stored at −80 °C.

To purify $MafR_{V583}$, bacterial cells were concentrated 40-fold in buffer VL containing 200 mM NaCl and a protease inhibitor cocktail (Roche). To purify His-tagged proteins, an EDTA-free protease inhibitor cocktail (Roche) was used. Cells were lysed using a prechilled French pressure cell, and the whole-cell extract was centrifuged to remove cell debris. The clarified extract was mixed with 0.2% poly(ethyleneimine), kept on ice for 30 min, and centrifuged at 10 414 ***g*** in an Eppendorf F-34-6-38 rotor for 20 min at 4 °C. $MafR_{V583}$, as well as the His-tagged variants, were recovered in the poly(ethyleneimine) pellet, which was subsequently washed twice with buffer VL containing 200 mM NaCl. $MafR_{V583}$ and the His-tagged variants were eluted with buffer VL containing 500 mM NaCl. Proteins were precipitated with ammonium sulfate (70% saturation; 60 min on ice), followed by a centrifugation step (10 414 ***g*** for 20 min at 4 °C).

In the case of $MafR_{V583}$, the ammonium sulfate precipitate was dissolved in buffer VL containing 200 mM NaCl, and dialyzed against the same buffer at 4 °C. The protein preparation was loaded onto a heparin affinity column (HiPrep Heparin, GE Healthcare, Chicago, IL, USA). To elute $MafR_{V583}$, a linear gradient of NaCl (200–600 mM) in buffer VL was used. Fractions containing $MafR_{V583}$ were identified by Coomassie-stained SDS-polyacrylamide (12%) gels, pooled, and dialyzed against buffer VL containing 200 mM NaCl. Then, the protein preparation was concentrated (Vivaspin-20, GE Healthcare), applied to a HiLoad Superdex 200 gel filtration column, and subjected to fast-pressure liquid chromatography (Biologic Duoflow, Bio-Rad) at 4 °C. The running buffer contained 200 mM NaCl. Protein fractions containing $MafR_{V583}$ were pooled, concentrated and stored at −80 °C.

In the case of His-tagged proteins, the ammonium sulfate precipitate was dissolved in buffer S-His (10 mM Tris-HCl, pH 7.6, 5% glycerol, 300 mM NaCl, 1 mM DTT), and dialyzed against the same buffer at 4 °C. Imidazole (10 mM) was added to the protein preparation, which was then loaded onto a nickel affinity column (HisTrap HP column, GE Healthcare). To elute the His-tagged protein, a linear gradient of imidazole (10–250 mM) in buffer S-His was used. Fractions containing the His-tagged protein were pooled, dialyzed against buffer VL containing 200 mM NaCl, concentrated, and stored at −80 °C.

Protein concentration was determined using a NanoDrop ND-2000 Spectrophotometer (Thermo Scientific). $MafR_{V583}$, $MafR_{OG1RF}$-His and $MafR_{OG1RF}\Delta 3N$-His were subjected to amino terminal sequencing by Edman degradation using a Procise 494 Sequencer (Perkin Elmer, Waltham, MA, USA).

Gel filtration chromatography

Protein $MafR_{V583}$ was injected into a HiLoad Superdex 200 gel filtration column using a Biologic Duoflow system (Bio-Rad). Buffer VL containing 200 mM NaCl was used to equilibrate the column and as running buffer. The column was calibrated with various proteins of known Stokes radius: alcohol dehydrogenase (ADH; 45 Å), albumin (A; 35.5 Å), ovalbumin (O; 30.5 Å), and carbonic anhydrase (CA; 20.1 Å). Elution positions were monitored at 280 nm. The K_{av} value was calculated as $(V_e - V_o)/(V_t - V_o)$, where V_e is the elution volume, V_o is the void volume (determined by elution of blue dextran), and V_t is the total volume of the packed bed. Data were plotted according to Siegel and Monty [29].

Analytical ultracentrifugation

Experiments were performed at 12 °C in an Optima XL-I analytical ultracentrifuge (Beckman-Coulter, Brea, CA, USA) equipped with an UV-visible optical detection system, using an An50Ti rotor and Epon-charcoal standard

double sector centerpieces (12 mm optical path). Sedimentation velocity assays were carried out at 185 795 ***g*** $MafR_{V583}$ (5 and 10 μM; 350 μL) was equilibrated in buffer AU (50 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.1 mM DTT; 3% glycerol, and 200 mM NaCl). The $MafR_{V583}$ sedimentation coefficient was estimated applying a direct linear least squares boundary modeling of the sedimentation velocity data using the SEDFIT program (version 12.0) [30]. The sedimentation coefficient was corrected to standard conditions to obtain the corresponding $S_{20,W}$ value using the SEDNTERP program [31]. The translational frictional coefficient (f) of $MafR_{V583}$ was determined from the molecular mass and sedimentation coefficient values [32], whereas the frictional coefficient of the equivalent hydrated sphere (f_0) was estimated using a hydration of 0.37 g H_20/g protein [33]. With these parameters the translational frictional ratio (f/f_0) was calculated ($MafR_{V583}$ hydrodynamic shape). Sedimentation equilibrium assays were performed at two protein concentrations (5 and 10 μM). Samples (90 μL) were centrifuged at two successive speeds (5161 and 8064 ***g***) and absorbance readings were done after the sedimentation equilibrium was reached. The absorbance scans were taken at 280 and 291 nm, depending on $MafR_{V583}$ concentration. In all cases, the baseline signals were measured after high-speed centrifugation (129 024 ***g***). Apparent average molecular masses of $MafR_{V583}$ were determined using the HETEROANALYSIS program [34]. The partial specific volume of $MafR_{V583}$ was 0.742 $mL{\cdot}g^{-1}$, calculated from the amino acid composition with the SEDNTERP program [31].

Radioactive labeling of DNA fragments

Oligonucleotides were radioactively labeled at the 5′ end using [γ-^{32}P]-ATP (3000 Ci/mmol; Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs) as described [10]. The 5′ labeled oligonucleotides were used for PCR amplification (labeling at either the coding or the noncoding strand).

Electrophoretic mobility shift assays

In general, binding reactions (10–20 μL) contained either nonlabeled DNA (10 nM) or ^{32}P-labeled DNA (1–2 nM) and different amounts of the purified protein. The binding buffer contained 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50 mM NaCl, and 0.5 $mg{\cdot}mL^{-1}$ BSA. When indicated, nonlabeled competitor calf thymus DNA and ^{32}P-labeled DNA were added simultaneously to the binding reaction. Reaction mixtures were incubated at room temperature for 20 min. Free and bound DNA forms were separated by electrophoresis on native polyacrylamide (6% or 8%) gels (Mini-PROTEAN system, Bio-Rad). Gels were pre-electrophoresed (20 min) and run at 100 V and room temperature. Labeled DNA was visualized using a Fujifilm Image Analyzer FLA-3000.

DNase I footprinting assays

Binding reactions (50 μL) contained 2 nM of ^{32}P-labeled DNA, 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50 mM NaCl, 0.5 $mg{\cdot}mL^{-1}$ BSA, 1 mM $CaCl_2$, 10 mM $MgCl_2$, and different concentrations of $MafR_{V583}$. Reaction mixtures were incubated at room temperature for 20 min. Then, DNA was digested using 0.04 units of DNase I (Roche Applied Science). After 5 min at room temperature, reactions were stopped by adding 25 μL of Stop DNase I buffer (2 M ammonium acetate; 0.8 mM sodium acetate, 0.15 M EDTA). DNA was precipitated with ethanol, dried and dissolved in 5 μL of loading buffer (80% formamide, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol, 1 mM EDTA). After heating at 95 °C for 5 min, samples were loaded onto 8 M urea-6% polyacrylamide gels. Dideoxy-mediated chain termination sequencing reactions were run in the same gel. Labeled products were visualized using a Fujifilm Image Analyzer FLA-3000. The intensity of the bands was quantified using the Quantity One software (Bio-Rad).

In silico prediction of intrinsic curvature

The bendability/curvature propensity plots were calculated with the bend.it server [35] (http://hydra.icgeb.trieste.it/dna/bend_it.html) as described previously [10].

Results

Purification of an untagged form of the $MafR_{V583}$ protein

To overproduce and then purify an untagged form of the $MafR_{V583}$ protein, we used a heterologous system based on the *E. coli* strain BL21 (DE3) and the inducible expression vector pET24b. The procedure used to purify $MafR_{V583}$ involved basically four steps: (a) precipitation of DNA and $MafR_{V583}$ with poly(ethyleneimine) using a low ionic strength buffer; (b) elution of $MafR_{V583}$ from the poly(ethyleneimine) pellet using a higher ionic strength buffer; (c) fractionation of proteins by heparin chromatography; and (d) fractionation of proteins by gel filtration chromatography (Fig. 1A). Purified $MafR_{V583}$ was analyzed by SDS/PAGE (12%). It migrated between standard proteins of 45 and 66 kDa, which is consistent with the theoretical molecular weight of the $MafR_{V583}$ monomer (56 247 Da). Further determination of the N-terminal amino acid sequence (eight residues) of $MafR_{V583}$ by

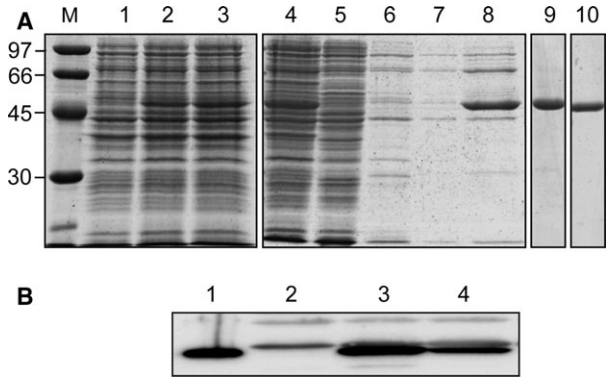


Fig. 1. (A) Purification of $MafR_{V583}$. Proteins were analyzed by SDS/PAGE (12%). Gels were stained with Coomassie Blue. Lane M: molecular weight standards (in kDa) were run in the same gel (LMW Marker, GE Healthcare). Lane 1: total proteins from cells nontreated with IPTG. Lane 2: total proteins from cells treated with IPTG for 25 min. Lane 3: total proteins from cells treated with IPTG (25 min) and then with rifampicin (60 min). Purification steps (lanes 4–10): cleared cell lysate (lane 4); supernatant after DNA precipitation with poly(ethyleneimine) in the presence of 200 mM NaCl (lane 5); proteins eluted from the poly(ethyleneimine) pellet using a buffer that contains 200 mM NaCl (lanes 6, 7); proteins eluted from the poly(ethyleneimine) pellet using a buffer that contains 500 mM NaCl (lane 8); protein preparation after heparin affinity chromatography (lane 9); protein preparation after gel filtration chromatography (lane 10). (B) Detection of $MafR_{V583}$ in *E. faecalis* cell extracts by western blotting. Rabbit polyclonal antibodies against $MafR_{V583}$ were used. Total proteins from JH2-2 cells harboring plasmid pDLF (lane 2), plasmid pDLF*mafR*$_{V583}$ (lane 3), and plasmid pDLS*mafR*$_{V583}$ (lane 4) were separated by SDS/PAGE (10%). Purified $MafR_{V583}$ (lane 1) was run in the same gel.

Edman degradation showed that the first Met residue was not removed.

Next, we obtained polyclonal antibodies against $MafR_{V583}$ and demonstrated that they are suitable for detection of $MafR_{V583}$ in enterococcal whole-cell extracts by western blotting (Fig. 1B). Specifically, we inserted the *mafR*$_{V583}$ gene into the expression vectors pDLF [6] and pDLS (this work), which are based on the enterococcal *P2493* promoter and the pneumococcal *PsulA* promoter, respectively [26]. Each recombinant plasmid (pDLF*mafR*$_{V583}$ and pDLS*mafR*$_{V583}$) was introduced into the enterococcal JH2-2 strain, which is a plasmid-free strain. As internal control, plasmid pDLF ('empty' vector) was introduced into JH2-2. Compared to cells harboring plasmid pDLS*mafR*$_{V583}$, the amount of $MafR_{V583}$ was ~ 3-fold higher in cells harboring plasmid pDLF*mafR*$_{V583}$ (Fig. 1B). This result is consistent with our previous results, which showed that the activity of the *P2493* promoter is ~ 3.8-fold higher than the activity of the *PsulA* promoter in *E. faecalis* JH2-2 cells [26].

Hydrodynamic behavior of $MafR_{V583}$

Gel filtration chromatography allowed us to determine the molecular size (Stokes radius) of the untagged $MafR_{V583}$ protein. $MafR_{V583}$ eluted from the column as a single peak (Fig. 2A). The elution volume was used to calculate the K_{av} value as indicated in Materials and Methods. A calibration curve was obtained by loading onto the column several standard proteins of known Stokes radius (Fig. 2A). The Stokes radius of $MafR_{V583}$ determined from the calibration curve was 43 Å, slightly lower than the value of the alcohol dehydrogenase standard protein (45 Å, 150 kDa). This result indicated that $MafR_{V583}$ behaves as a dimer in solution. This conclusion was further confirmed by analytical ultracentrifugation experiments (sedimentation velocity and sedimentation equilibrium) (Fig. 2B). At 5 and 10 μM of $MafR_{V583}$, the sedimentation velocity profiles showed a major peak (92–98%) with an $S_{20,w}$ value of 6 S. A minor peak (4–6%) corresponding to a molecular species of higher sedimentation coefficient ($S_{20,w}$ = 6.4 S) was also observed. Sedimentation equilibrium assays showed that, at 5 μM of $MafR_{V583}$, the experimental data are best fit to an average molecular mass ($M_{w,a}$) of 118 000 ± 1000 Da, a value that corresponds with the theoretical mass of a $MafR_{V583}$ dimer (112 495 Da). A similar average molecular mass was determined at 10 μM (121 000 ± 1000 Da). Thus, under the conditions tested, the $MafR_{V583}$ dimer is the major molecular species in the protein preparation. The frictional ratio (f/f_0) calculated from the analytical ultracentrifugation assays was 1.34, indicating that the hydrodynamic behavior of the $MafR_{V583}$ dimer deviates from the behavior corresponding to a rigid spherical particle (f/f_0 = 1.0). From these results we conclude that the shape of the $MafR_{V583}$ dimer is an ellipsoid.

$MafR_{V583}$ generates multimeric complexes on linear dsDNA

The pneumococcal Mga*Spn* regulator, a member of the Mga/AtxA family, was shown to generate multimeric complexes on linear dsDNAs [10]. In this work, we analyzed whether $MafR_{V583}$ has a similar ability. Firstly, we performed electrophoretic mobility shift assays (EMSAs) using a 217-bp DNA fragment (coordinates 2888932–2889148 of the V583 chromosome) (Fig. 3A), which contains the promoter of the *mafR* gene (promoter *Pma*) [6]. The ^{32}P-labeled DNA fragment (2 nM) was incubated with increasing concentrations of $MafR_{V583}$ in the presence of nonlabeled competitor calf thymus DNA (5 μg·mL^{-1}) (Fig. 4A). At 20 nM of $MafR_{V583}$, free

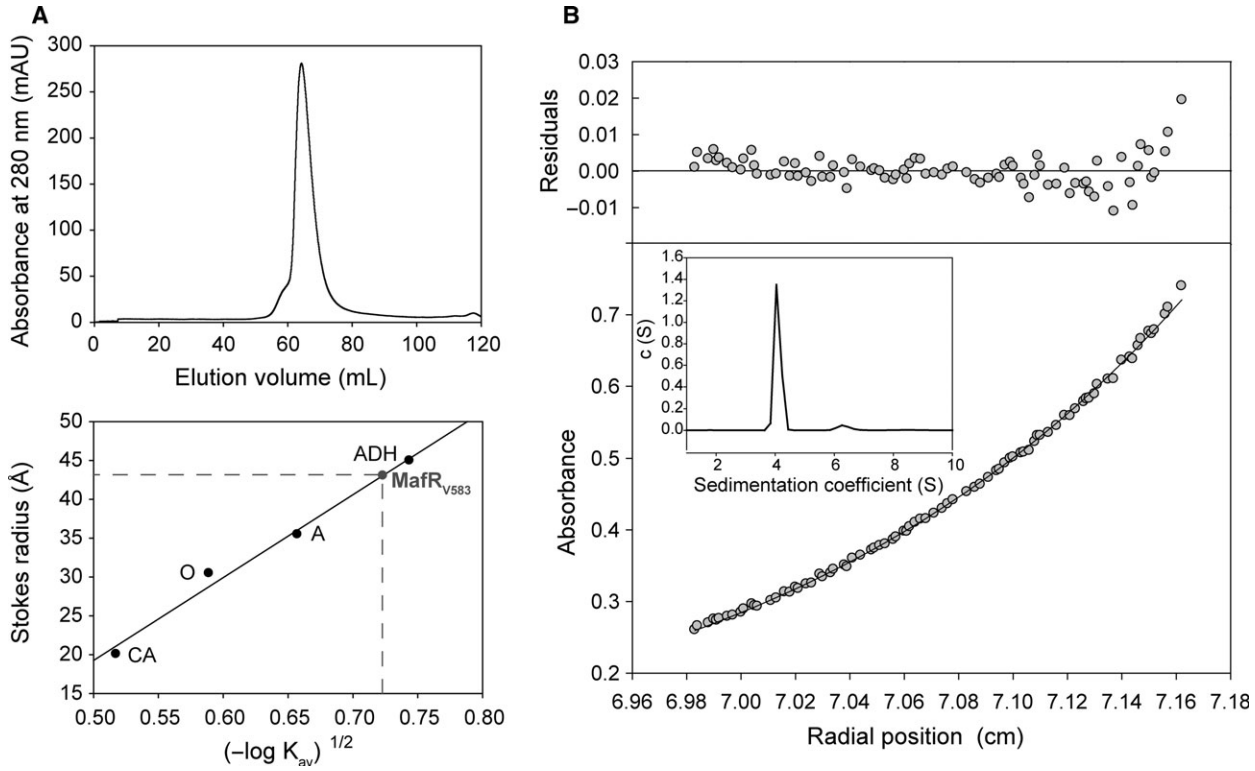


Fig. 2. (A) Stokes radius of MafR$_{V583}$ determined by gel filtration chromatography. Upper part: elution profile of MafR$_{V583}$ on a HiLoad Superdex 200 gel filtration column. Lower part: four standard proteins of known Stokes radius were loaded on the same column. CA: carbonic anhydrase. O: ovalbumin. A: albumin. ADH: alcohol dehydrogenase. (B) Analytical ultracentrifugation analysis of MafR$_{V583}$. The sedimentation equilibrium profile of MafR$_{V583}$ (5 μM) was taken at 280 nm. The lower part shows the experimental data (circles) and the best fit (continuous line) to a single species with a $M_{w,a}$ = 118 000 Da. The residuals to the fit are shown in the upper part. Inset: sedimentation velocity profile of the same MafR$_{V583}$ sample.

DNA and a protein–DNA complex (C1) were observed. However, as the concentration of MafR$_{V583}$ was increased, complexes of lower electrophoretic mobility appeared sequentially and faster moving complexes disappeared gradually. Secondly, we performed dissociation experiments (Fig. 4B). The ^{32}P-labeled DNA fragment (2 nM) was incubated with MafR$_{V583}$ (260 nM) in the absence of competitor DNA to generate higher order protein–DNA complexes (Fig. 4B, lane 1). Then, different amounts of calf thymus DNA were added to the reaction mixtures. As the concentration of competitor DNA was increased, protein–DNA complexes moving faster appeared gradually, as well as free DNA molecules. Taken together, these results indicated that multiple units of MafR$_{V583}$ bind orderly on the same linear DNA molecule generating multimeric complexes. Moreover, the MafR$_{V583}$ units are able to dissociate orderly from the higher order complexes.

Further EMSAs using different linear dsDNAs indicated that MafR$_{V583}$, like the Mga*Spn* regulator [10], binds to DNA with little or no sequence specificity. Specifically, we used a 321-bp DNA fragment from the enterococcal V583 chromosome (coordinates 94488–94808) and a 322-bp DNA fragment from the pneumococcal R6 chromosome (coordinates 1598010–1598331). Although both DNAs have a similar A + T content (72.3% and 71.1%, respectively), sequence similarities between them are not apparent. As shown in Fig. S1 (Supplementary material), MafR$_{V583}$ was able to form multimeric complexes on both DNA fragments. Moreover, the pattern of complexes was similar to that generated by MafR$_{V583}$ on the 217-bp DNA fragment (Fig. 4A).

By EMSA, we also analyzed the ability of a His-tagged MafR$_{V583}$ protein (MafR$_{V583}$-His) to interact with linear dsDNAs. This variant of MafR$_{V583}$ carries the Leu-Glu-6xHis peptide (His-tag) fused to its C terminus. As shown in Fig. S2 (Supplementary material), MafR$_{V583}$-His was able to generate multiple complexes on a 260-bp DNA fragment (coordinates 2888858–2889117 of the V583 chromosome) that contains the *Pma* promoter (Fig. 3A). Thus, the presence of the His-tag at the C-terminal end of MafR$_{V583}$ does not affect its ability to generate multimeric complexes.

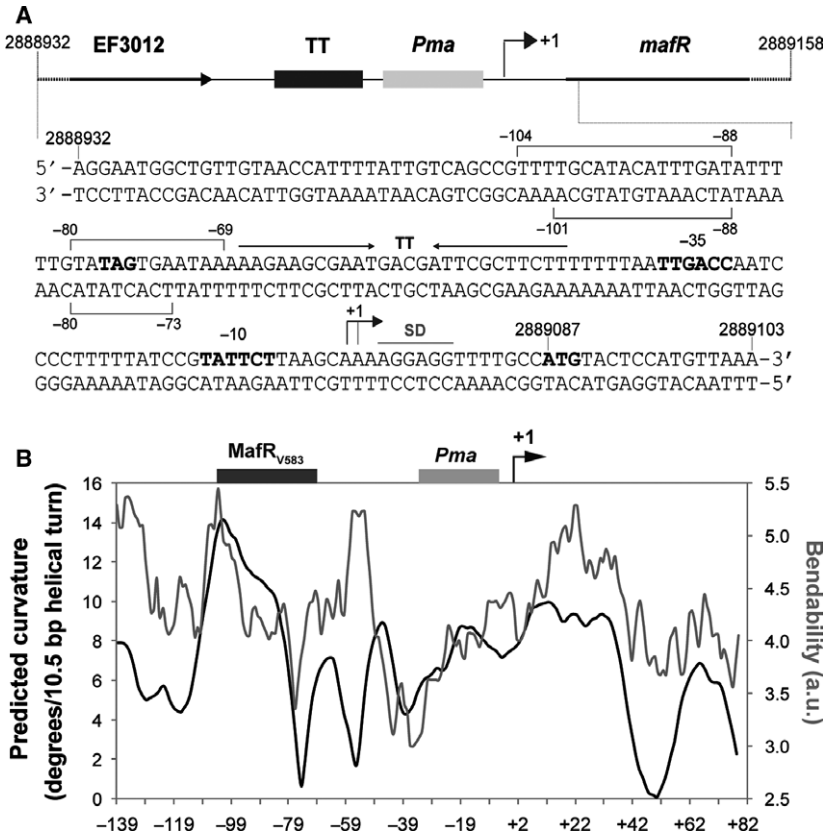


Fig. 3. (A) Relevant features of the region located upstream of the *mafR* gene. The nucleotide sequence of the region that spans coordinates 2888932–2889103 of the *E. faecalis* V583 chromosome is shown. The stop codon (TAG) of EF3012, the main sequence elements of the *Pma* promoter (−35 box and −10 box), and the start codon (ATG) of *mafR* are indicated in boldface letters. The transcriptional terminator (TT) of EF3012, the transcription start site (+ 1 position) of *mafR*, and the putative Shine-Dalgarno sequence (SD) of *mafR* [6] are shown. The brackets denote the primary MafR$_{V583}$-binding site defined by DNase I footprinting assays in this work. (B) The region upstream of the *Pma* promoter is potentially curved. The bendability/curvature propensity plot of the 227-bp DNA fragment (coordinates 2888932–2889158 of the V583 chromosome) was calculated using the bend.it program [35]. The transcription start site (+ 1) of the *mafR*$_{V583}$ gene is indicated with an arrow. The location of the main sequence elements of the *Pma* promoter (−35 and −10 elements) is shown (gray box). The site recognized preferentially by MafR$_{V583}$ on the 227-bp DNA fragment (this work, Fig. 5) is indicated (black box).

MafR$_{V583}$ binds to a potentially curved DNA region

Regulators of the Mga/AtxA family appear to bind DNA with low sequence specificity. It has been shown that all established Mga-binding sites exhibit only 13.4% identity [16]. In the case of AtxA, sequence similarities in its target promoters are not apparent, and *in silico* and *in vitro* analyses revealed that the promoter regions of several target genes are intrinsically curved [20]. Moreover, *in vitro* studies indicated that Mga*Spn* recognizes structural features in its DNA targets rather than specific nucleotide sequences [10,19]. To further investigate the DNA-binding properties of MafR$_{V583}$, we performed DNase I footprinting experiments. We used a 227-bp DNA fragment (coordinates 2888932–2889158) that contains the *Pma* promoter (Fig. 3A). Figure 3B shows the bendability/curvature propensity plot of this DNA fragment according to the bend.it program [35]. The highest magnitude of curvature propensity (~ 14 degrees per helical turn) is located at position −104 relative to the transcription start site of the *mafR*$_{V583}$ gene. Thus, the region upstream of the *Pma* promoter is potentially curved. Moreover, such a predicted curvature is located at a region of conspicuous bendability.

The 227-bp DNA fragment (2 nM) was ^{32}P-labeled either at the 5′-end of the coding strand or at the 5′-end of the noncoding strand (Fig. 5). On the coding strand and at 80 nM of MafR$_{V583}$, two regions protected against DNase I digestion were observed, from position −69 to −80 and from position −88 to −104. On the

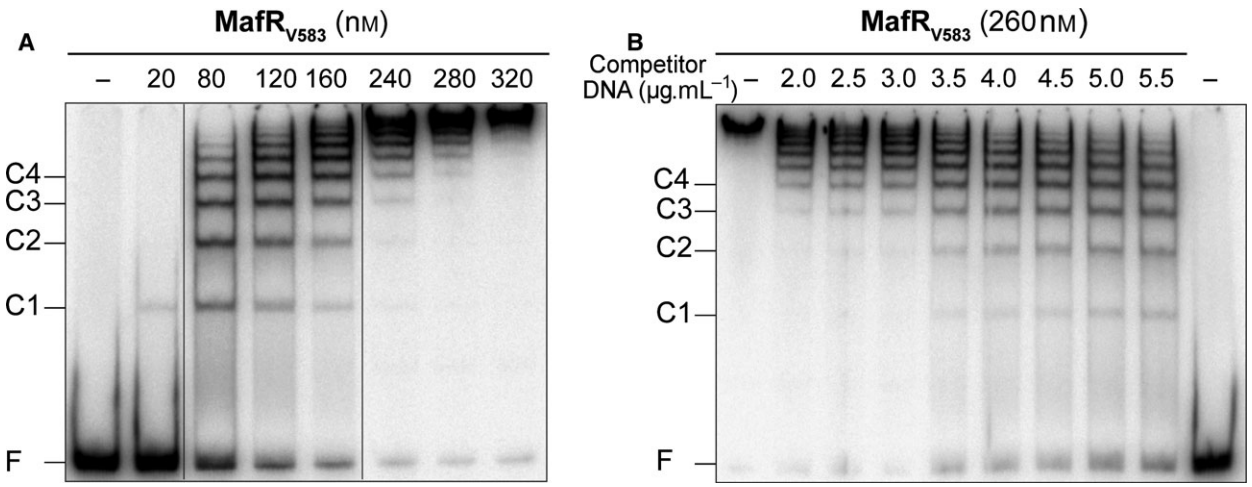


Fig. 4. (A) Formation of MafR$_{V583}$–DNA complexes. The ^{32}P-labeled 217-bp DNA fragment (2 nM) was incubated with the indicated concentration of MafR$_{V583}$ in the presence of nonlabeled calf thymus DNA (5 μg·mL^{-1}). Reaction mixtures were loaded onto a native gel (6% polyacrylamide). All the lanes came from the same gel. Bands corresponding to free DNA (F) and some MafR$_{V583}$–DNA complexes (C1, C2, C3, and C4) are indicated. (B) Dissociation of MafR$_{V583}$-DNA complexes. The indicated concentration of nonlabeled competitor calf thymus DNA was added to preformed MafR$_{V583}$–DNA complexes (260 nM MafR$_{V583}$, 2 nM ^{32}P-labeled 217-bp DNA).

noncoding strand and at 120 nM of MafR$_{V583}$, diminished cleavages were observed from −73 to −80 and from −88 to −101. These results indicated that, on the 227-bp DNA, MafR$_{V583}$ binds preferentially to a region located upstream of the *Pma* promoter (between positions −69 and −104) (Fig. 3A). Such a site is adjacent to the peak of the potential curvature (Fig. 3B). On both DNA strands and at 200 nM of MafR$_{V583}$, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the pattern of protein–DNA complexes observed by EMSA (Fig. 4A).

MafR$_{V583}$ binds to a 32-bp DNA but not to a 26-bp DNA

To define the minimum size of DNA needed for binding of MafR$_{V583}$, we used dsDNAs of 20, 26, 32, and 40 bp, which were obtained by annealing of complementary oligonucleotides. The sequence of such oligonucleotides (Table 1) was based on the pneumococcal R6 genome, and identical sequences were not found in the enterococcal V583 genome. The nonlabeled dsDNAs were incubated with increasing concentrations of MafR$_{V583}$. Reaction mixtures were then analyzed by electrophoresis on native polyacrylamide gels. In the case of the 20-bp DNA, a faint band was observed at 0.3 μM of MafR$_{V583}$ (Fig. 6A). However, its intensity did not change significantly as the protein concentration was increased. Most of the DNA moved as free DNA even at high protein concentrations (6–9 μM). In the case of the 26-bp DNA, most of the DNA moved as free DNA at 4 μM of MafR$_{V583}$ (Fig. 6B). Different results were obtained with the 32-bp DNA (Fig. 6B) and the 40-bp DNA (not shown). In both cases, a protein–DNA complex was detected at 0.2 μM of MafR$_{V583}$, and its amount increased as the amount of free DNA decreased. From these results we conclude that the minimum size of DNA required for MafR$_{V583}$ binding is between 26 bp and 32 bp. In the case of the pneumococcal Mga*Spn* regulator, the minimum DNA size for binding was reported to be between 20 bp and 26 bp [10].

Protein MafR$_{OG1RF}$-His, but not MafR$_{OG1RF}$Δ3N-His, binds to DNA

The genome sequence of the enterococcal OG1RF strain was published in 2008 [22]. Compared to strain V583 [21], the OG1RF genome contains 227 unique open reading frames but has fewer mobile genetic elements. Despite this difference between both genomes, the nucleotide sequence of the region that spans coordinates 2888932–2889103 in V583 is identical in OG1RF (Fig. 3A). Such a region contains the *Pma* promoter [6] and the primary binding site of MafR$_{V583}$ (this work). However, in contrast to MafR$_{V583}$, MafR encoded by the OG1RF genome (MafR$_{OG1RF}$) has five amino acid changes (Ala37Thr, Gln131Leu, Met145Thr, Ser193Asn, Ile388Ser), and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the predicted DNA-binding domain (residues 11–164).

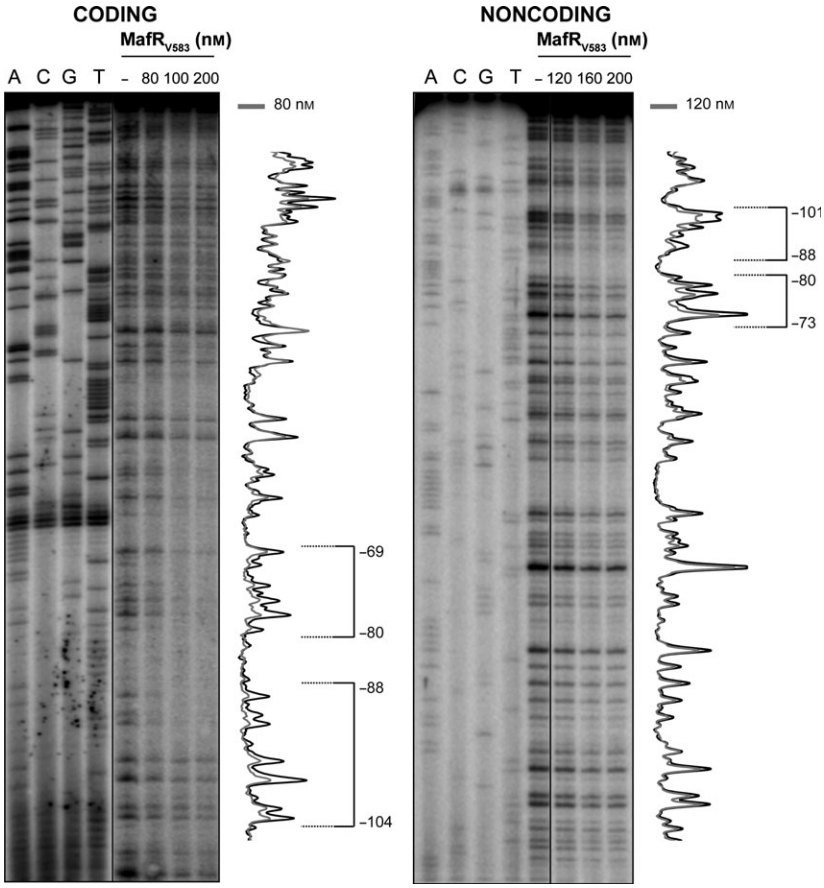


Fig. 5. DNase I footprints of complexes formed by MafR$_{V583}$ on the 227-bp DNA fragment. Either the coding or the noncoding strand relative to the *Pma* promoter was ^{32}P-labeled at the 5′-end. All the lanes displayed came from the same gel. Dideoxy-mediated chain termination sequencing reactions were run in the same gel (lanes A, C, G, T). Densitometer scans corresponding to DNA without protein (black line) and DNA with the indicated concentration of protein (gray line) are shown. The regions protected against DNase I digestion are indicated with brackets. The indicated positions are relative to the transcription start site of the *mafR*$_{V583}$ gene.

To analyze whether these changes affect the formation of protein–DNA complexes, we purified a His-tagged MafR$_{OG1RF}$ protein (MafR$_{OG1RF}$-His). MafR$_{OG1RF}$-His carries the His-tag fused to its C terminus. EMSA experiments showed that MafR$_{OG1RF}$-His is able to form multimeric complexes on the 217-bp DNA fragment that contains the *Pma* promoter (Fig. 7). The pattern of complexes was similar to that generated by MafR$_{V583}$ (Fig. 4A). Thus, the amino acid substitutions of MafR$_{OG1RF}$ do not affect its interaction with DNA.

We identified previously the transcription initiation site of the *mafR*$_{V583}$ gene at coordinate 2889071, and proposed that the first ATG codon is likely the translation start site [6] (Fig. 3A). Translation from this ATG generates a protein of 482 amino acid residues (MafR$_{V583}$ and MafR$_{OG1RF}$). However, there is a second ATG codon that might function as a translation start site. Translation from this second ATG would result in a variant that lacks the first three amino acid residues (here named MafR$_{OG1RF}$Δ3N). To analyze whether the lack of these residues (Met-Tyr-Ser) impairs the formation of protein–DNA complexes, we purified a His-tagged MafR$_{OG1RF}$Δ3N protein (MafR$_{OG1RF}$Δ3N-His). The sequence of its N-terminal end (eight residues) was confirmed by Edman degradation. EMSA experiments showed that MafR$_{OG1RF}$Δ3N-His has lost the capacity to interact with DNA (Fig. 7). From these results we conclude that (a) the first ATG codon is the translation start site, and (b) the first three amino acid residues of MafR$_{OG1RF}$ are crucial for its structure and/or function.

Discussion

Bacteria have evolved complex regulatory networks to rapidly adapt to environmental fluctuations. Global transcriptional regulators are key elements in such networks due to their ability to activate and/or repress the expression of multiple genes through a variety of mechanisms. In *E. faecalis*, various transcriptome analyses support the notion that bacterial adaptation is associated with global changes in gene expression [7,8,36]. Our previous work identified MafR as a protein involved in global regulation of gene expression. Moreover, we proposed that MafR is a member of the

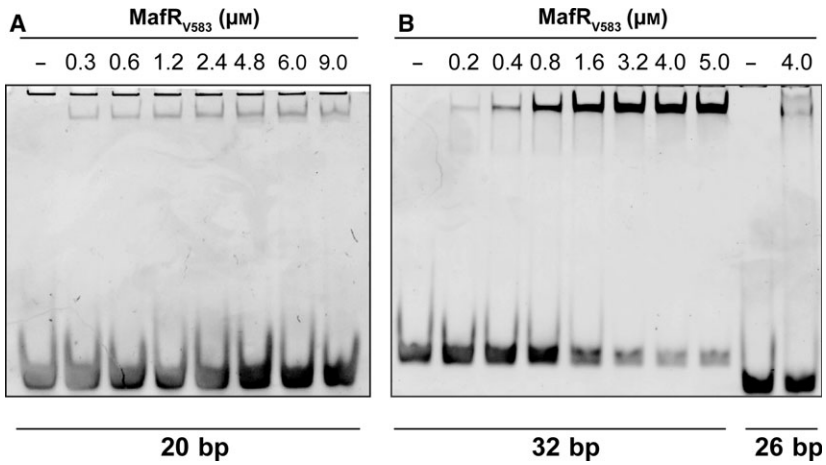


Fig. 6. EMSAs using MafR$_{V583}$ and small DNA fragments. DNA fragments were obtained by annealing of complementary oligonucleotides. (A) Oligonucleotides 20A/20B were used to generate the 20-bp DNA. (B) Oligonucleotides 26A/26B and 32A/32B were used to generate the 26- and 32-bp DNAs, respectively. The indicated concentration of MafR$_{V583}$ was mixed with 300 nM (20 and 26 bp) or 200 nM (32 bp) of DNA. Binding reactions were analyzed by native polyacrylamide (8%) gel electrophoresis. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad).

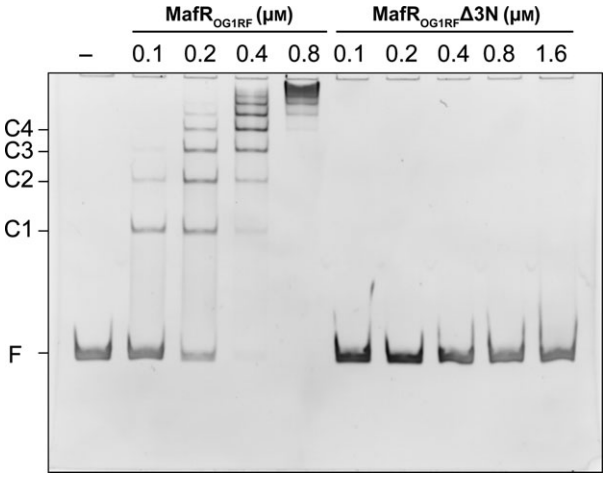


Fig. 7. EMSAs with MafR$_{OG1RF}$-His and MafR$_{OG1RF}$Δ3N-His The 217-bp DNA fragment (10 nM) was incubated with the indicated concentration of protein. Reaction mixtures were loaded onto a native gel (6% polyacrylamide). Bands corresponding to free DNA (F) and protein–DNA complexes (C1, C2, C3, and C4) are indicated. DNA was stained with GelRed (Biotium) and visualized using a Gel Doc system (Bio-Rad).

Mga/AtxA family of global regulators. This proposal was based on amino acid sequence similarities and predictions of functional domains [6]. However, the interaction of MafR with DNA has not previously been investigated.

In this work, we purified an untagged form of the MafR$_{V583}$ protein. Using MafR$_{V583}$ and various linear dsDNAs, we have found that the DNA-binding behavior of MafR$_{V583}$ is very similar to the one described for the pneumococcal Mga*Spn* regulator [10]. First, gel retardation assays indicated that MafR$_{V583}$ is able to bind DNA with low sequence specificity. Second, on a 227-bp DNA fragment that contains the *Pma* promoter, footprinting experiments showed that MafR$_{V583}$ binds preferentially to a site that is adjacent to the peak of a potential curvature. Such a binding site is located upstream of the *Pma* promoter (positions −69 to −104). In the case of Mga*Spn*, two primary binding sites were identified by footprinting experiments: the *P1623B* promoter site (positions −60 to −99) and the *Pmga* promoter site (positions −23 to + 21) [10]. Both binding sites have a low sequence identity but contain an intrinsic curvature flanked by regions of bendability. Regarding the Mga and AtxA regulators, a low sequence identity has been found in the target promoters [13,16,20]. Moreover, the promoter regions of some AtxA target genes are intrinsically curved [20].

Another interesting finding of our current study is that MafR$_{V583}$, like Mga*Spn* [10], is able to generate multimeric complexes on linear dsDNAs. In both cases, the pattern of complexes observed by EMSA is compatible with a sequential binding of multiple protein units (likely dimers) to the same DNA molecule. Furthermore, in both cases, the protein units dissociate sequentially from the higher order complexes in the presence of increasing amounts of competitor DNA. Formation of multimeric protein–DNA complexes has not been reported for the Mga regulator. Nevertheless,

a correlation between its ability to oligomerize in solution (without DNA) and its ability to activate transcription has been shown [18]. Regarding AtxA, protein–DNA interaction studies have not been reported, but protein–protein interaction analyses revealed that AtxA exists in a homo-oligomeric state [37].

The genome sequence of the *E. faecalis* V583 clinical isolate was published in 2003 [21]. Five years later the genome sequence of the OG1RF strain was published [22]. This strain is a derivative of the OG1 human isolate. At present, 557 *E. faecalis* genomes have been totally or partially sequenced (NCBI, Genome Assembly and Annotation Report, 22/12/2017). Among them, we have found that 146 *E. faecalis* genomes encode a protein that is identical to $MafR_{V583}$ (Table S1). Furthermore, we have found that 50 *E. faecalis* genomes encode a protein that is identical to $MafR_{OG1RF}$ (Table S2). Compared to $MafR_{V583}$, $MafR_{OG1RF}$ has five amino acid substitutions, and three of them (Ala37Thr, Gln131Leu, Met145Thr) are located within the putative DNA-binding domain (Table 2). Despite this difference between $MafR_{V583}$ and $MafR_{OG1RF}$, we have shown that both proteins are able to generate multimeric complexes on linear dsDNAs, which could be an indication of functional conservation. Moreover, we have shown that removal of the first three amino acids in $MafR_{OG1RF}$, and likely in $MafR_{V583}$, results in a protein unable to interact with DNA. Similar DNA-binding properties are expected for MafR from the strains that belong to the groups GA2, KS19, MTmid8, B653, X98, AZ19 and Com1 (Table S3). Compared to $MafR_{V583}$, MafR from such strains has one to four amino acid substitutions, and all of them are present in $MafR_{OG1RF}$ (Table 2).

Table 2. Amino acid substitutions in the indicated MafR proteins compared to $MafR_{V583}$. Data from the National Center for Biotechnology Information (Identical Protein Groups) (22/12/2017).

	Amino acid residue					
Strain[a]	37	131	145	193	388	Identical[b]
V583	A	Q	M	S	I	146
GA2	T					2
KS19	T		T			6
MTmid8			T		S	2
B653	T	L	T			1
X98	T		T		S	15
AZ19	T	L	T	N		8
Com1	T	L	T		S	7
OG1RF	T	L	T	N	S	50

[a]Name of the *E. faecalis* strain that represents the group.
[b]Number of strains that belong to the group (identical MafR) (Tables S1, S2, and S3).

In conclusion, recognition of specific sites across the genome by transcriptional regulators is essential for controlling gene expression. Different mechanisms for protein recognition of specific DNA sites have been characterized. In some cases, DNA-binding proteins recognize intrinsic DNA structural characteristics rather than particular nucleotide sequences [2,38]. Examples where both DNA base sequence and shape recognition are required for protein binding have been also reported [39–41]. Our study on the DNA-binding properties of MafR reinforces that recognition of particular DNA structures might be a general feature of the global regulators that constitute the Mga/AtxA family.

Acknowledgements

We thank Dr. Virtu Solano-Collado for helpful discussions. This work was supported by grants BIO2016-76412-C2-2-R (AEI/FEDER, UE) and BIO2015-69085-REDC from the Spanish Ministry of Economy and Competitiveness.

Author contributions

SR-C and AM-B performed experiments. SR-C, ME and AB designed the study and wrote the manuscript. All authors read and approved the final manuscript.

References

1 Siggers T and Gordân R (2014) Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res* **42**, 2099–2111.
2 Rohs R, Jin X, West SM, Joshi R, Honig B and Mann RS (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* **79**, 233–269.
3 Ramsey M, Hartke A and Huycke M (2014) The physiology and metabolism of Enterococci. In Enterococci: From Commensals to Leading Causes of Drug Resistant Infection (Internet) (Gilmore MS, Clewell DB, Ike Y and Shankar N, eds), pp. 1–43. Massachusetts Eye and Ear Infirmary, Boston.
4 Hollenbeck BL and Rice LB (2012) Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* **3**, 421–433.
5 Fisher K and Phillips C (2009) The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* **155**, 1749–1757.
6 Ruiz-Cruz S, Espinosa M, Goldmann O and Bravo A (2016) Global regulation of gene expression by the MafR protein of *Enterococcus faecalis*. *Front Microbiol* **6**, 1521.

7 Vebø HC, Snipen L, Nes IF and Brede DA (2009) The transcriptome of the nosocomial pathogen *Enterococcus faecalis* V583 reveals adaptive responses to growth in blood. *PLoS One* **4**, e7660.
8 Vebø HC, Solheim M, Snipen L, Nes IF and Brede DA (2010) Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. *PLoS One* **5**, e12489.
9 Hondorp ER, Hou SC, Hause LL, Gera K, Lee C-E and McIver KS (2013) PTS phosphorylation of Mga modulates regulon expression and virulence in the group A streptococcus. *Mol Microbiol* **88**, 1176–1193.
10 Solano-Collado V, Lurz R, Espinosa M and Bravo A (2013) The pneumococcal Mga*Spn* virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. *Nucleic Acids Res* **41**, 6975–6991.
11 Hammerstrom TG, Horton LB, Swick MC, Joachimiak A, Osipiuk J and Koehler TM (2015) Crystal structure of *Bacillus anthracis* virulence regulator AtxA and effects of phosphorylated histidines on multimerization and activity. *Mol Microbiol* **95**, 426–441.
12 Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–D285.
13 Hondorp ER and McIver KS (2007) The Mga virulence regulon: infection where the grass is greener. *Mol Microbiol* **66**, 1056–1065.
14 Solano-Collado V, Espinosa M and Bravo A (2012) Activator role of the pneumococcal Mga-like virulence transcriptional regulator. *J Bacteriol* **194**, 4197–4207.
15 Ribardo DA and McIver KS (2006) Defining the Mga regulon: comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. *Mol Microbiol* **62**, 491–508.
16 Hause LL and McIver KS (2012) Nucleotides critical for the interaction of the *Streptococcus pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. *J Bacteriol* **194**, 4904–4919.
17 McIver KS, Heath AS, Green BD and Scott JR (1995) Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. *J Bacteriol* **177**, 6619–6624.
18 Hondorp ER, Hou SC, Hempstead AD, Hause LL, Beckett DM and McIver KS (2012) Characterization of the group A streptococcus Mga virulence regulator reveals a role for the C-terminal region in oligomerization and transcriptional activation. *Mol Microbiol* **83**, 953–967.
19 Solano-Collado V, Hüttener M, Espinosa M, Juárez A and Bravo A (2016) Mga*Spn* and H-NS: two unrelated global regulators with similar DNA-binding properties. *Front Mol Biosci* **3**, 60.
20 Hadjifrangiskou M and Koehler TM (2008) Intrinsic curvature associated with the coordinately regulated anthrax toxin gene promoters. *Microbiology* **154**, 2501–2512.
21 Paulsen IT, Banerjei L, Myers GSA, Nelson KE, Seshadri R, Read TD, Fouts DE, Eisen JA, Gill SR, Heidelberg JF *et al.* (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **299**, 2071–2074.
22 Bourgogne A, Garsin DA, Qin X, Singh KV, Sillanpaa J, Yerrapragada S, Ding Y, Dugan-Rocha S, Buhay C, Shen H *et al.* (2008) Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol* **9**, R110.
23 Hoskins J, Alborn WE Jr, Arnold J, Blaszczak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu D-J, Fuller W *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**, 5709–5717.
24 Jacob AE and Hobbs SJ (1974) Conjugal transfer of plasmid-borne multiple antibiotic resistance in *Streptococcus faecalis* var. *zymogenes*. *J Bacteriol* **117**, 360–372.
25 LeBlanc DJ, Chen YY and Lee LN (1993) Identification and characterization of a mobilization gene in the streptococcal plasmid, pVA380-1. *Plasmid* **30**, 296–302.
26 Ruiz-Cruz S, Solano-Collado V, Espinosa M and Bravo A (2010) Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. *J Microbiol Methods* **83**, 156–163.
27 Dower WJ, Miller JF and Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127–6145.
28 Shepard BD and Gilmore MS (1995) Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. *Methods Mol Biol* **47**, 217–226.
29 Siegel LM and Monty KJ (1966) Determination of molecular weights and frictional ratios of proteins in impure systems by the use of gel filtration and density gradient centrifugation. Applications to crude preparations of sulfite and hydroxylamine reductases. *Biochim Biophys Acta* **112**, 346–362.
30 Schuck P and Rossmanith P (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54**, 328–341.
31 Laue TM, Shah BD, Ridgeway TM and Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer

Sciences (Harding SE, Rowe A and Horton JC, eds), pp. 90–125. Royal Society of Chemistry, Cambridge.
32 van Holde KE (1985) Physical Biochemistry. Prentice-Hall, Englewood Cliffs.
33 Pessen H and Kumosinski TF (1985) Measurements of protein hydration by various techniques. *Methods Enzymol* **117**, 219–255.
34 Cole JL (2004) Analysis of heterogeneous interactions. *Methods Enzymol* **384**, 212–232.
35 Vlahovicek K, Kaján L and Pongor S (2003) DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res* **31**, 3686–3687.
36 Lindenstrauss AG, Ehrmann MA, Behr J, Landstorfer R, Haller D, Sartor RB and Vogel RF (2014) Transcriptome analysis of *Enterococcus faecalis* toward its adaption to surviving in the mouse intestinal tract. *Arch Microbiol* **196**, 423–433.
37 Hammerstrom TG, Roh JH, Nikonowicz EP and Koehler TM (2011) *Bacillus anthracis* virulence regulator AtxA: oligomeric state, function and CO_2 -signalling. *Mol Microbiol* **82**, 634–647.
38 Abe N, Dror I, Yang L, Slattery M, Zhou T, Bussemaker HJ, Rohs R and Mann RS (2015) Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307–318.
39 Deng Z, Wang Q, Liu Z, Zhang M, Machado ACD, Chiu TP, Feng C, Zhang Q, Yu L, Qi L *et al.* (2015) Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nat Commun* **6**, 7642.
40 Ding P, McFarland KA, Jin S, Tong G, Duan B, Yang A, Hughes TR, Liu J, Dove SL, Navarre WW *et al.* (2015) A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. *PLoS Pathog* **11**, e1004967.
41 Al-Zyoud WA, Hynson RMG, Ganuelas LA, Coster ACF, Duff AP, Baker MAB, Stewart AG, Giannoulatou E, Ho JWK, Gaus K *et al.* (2016) Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. *Nucleic Acids Res* **44**, 1411–1420.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Supplementary material.

SCIENTIFIC REPORTS

OPEN

Transcriptional activation by MafR, a global regulator of *Enterococcus faecalis*

Received: 7 January 2019
Accepted: 1 April 2019
Published online: 16 April 2019

Sofía Ruiz-Cruz, Ana Moreno-Blanco, Manuel Espinosa & Alicia Bravo

Proteins that act as global transcriptional regulators play key roles in bacterial adaptation to new niches. These proteins recognize multiple DNA sites across the bacterial genome by different mechanisms. *Enterococcus faecalis* is able to survive in various niches of the human host, either as a commensal or as a leading cause of serious infections. Nonetheless, the regulatory pathways involved in its adaptive responses remain poorly understood. We reported previously that the MafR protein of *E. faecalis* causes genome-wide changes in the transcriptome. Here we demonstrate that MafR functions as a transcription activator. *In vivo*, MafR increased the activity of the *P12294* and *P11486* promoters and also the transcription levels of the two genes controlled by those promoters. These genes are predicted to encode a calcium-transporting P-type ATPase and a QueT transporter family protein, respectively. Thus, MafR could have a regulatory role in calcium homeostasis and queuosine synthesis. Furthermore, MafR recognized *in vitro* specific DNA sites that overlap the −35 element of each target promoter. The MafR binding sites exhibit a low sequence identity, suggesting that MafR uses a shape readout mechanism to achieve DNA-binding specificity.

Global transcriptional regulators play crucial roles during bacterial adaptation to specific niches. They activate and/or repress the transcription of multiple genes and, therefore, make possible to rapidly adjust the gene expression pattern to new environmental situations. *Enterococcus faecalis* is usually found as a harmless commensal in the human gastrointestinal tract. However, this Gram-positive bacterium is able to colonize other niches of the human host and cause a variety of life-threatening infections, such as urinary tract infections, endocarditis or bacteraemia[1–3]. Despite the pathogenic potential of *E. faecalis*, our understanding of the regulatory circuits involved in its adaptive responses is still very limited.

The MafR protein (482 amino acids) of *E. faecalis* is highly conserved among the strains whose genomes have been totally or partially sequenced[4]. Genome-wide microarray assays showed that MafR is involved in global regulation of gene expression[5]. In such experiments, the transcriptional profiles of strains OG1RF (wild-type) and OG1RFΔ*mafR* (*mafR* deletion mutant) were compared, demonstrating that MafR activates, directly or indirectly, the expression of at least 87 genes. Many of them are organized in operons and encode proteins involved in the utilization of carbon sources (*e.g.* mannitol, glycerol, gluconate, maltose and citrate). Furthermore, compared to OG1RF, the OG1RFΔ*mafR* strain was shown to induce a lower degree of inflammation in the peritoneal cavity of mice. Because of these findings, we proposed that MafR could facilitate the growth of *E. faecalis* in particular human niches and, consequently, could contribute to its potential virulence[5].

Different protein-DNA recognition mechanisms have been characterized. In some cases, proteins recognize a sequence-dependent DNA shape (shape readout mechanism) rather than the unique chemical signatures of the DNA bases (base readout mechanism)[6,7]. MafR is a new member of the Mga/AtxA family of global transcriptional regulators[4,5]. This family includes AtxA from *Bacillus anthracis*, Mga*Spn* from *Streptococcus pneumoniae*, and Mga from *S. pyogenes*. Like these three regulatory proteins[8], MafR has two putative helix-turn-helix DNA-binding motifs within the N-terminal region, the so-called HTH_Mga (residues 11–69) and Mga (residues 76–164) motifs[5]. In the Mga regulator, both motifs were found to be required for DNA-binding and transcriptional activation[9,10]. *In vitro* protein-DNA interaction studies have shown that MafR binds to linear double-stranded DNAs with little or no sequence specificity. Furthermore, MafR was able to generate multimeric complexes on linear double-stranded DNAs[4]. Similar DNA-binding properties have been described for the pneumococcal Mga*Spn* regulator. Mga*Spn* has a preference for AT-rich DNA sites, as well as a high affinity for a naturally occurring

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain. Correspondence and requests for materials should be addressed to A.B. (email: abravo@cib.csic.es)

curved DNA[11–13]. On DNA fragments that contain the promoter of the *mafR* gene (*Pma* promoter), MafR recognizes a potentially curved DNA region, which is located upstream of the promoter (positions −69 to −104)[4]. We hypothesized that MafR, and most likely the regulators of the Mga/AtxA family, recognizes structural features in its target DNAs rather than specific nucleotide sequences[4]. Nevertheless, verification of this hypothesis requires the identification of additional MafR binding sites across the bacterial genome.

A further DNA microarray assay using an OG1RFΔ*mafR* derivative that overproduces MafR (plasmid-encoded MafR) allowed us to identify two new potential MafR target genes: *OG1RF_12294* and *OG1RF_11486*. In the presence of plasmid-encoded MafR, the highest increase in gene expression corresponded to both genes (our unpublished results). In this manuscript, we addressed the validation of such a finding by *in vivo* and *in vitro* approaches. Gene *OG1RF_12294* encodes a putative phosphorylated intermediate-type ATPase (P-type ATPase) transporter, which could contribute to maintain calcium homeostasis. Gene *OG1RF_11486* encodes a putative QueT transporter family protein, which could be involved in uptake of a queuosine biosynthetic intermediate. Here we demonstrate that MafR activates directly the transcription of both genes by binding to a specific DNA site overlapping the core promoter. Such sites exhibit a low sequence identity. This study shows, for the first time, that MafR functions as a transcription activator. Moreover, it supports that MafR might recognize particular DNA shapes.

Results

Transcription of *mafR* under laboratory conditions. The genome of the *E. faecalis* strain OG1RF has been totally sequenced (GenBank CP002621.1)[14]. By quantitative RT-PCR (qRT-PCR) assays and using the comparative C_T method[15], we determined the relative expression of the regulatory *mafR* gene (locus_tag *OG1RF_12293*) in cells grown under laboratory conditions (Brain Heart Infusion (BHI) broth, 37 °C, without aeration) to both logarithmic and stationary phases. Transcription of *mafR* was found to be higher at logarithmic phase. Compared to stationary phase, the fold change ($\log_2$FC) in *mafR* expression was ∼4. Therefore, all the experiments shown in this work were performed at the logarithmic growth phase.

Gene *OG1RF_12294* encodes a putative P-type ATPase cation transporter. P-type ATPases constitute a large superfamily of cation and lipid pumps that use ATP hydrolysis for energy. They are integral, multispanning membrane proteins that are found in bacteria and in a number of eukaryotic plasma membranes and organelles[16]. The enterococcal *OG1RF_12294* gene, which is adjacent to *mafR* (Fig. 1A), encodes a putative P-type ATPase cation transporter. Such a gene has been annotated as *pmr1* (GeneID: 12289043) because it encodes a protein (850 amino acids) that has sequence similarity (∼52%) to eukaryotic PMR1 (plasma membrane ATPase related) P-type ATPases (Supplementary Table S1). Some PMR1-type pumps are able to transport calcium, as well as manganese, into the Golgi apparatus[17–19].

In addition to OG1RF_12294, the OG1RF genome encodes two putative calcium-transporting ATPases: OG1RF_10600 and OG1RF_11602 (Supplementary Table S2). Using the BLASTP protein sequence alignment program[20], we found that OG1RF_12294 has sequence similarity (∼53–56%) to both ATPases (Supplementary Table S1). Furthermore, OG1RF_12294 has sequence similarity (∼53–56%) to several prokaryotic proteins characterized as calcium P-type ATPases (Supplementary Table S1)[21–25]. Thus, protein OG1RF_12294 might contribute to maintain calcium homeostasis in enterococcal cells.

MafR influences positively the transcription of *OG1RF_12294*. To analyse whether MafR regulates the expression of the *OG1RF_12294* gene, we determined its relative expression in OG1RF (wild-type) and OG1RFΔ*mafR* (deletion mutant) by qRT-PCR. The $\log_2$FC in *OG1RF_12294* expression due to the presence of MafR was ∼3, indicating that MafR has a positive effect on the transcription of such a gene. This conclusion was further confirmed by increasing the intracellular level of MafR. Specifically, we determined the relative expression of *OG1RF_12294* in two strains: OG1RFΔ*mafR* harbouring pDLF (absence of MafR) and OG1RFΔ*mafR* harbouring pDLF*mafR* (plasmid-encoded MafR). In addition, we determined the relative expression of the *OG1RF_10600* and *OG1RF_11602* genes, which encode putative calcium-transporting ATPases (Supplementary Table S1). In the presence of plasmid-encoded MafR, only transcription of *OG1RF_12294* was increased ($\log_2$FC ∼4). Thus, MafR influences positively and specifically the transcription of the *OG1RF_12294* gene.

MafR activates the *P12294* promoter *in vivo*. In the OG1RF genome[14], the ATG codon at coordinate 2425611 is likely the translation start site of the *OG1RF_12294* gene (Fig. 1A). It is preceded by a putative ribosome binding site sequence (AGGAGG). Upstream of such a sequence there is a putative promoter (here named *P12294*) that has a canonical −10 element (**TATAAT**) but lacks a potential −35 element (consensus **TTGACA**) at the optimal length of 17 nucleotides. Nevertheless, there is a near-consensus −35 element (**TC**G**AC**C) at the suboptimal spacer length of 22 nucleotides. These features suggested that promoter *P12294* could be recognized by a σ factor similar to the *Escherichia coli* σ^{70} and that its activity could be enhanced by regulatory proteins. Sequence analysis of the region located between the TAA stop codon of the *OG1RF_12295* gene (coordinate 2425761) and the *P12294* promoter revealed the existence of an inverted-repeat (IR) that may function as a Rho-independent transcriptional terminator (Fig. 1A).

To characterize the *P12294* promoter, a 255-bp DNA fragment (coordinates 2425885 to 2425631) (Fig. 2) was inserted into the pASTT promoter-probe vector, which is based on the *gfp* reporter gene. The recombinant plasmid (pASTT-*P12294*) was first introduced into OG1RF and OG1RFΔ*mafR*. In these strains, the expression of *gfp* (0.32 ± 0.02 and 0.26 ± 0.04 units, respectively) was similar to the basal level (OG1RF harbouring pASTT; 0.38 ± 0.02 units). Different results were obtained when pASTT-*P12294* was introduced into OG1RFΔ*mafR* harbouring either pDLF or pDLF*mafR* (plasmid-encoded MafR) (Fig. 2). The expression of *gfp* was ∼2.5-fold higher in the presence of plasmid-encoded MafR. This result indicated that the 255-bp DNA fragment contains a MafR-dependent promoter activity. Removal of the −10 element of the *P12294* promoter resulted in loss of such

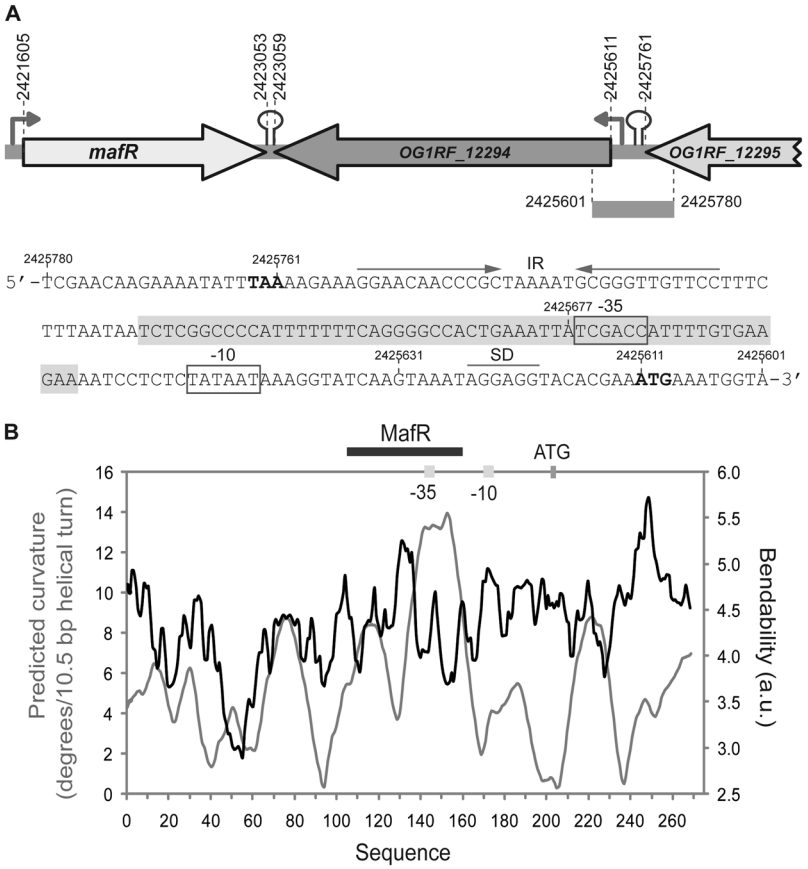


Figure 1. Relevant features of the *P12294* promoter region. (**A**) Genetic organization of the chromosome region that contains *OG1RF_12294*. Coordinates of the translation start and stop codons are indicated. Stem-loop elements represent potential transcriptional terminators. Arrows upstream of the genes represent promoters. The nucleotide sequence of the region spanning coordinates 2425780 to 2425601 is shown. The stop codon (TAA) of *OG1RF_12295* and the start codon (ATG) of *OG1RF_12294* are indicated in boldface letters. IR: inverted-repeat. SD: Shine-Dalgarno sequence. The main sequence elements (−35 box and −10 box) of the *P12294* promoter are indicated. The MafR binding site defined in this work is shown (shadowed box). Genes *OG1RF_12294* and *OG1RF_12295* correspond to genes *EF3014* and *EF3015* in *E. faecalis* strain V583. (**B**) Bendability/curvature propensity plot of the region spanning coordinates 2425817 to 2425548. The location of the *P12294* core promoter, the start codon of *OG1RF_12294* and the MafR binding site are indicated.

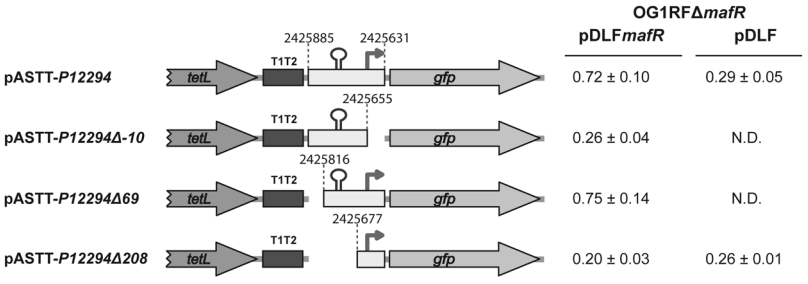


Figure 2. Effect of plasmid-encoded MafR on the activity of the *P12294* promoter. Four regions from the OG1RF chromosome were inserted independently into the *Sac*I site of pASTT. The coordinates of such regions are indicated. Gene *tetL*: tetracycline resistance determinant. Gene *gfp*: green fluorescent protein. The T1T2 box represents the tandem transcriptional terminators T1 and T2 of the *Escherichia coli rrnB* rRNA operon. The stem-loop element represents the inverted-repeat located upstream of the *P12294* promoter (see Fig. 1A). The arrow represents the canonical −10 element of the *P12294* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} = 0.4). In each case, three independent cultures were analysed. N.D.: non-determined.

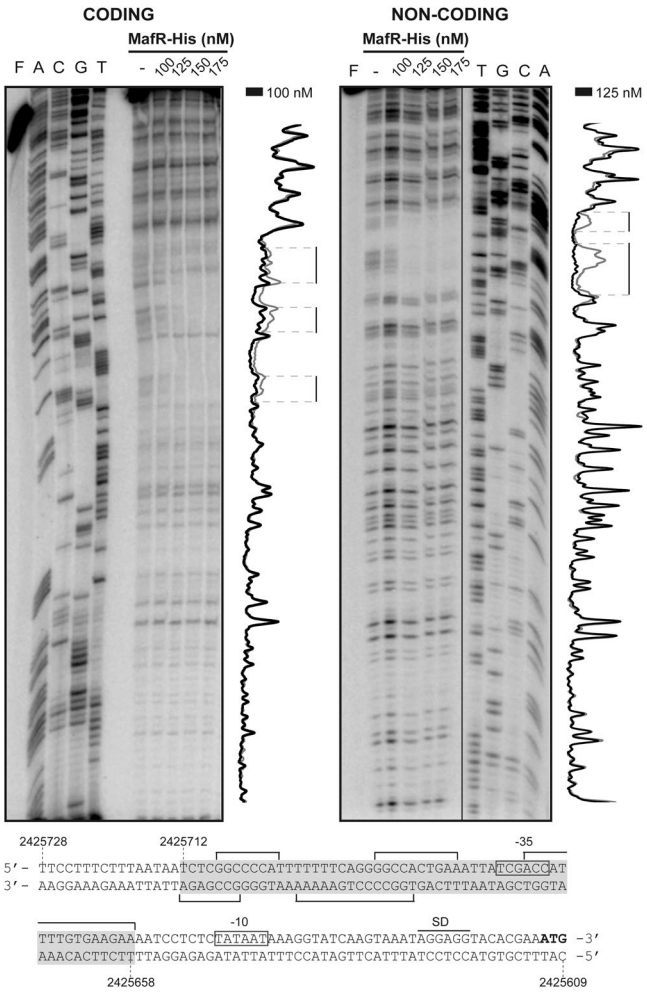


Figure 3. DNase I footprints of complexes formed by MafR-His on the 270-bp DNA fragment that contains the *P12294* promoter. ^{32}P-labelled DNA (2 nM) was incubated with the indicated concentrations of MafR-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (A, C, G, T) were run in the same gel. All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to free DNA (grey line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 2425728 to 2425609 is shown. The −35 and −10 boxes of the *P12294* promoter are indicated. SD: Shine-Dalgarno sequence. Brackets indicate regions protected against DNase I digestion. The site recognized by MafR-His (coordinates 2425712-2425658) is indicated with a grey box.

an activity (plasmid pASTT-*P12294Δ-10*). A further deletion analysis allowed us to conclude that the 186-bp region between coordinates 2425816 and 2425631 contains both the *P12294* promoter and the site required for its activation by MafR (plasmids pASTT-*P12294Δ69* and pASTT-*P12294Δ208*) (Fig. 2).

MafR binds to the *P12294* promoter region *in vitro*. To investigate whether MafR activates directly the expression of the *OG1RF_12294* gene, we performed DNase I footprinting experiments. We used a His-tagged MafR protein (MafR-His) and a 270-bp DNA fragment (coordinates 2425817 to 2425548). This fragment contains the *P12294* promoter and the site required for its activation by MafR *in vivo* (Fig. 2). The presence of a His-tag at the C-terminal end of MafR does not affect its DNA-binding properties[4]. The 270-bp DNA fragment was radioactively labelled either at the 5′-end of the coding strand or at the 5′-end of the non-coding strand (Fig. 3). On the coding strand and at 100 nM of MafR-His, protections against DNase I digestion were observed within the region spanning coordinates 2425708 and 2425658. On the non-coding strand and at 125 nM of MafR-His, diminished cleavages were observed between coordinates 2425712 and 2425686. Thus, MafR-His recognizes a site overlapping the −35 element of the *P12294* promoter (Fig. 3). This result allowed us to conclude that MafR activates directly the transcription of the *OG1RF_12294* gene.

Figure 1B shows the bendability/curvature propensity plot of the 270-bp DNA fragment according to the bend.it program[26]. The profile contains an intrinsic curvature of high magnitude (~13 degrees per helical turn), which is adjacent to the MafR binding site. In addition, the site recognized by MafR contains a region of potential bendability (~5.2 units).

Gene *OG1RF_11486* encodes a putative QueT transporter family protein. Energy-coupling factor (ECF) transporters are a family of ATP-binding cassette (ABC) transporters that are responsible for the uptake of essential micronutrients in prokaryotes. They consist of a membrane-embedded S-component that provides substrate specificity and a three-subunit ECF module that couples ATP hydrolysis to transport. In the so-called group II ECF transporters, different S-components share the same ECF module. Furthermore, the S-component genes are not located in the same operon as the genes for the ECF module[27–29].

The enterococcal *OG1RF_11486* gene encodes a putative QueT transporter family protein (GenBank AEA94173.1). Proteins identical to OG1RF_11486 (173 residues) are encoded by *Mycobacterium abscessus* (CPW17925.1), *Listeria monocytogenes* (CWW42654.1; 172 up to 173 residues are identical) and *S. agalactiae* (KLL29182.1). In the two former bacteria, the corresponding protein has been annotated as queuosine precursor ECF transporter S-component QueT. Therefore, protein OG1RF_11486 could be involved in the uptake of a queuosine biosynthetic intermediate. Using the BLASTP program[20], we found that the OG1RF genome encodes an additional QueT transporter family protein (OG1RF_12031; 168 residues; AEA94718.1). It has 55% of similarity to the OG1RF_11486 protein.

MafR activates the *P11486* promoter *in vivo*. By qRT-PCR assays, we found that MafR has a positive effect on the transcription of *OG1RF_11486*. Compared to strain OG1RFΔ*mafR*, the relative expression of *OG1RF_11486* was slightly higher in strain OG1RF ($\log_2$FC ∼0.9). Moreover, the relative expression of *OG1RF_11486* was higher in strain OG1RFΔ*mafR* harbouring pDLF*mafR* (plasmid-encoded MafR) than in strain OG1RFΔ*mafR* harbouring pDLF ($\log_2$FC ∼2.4).

The BPROM program (*Softberry, Inc.*) predicts a promoter sequence (named *P11486* herein) upstream of the *OG1RF_11486* gene. The −35 (**TT**T**ACA**) and −10 (**TA**AC**AT**) elements of this promoter are separated by 17 nucleotides (Fig. 4A). By primer extension using total RNA from OG1RF cells, we demonstrated that the *P11486* promoter is functional *in vivo* (Fig. 5). Oligonucleotide R*11486*-D was used as primer (Table 1). A cDNA product of 130 nucleotides was detected, indicating that transcription of *OG1RF_11486* starts at coordinate 1543115 (Fig. 4A).

To further characterize the *P11486* promoter, we constructed several transcriptional fusions (Fig. 6). A 284-bp DNA fragment (coordinates 1542902 to 1543185) was inserted into pASTT. The recombinant plasmid (pASTT-*P11486*) was first introduced into OG1RF and OG1RFΔ*mafR*. In both strains, *gfp* expression (1.48 ± 0.10 and 1.51 ± 0.16 units, respectively) was ∼4-fold higher than the basal level (OG1RF harbouring pASTT). This result indicated that the 284-bp DNA fragment has promoter activity, however, the chromosomal copy of *mafR* is not sufficient to activate such a promoter located on pASTT (multicopy plasmid). Next, we introduced pASTT-*P11486* into OG1RFΔ*mafR* harbouring pDLF*mafR* (plasmid-encoded MafR). In this strain, *gfp* expression was ∼3-fold higher than in the control strain (OG1RFΔ*mafR* harbouring pDLF) (Fig. 6). Similar results were obtained with plasmids pASTT-*P11486Δ*66 and pASTT-*P11486Δ145*, which allowed us to conclude that the 139-bp region between coordinates 1543047 and 1543185 contains both the *P11486* promoter and the site required for its activation by MafR. A further deletion analysis showed that sequences between coordinates 1543047 and 1543071 (plasmid pASTT-*P11486Δ169*) are needed for MafR-mediated activation of the *P11486* promoter but not for promoter activity. Moreover, deletion of the region that spans coordinates 1543071 and 1543090 (plasmid pASTT-*P11486Δ188*) removes the −35 element of the *P11486* promoter and, consequently, reduces the expression of *gfp* to basal levels (Fig. 6).

MafR binds to the *P11486* promoter region *in vitro*. By DNase I footprinting assays, we analysed whether MafR-His binds to the *P11486* promoter region (Fig. 7). We used a 275-bp DNA fragment (coordinates 1542969 to 1543243), which contains both the *P11486* promoter and the site required for its activation by MafR *in vivo* (Fig. 6). On the coding strand and at 350 nM of MafR-His, changes in DNase I sensitivity (diminished cleavages) were observed within the region spanning coordinates 1543047 and 1543110. On the non-coding strand and at 300 nM of MafR-His, diminished cleavages were observed between coordinates 1543043 and 1543110. On both strands and at 400 nM of MafR-His, regions protected against DNase I digestion were observed along the DNA fragment, which is consistent with the ability of MafR-His to generate multimeric complexes[4]. Therefore, MafR-His recognizes preferentially a DNA site overlapping the *P11486* core promoter. Such a DNA site includes sequences needed for MafR-mediated activation of the *P11486* promoter *in vivo* (Fig. 6). According to the bendability/curvature propensity plot of the 275-bp DNA fragment, the MafR binding site contains regions of potential bendability (Fig. 4B).

Discussion

Gene regulation plays a key role during bacterial adaptation to environmental fluctuations. The ability of enterococci to metabolize numerous carbohydrates enables them to colonize diverse environments[1]. Our previous work showed that MafR activates, directly or indirectly, the transcription of numerous genes on a genome-wide scale. Many of such genes encode proteins involved in transport or metabolism of carbon sources[5]. Now, by qRT-PCR, transcriptional fusions and DNase I footprinting experiments, we have demonstrated that MafR functions as a transcription activator. It activates directly the transcription of the *OG1RF_12294* and *OG1RF_11486* genes. Gene *OG1RF_12294* encodes a protein that has sequence similarity to several eukaryotic and prokaryotic proteins characterized as calcium P-type ATPases (Supplementary Table S1). This finding suggests that MafR could have a regulatory role in maintaining cellular calcium homeostasis. Calcium ions are known to affect different physiological processes in prokaryotic organisms, such as division, secretion, transport, and stress response[30]. Gene *OG1RF_11486* encodes a putative ECF transporter S-component, likely involved in the uptake of a queuosine precursor. Thus, MafR could have an additional regulatory role in the biosynthesis of queuosine, a modified nucleoside found at the wobble position of particular transfer RNAs[31]. There is evidence that queuosine contributes

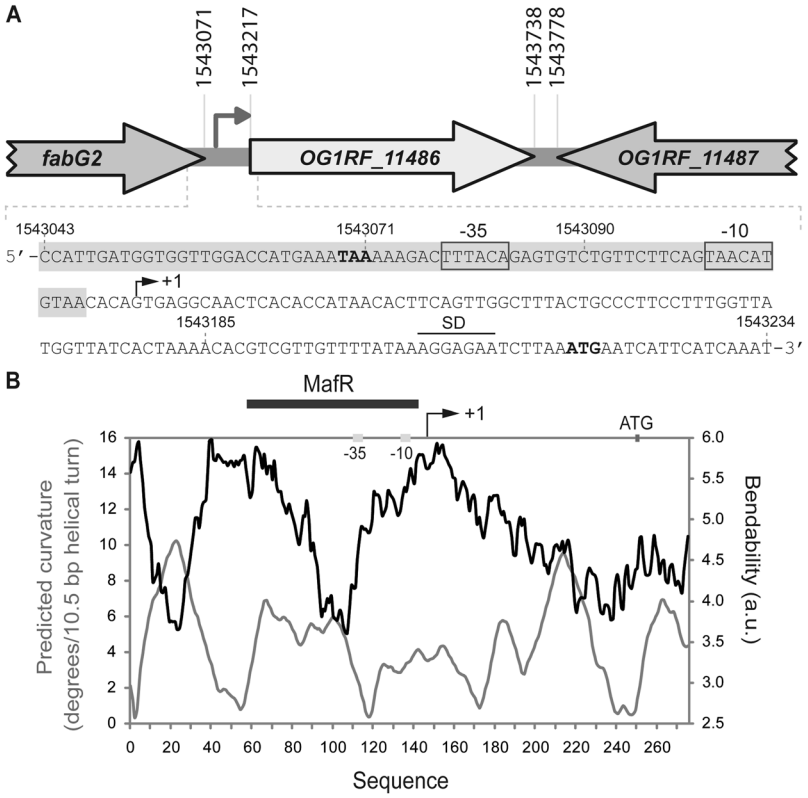


Figure 4. Relevant features of the *P11486* promoter region. (**A**) Genetic organization of the chromosome region that contains *OG1RF_11486*. Coordinates of the translation start and stop codons are indicated. The arrow upstream of the *OG1RF_11486* gene represents its promoter. The nucleotide sequence of the region spanning coordinates 1543043 to 1543234 is shown. The stop codon (TAA) of *fabG2* and the start codon (ATG) of *OG1RF_11486* are indicated in boldface letters. SD: Shine-Dalgarno sequence. The transcription start site (+1 position) of the *OG1RF_11486* gene, and the main sequence elements (−35 box and −10 box) of the *P11486* promoter identified in this work are indicated. The MafR binding site defined in this work is shown (shadowed box). Genes *OG1RF_11486* and *OG1RF_11487* correspond to genes *EF1774* and *EF1775* in *E. faecalis* strain V583. (**B**) Bendability/curvature propensity plot of the region spanning coordinates 1542969 to 1543243. The location of the *P11486* core promoter, the start codon of *OG1RF_11486* and the MafR binding site are indicated.

to the efficiency of protein synthesis. In *Shigella flexneri*, the intracellular concentration of the virulence-related transcriptional regulator VirF is reduced in the absence of queuosine[32]. Moreover, it has been reported that the lack of queuosine affects the growth of some bacteria under stress conditions[33,34].

Bacteria use a variety of mechanisms to activate transcription from specific promoters. Genetic and biochemical studies have shown that some proteins stimulate transcription by binding to a specific DNA site either upstream of or overlapping the core promoter[35]. By DNase I footprinting experiments, we have found that MafR recognizes a site overlapping the *P12294* core promoter, as well as a site overlapping the *P11486* core promoter (this work). These results suggest that MafR might enhance the efficiency of both promoters by recruitment of RNA polymerase through direct interactions with the sigma factor. In addition, MafR might induce conformational changes in the target promoters, as it has been described for some transcription activators[35]. Transcriptional activation from specific promoters has also been reported for other members of the Mga/AtxA family. The pneumococcal Mga*Spn* regulator stimulates transcription of a four-gene operon (*spr1623-spr1626*) by binding to a specific DNA site upstream of the promoter (positions −60 to −99)[12]. Regarding the Mga regulator from *S. pyogenes*, the position of its DNA-binding site with respect to the start of transcription varies among the promoters tested. Nevertheless, the majority of the promoters contain an Mga binding site located around position −54, thereby overlapping the −35 element of the promoter[8].

Simple protein-DNA recognition mechanisms do not exist[36]. Based on the structures of various protein-DNA complexes, Rohs *et al*. proposed that particular proteins use likely a combination of readout mechanisms: base readout and shape readout[6]. The DNA sites recognized by MafR on the *P12294* and *P11486* promoters have a low sequence identity: they share the **GG**(C/A)**C**(A/C)(C/A)**TGAAAT**(T/A)**A** sequence element (Fig. 8A). Moreover, both MafR binding sites contain regions of potential bendability (Figs 1B and 4B). We have also shown that MafR recognizes a DNA site upstream of the *Pma* promoter (positions −69 to −104)[4]. The function of this interaction remains unknown. Such a MafR binding site is adjacent to the peak of a potential intrinsic curvature[4] and shares a short DNA sequence motif (**TGA**T**AT**) with the two MafR binding sites identified in this work (Fig. 8B).

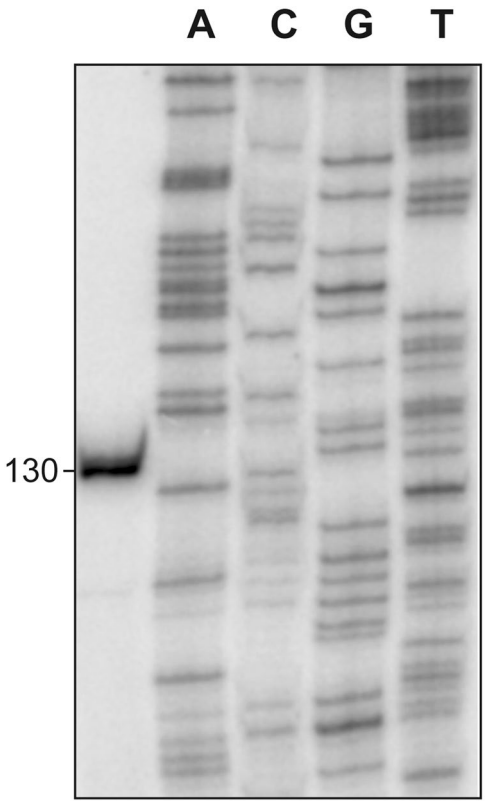


Figure 5. Transcription initiation site of the *OG1RF_11486* gene. Primer extension reactions were carried out using total RNA from OG1RF cells. Oligonucleotide R*11486*-D (coordinates 1543222-1543243) was used as primer. The size (in nucleotides) of the cDNA product is indicated on the left of the gel. Dideoxi-mediated chain termination sequencing reactions were used as DNA size markers (lanes A, C, G, T).

Therefore, MafR does not seem to recognize a specific nucleotide sequence. Several findings suggest that recognition of particular DNA shapes could be a characteristic of the global regulators that constitute the Mga/AtxA family. Mga*Spn* from *S. pneumoniae* recognizes a DNA site upstream of the *P1623B* promoter (positions −60 to −99), as well as a DNA site overlapping the *Pmga* promoter (positions −23 to +21)[12]. The former interaction enhances the efficiency of the promoter[11], whereas the function of the latter remains unknown. Such Mga*Spn* binding sites have a low sequence identity and, according to predictions, they contain an intrinsic curvature flanked by regions of bendability[12]. Furthermore, Mga*Spn* was shown to have a preference for AT-rich DNA regions[13]. Concerning Mga from *S. pyogenes*, several DNA-binding sites have been identified. These sites exhibit a low sequence identity (13.4%)[37], although a consensus Mga binding sequence was initially proposed[38]. In the case of AtxA from *B. anthracis*, *in vitro* protein-DNA interaction studies have not been reported. Nevertheless, sequence similarities are not apparent in its target promoter regions, and some of them are intrinsically curved[39].

In conclusion, our study shows for the first time that MafR is a transcription activator. It stimulates transcription from the *P12294* and *P11486* promoters *in vivo*. Moreover, MafR binds *in vitro* to a specific DNA site that overlaps the −35 element of each promoter. The two MafR binding sites have a low sequence identity but share a six-base pair motif. We propose that MafR would recognize intrinsic DNA structural features rather than particular DNA sequences on its target DNAs.

Materials and Methods

Oligonucleotides, bacterial strains, and plasmids. Oligonucleotides used in this work are listed in Table 1. *E. faecalis* strains OG1RF[14] and OG1RFΔ*mafR*[5] were used. Plasmids pDLF (expression vector) and pDLF*mafR* were described[5]. These plasmids carry a kanamycin resistance gene. Plasmid pASTT (D. García-Rincón, V. Solano-Collado and A. Bravo, unpublished results) is based on the pAST promoter-probe vector[40], which carries a tetracycline resistance gene. Plasmid pASTT carries the *TrsiV* transcriptional terminator[40] downstream of the *gfp* reporter gene. The following pASTT-derivatives were constructed in this work. In all cases, a region of the OG1RF chromosome was amplified by PCR using the indicated primers, digested with *Sac*I, and inserted into pASTT: pASTT-*P12294* (primers F*12294* and R*12294*, 260-bp restriction fragment), pASTT-*P12294Δ-10* (primers F*12294* and R*12294Δ-10*, 236-bp restriction fragment), pASTT-*P12294Δ69* (primers F*12294Δ69* and R*12294*, 192-bp restriction fragment), pASTT-*P12294Δ208* (primers F*12294Δ208* and R*12294*, 53-bp restriction fragment), pASTT-*P11486* (primers F*11486* and R*11486*, 290-bp restriction fragment), pASTT-*P11486Δ*66 (primers F*11486Δ66* and R*11486*, 224-bp restriction fragment), pASTT-*P11486Δ145* (primers F*11486Δ145* and R*11486*, 145-bp restriction fragment), pASTT-*P11486Δ169* (primers F*11486Δ169* and

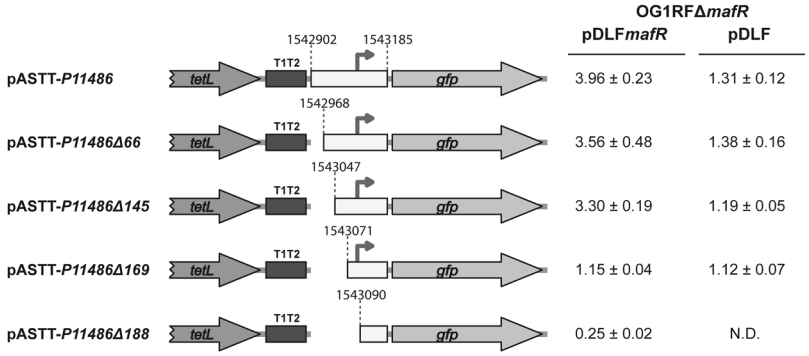


Figure 6. Effect of plasmid-encoded MafR on the activity of the *P11486* promoter. Five regions from the OG1RF chromosome were inserted independently into the *Sac*I site of pASTT. The coordinates of such regions are indicated. Gene *tetL*: tetracycline resistance determinant. Gene *gfp*: green fluorescent protein. The T1T2 box represents the tandem transcriptional terminators T1 and T2 of the *Escherichia coli rrnB* rRNA operon. The arrow represents the −35 element of the *P11486* promoter. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} = 0.4). In each case, three independent cultures were analysed. N.D.: non-determined.

R*11486*, 121-bp restriction fragment), pASTT-*P11486Δ188* (primers F*11486Δ188* and R*11486*, 102-bp restriction fragment).

Growth and transformation of bacteria. *E. faecalis* was grown in BHI medium, which was supplemented with tetracycline (4 μg/ml) and/or with kanamycin (250 μg/ml) when strains carrying plasmids were used. Experiments were performed at 37 °C without aeration. The protocol used to transform *E. faecalis* by electroporation was described[41].

DNA and RNA isolation. Genomic DNA was prepared using the Bacterial Genomic Isolation Kit (Norgen Biotek Corporation). Plasmid DNA was prepared using the High Pure Plasmid Isolation Kit (Roche Applied Science) as described[5]. Total RNA was isolated using the RNeasy mini Kit (QIAGEN). In general, bacteria were grown to an optical density at 650 nm (OD_{650}) of 0.4 (logarithmic growth phase). For stationary phase, bacteria were grown to an OD_{650} of 0.8 and then incubated for two hours at the same temperature. Then, cultures were processed as reported[5]. The integrity of rRNAs was analysed by agarose gel electrophoresis. RNA concentration was determined using a NanoDrop ND-2000 Spectrophotometer.

Polymerase chain reaction (PCR). The Phusion High-Fidelity DNA polymerase (Thermo Scientific) and the Phusion HF buffer were used. Reaction mixtures (50 μl) contained 5–30 ng of template DNA, 20 pmol of each primer, 200 μM each deoxynucleoside triphosphate (dNTP), and one unit of DNA polymerase. PCR conditions were reported[40]. To amplify the 270-bp DNA fragment (promoter *P12294*) used in footprinting experiments, the Phusion GC buffer was used. In this case, reaction mixtures were supplemented with 7% DMSO and the annealing step was performed at 59 °C. PCR products were purified with the QIAquick PCR purification kit (QIAGEN).

Quantitative RT-PCR (qRT-PCR). For cDNA synthesis with random primers, the iScript Select cDNA Synthesis kit (Bio-Rad) was used as described[5]. Quantitative PCRs were performed using the iQ SYBR Green Supermix (Bio-Rad) and a iCycler Thermal Cycler (Bio-Rad) as reported[5]. Forward (F*gene*-q) and reverse (R*gene*-q) primers used in the quantitative PCRs are listed in Table 1. Relative quantification of gene expression was performed using the comparative C_T method[15] as described[5]. Except for gene *mafR*, the internal control gene was *recA* (OG1RF_12439; recombination protein RecA). In the case of *mafR*, the internal control gene was *zwf* (OG1RF_10737; glucose-6-phosphate 1-dehydrogenase) because its expression level was similar at the logarithmic and stationary growth phases.

Primer extension. Oligonucleotide R*11486*-D was radioactively labelled at the 5′-end using [γ-^{32}P]-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs) as reported[12]. Primer extension reactions (20 μl) contained 1.2 pmol of ^{32}P-labelled oligonucleotide and 5 μg of total RNA isolated from strain OG1RF. The ThermoScript Reverse Transcriptase enzyme (Invitrogen) was used. Reactions were incubated at 55 °C for 45 min. After heating at 85 °C for 5 min, samples were ethanol precipitated and dissolved in loading buffer (80% formamide, 1 mM EDTA, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol). cDNA products were analysed by sequencing gel (8 M urea, 6% polyacrylamide) electrophoresis. Dideoxy-mediated chain termination sequencing reactions were run in the same gel. Labelled products were visualized using a Fujifilm Image Analyser FLA-3000.

Fluorescence assays. Plasmid-carrying cells were grown to an OD_{650} of 0.4 (logarithmic phase). Then, different volumes of culture (0.4 to 1 ml) were centrifuged, and cells were resuspended in 200 μl of

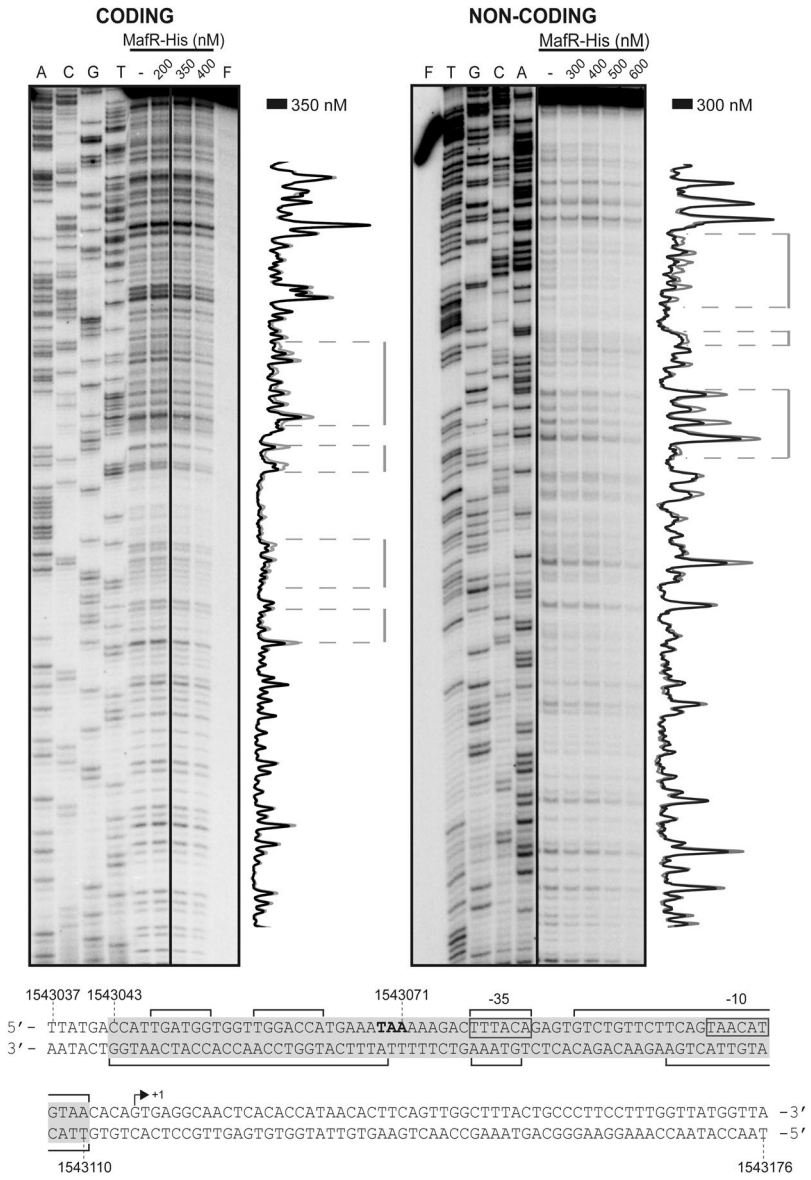


Figure 7. DNase I footprints of complexes formed by MafR-His on the 275-bp DNA fragment that contains the *P11486* promoter. ^{32}P-labelled DNA (4 nM) was incubated with the indicated concentrations of MafR-His and then it was digested with DNase I. Non-digested DNA (F) and dideoxy-mediated chain termination sequencing reactions (A, C, G, T) were run in the same gel. All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to free DNA (grey line) and DNA with protein (black line) are shown. The nucleotide sequence of the region spanning coordinates 1543037 to 1543176 is shown. The transcription initiation site (+1 position) of *OG1RF_11486* is shown. The −35 and −10 elements of the *P11486* promoter are indicated. Brackets indicate regions protected against DNase I digestion. The site recognized by MafR-His (coordinates 1543043-1543110) is indicated with a grey box.

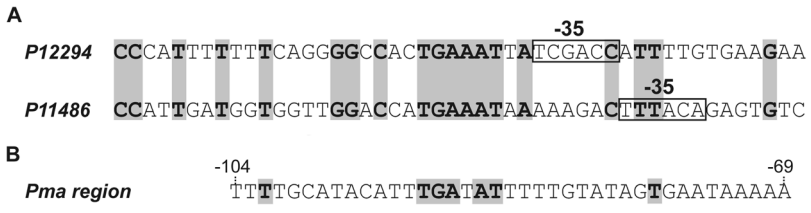


Figure 8. DNA sites recognized by MafR. (**A**) Nucleotide sequence alignment of the DNA sites recognized by MafR on the *P12294* and *P11486* promoter regions. Identical nucleotides are highlighted in grey boxes. (**B**) Nucleotide sequence of the DNA site recognized by MafR on the *Pma* promoter region (positions −69 to −104)[4]. Nucleotides shared with the MafR binding sites shown in (**A**) are highlighted in grey boxes.

Name	Sequence (5′ to 3′)[a]
F*mafR*-q	ACTTTATCAACCGTCCTTGG
R*mafR*-q	GTTTCGCCATAGACATTATC
F*zwf*-q	CGGTCAAGGGTTCAATACAACT
R*zwf*-q	CCAAGATTGGGCAACTTCGTCCCA
F*12294*-q	TCCTCTACCGTTGACACCTG
R*12294*-q	TGCCTTCGTTGACATCTCTTG
F*recA*-q	GCAACGAAATGGTGGAACAG
R*recA*-q	AAGGCATCGGCAATCTCTAAG
F*10600*-q	GCGTAGAAGAGTCAGCACTA
R*10600*-q	GCCATTCACAACGGTACAGC
F*11602*-q	CAACACCTCATTAGCGAAAC
R*11602*-q	GTCAATCATACCGACTAAACCA
F*11486*-q	TGGTTACCGCTTTGTATGTTG
R*11486*-q	CCCTAACGTAATGGACCAGAT
F*12294*	GAAACAGCGTTGA**G**CTCTTCTAGTGAC
R*12294*	CATTTCGTGTACCTCCGA**G**CTCCTTGATACCT
F*12294Δ69*	GTAAAAATGGTGAAA**G**A**G**CT**C**ATGTCAAAGCGT
F*12294Δ208*	CAGGGGCCACTGA**G**CT**C**ATCGACCATT
R*12294Δ-10*	CCTTTATTATAGAG**CTC**ATTTTCTTCACA
F*11486*	ACACCCATGAACG**G**AGCT**C**ATTTTGTA
R*11486*	ATAAAACAAC**G**AGCT**C**TTTTAGTGATAACC
F*11486Δ66*	GGGCCGTTG**A**GC**T**CAGCCACAGGAAGTA
F*11486Δ145*	GGCACAGTTATGAGC**T**CTGATGGTGGT
F*11486Δ169*	GTTGGACCATGAGCTCAAAAGACTTTACA
F*11486Δ188*	GACTTTACAGAG**CTC**CTGTTCTTCAGTA
F*12294*-D	GATGTCAAAGCGTTAATTGGCA
R*12294*-D	GACCCGTTTGCTTCGTCTTAGT
F*11486*-D	GCCACAGGAAGTAGCAAAACT
R*11486*-D	GGTTTGTGGATTTGATGAATGA

Table 1. Oligonucleotides used in this work. [a]Restriction sites are underlined, and the base changes that generate restriction sites are in bold.

phosphate-buffered saline (PBS). In each case, three independent cultures were analysed. Fluorescence intensity was measured using a Thermo Scientific Varioskan Flash instrument (excitation at 488 nm and emission at 515 nm). The fluorescence corresponding to 200 µl of PBS buffer without cells was ~0.03 arbitrary units.

Purification of MafR-His. The procedure to overproduce and purify a His-tagged $MafR_{OG1RF}$ protein (herein MafR-His) was reported[4]. MafR-His carries the Leu-Glu-6xHis peptide (His-tag) fused to its C terminus. Protein concentration was determined using a NanoDrop ND-2000 Spectrophotometer (Thermo Scientific).

DNase I footprinting assays. Oligonucleotides were ^{32}P-labelled at the 5′-end as described[12]. ^{32}P-labelled oligonucleotides were used for PCR amplification to obtain double-stranded DNA fragments labelled at either the coding or the non-coding strand. Two regions of the OG1RF chromosome were amplified: a 270-bp region (coordinates 2425817-2425548) using the F*12294*-D and R*12294*-D oligonucleotides, and a 275-bp region (coordinates 1542969-1543243) using the F*11486*-D and R*11486*-D oligonucleotides. Binding reactions (8 µl) contained 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 1 mg/ml BSA, 1.25% glycerol, 0.25 mM EDTA, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 2–4 nM ^{32}P-labelled DNA and different concentrations of MafR-His (100 to 600 nM). Reaction mixtures were incubated at room temperature for 20 min. Then, 0.015 units of DNase I (Roche Applied Science) was added and the reaction proceeded for 5 min at the same temperature. DNase I digestion was stopped by adding 1 µl of 250 mM EDTA. Then, 4 µl of loading buffer (80% formamide, 1 mM EDTA, 10 mM NaOH, 0.1% bromophenol blue and 0.1% xylene cyanol) was added. Samples were heated at 95 °C for 5 min and loaded onto sequencing gels (6% polyacrylamide, 8 M urea). Dideoxy-mediated chain termination sequencing reactions were run in the same gel. Labelled products were visualized using a Fujifilm Image Analyser FLA-3000. The intensity of the bands was quantified using the Quantity One software (Bio-Rad).

***In silico* prediction of intrinsic curvature.** The bendability/curvature propensity plots were calculated with the bend.it server[26] (http://hydra.icgeb.trieste.it/dna/bend_it.html) as described previously[12].

References

1. Ramsey, M., Hartke, A. & Huycke, M. The Physiology and Metabolism of Enterococci. In: Gilmore MS, Clewell DB, Ike Y, *et al.* editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection (Internet)*. Boston, Massachusetts Eye and Ear Infirmary, 1–43 (2014).
2. Kim, S., Covington, A. & Pamer, E. G. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. *Immunol. Rev.* **279**, 90–105 (2017).

3. Beganovic, M. *et al*. A review of combination antimicrobial therapy for *Enterococcus faecalis* bloodstream infections and infective endocarditis. *Clin. Infect. Dis.* **67**, 303–309 (2018).
4. Ruiz-Cruz, S., Moreno-Blanco, A., Espinosa, M. & Bravo, A. DNA-binding properties of MafR, a global regulator of *Enterococcus faecalis*. *FEBS Lett.* **592**, 1412–1425 (2018).
5. Ruiz-Cruz, S., Espinosa, M., Goldmann, O. & Bravo, A. Global regulation of gene expression by the MafR protein of *Enterococcus faecalis*. *Front. Microbiol.* **6**, 1521 (2016).
6. Rohs, R. *et al*. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
7. Abe, N. *et al*. Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307–318 (2015).
8. Hondorp, E. R. & McIver, K. S. The Mga virulence regulon: infection where the grass is greener. *Mol. Microbiol.* **66**, 1056–1065 (2007).
9. McIver, K. S. & Myles, R. L. Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Mol. Microbiol.* **43**, 1591–1601 (2002).
10. Vahling, C. M. & McIver, K. S. Domains required for transcriptional activation show conservation in the Mga family of virulence gene regulators. *J. Bacteriol.* **188**, 863–873 (2006).
11. Solano-Collado, V., Espinosa, M. & Bravo, A. Activator role of the pneumococcal Mga-like virulence transcriptional regulator. *J. Bacteriol.* **194**, 4197–4207 (2012).
12. Solano-Collado, V., Lurz, R., Espinosa, M. & Bravo, A. The pneumococcal Mga*Spn* virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. *Nucleic Acids Res.* **41**, 6975–6991 (2013).
13. Solano-Collado, V., Hüttener, M., Espinosa, M., Juárez, A. & Bravo, A. Mga*Spn* and H-NS: Two unrelated global regulators with similar DNA-binding properties. *Front. Mol. Biosci.* **3**, 60 (2016).
14. Bourgogne, A. *et al*. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* **9**, R110 (2008).
15. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).
16. Palmgren, M. G. & Nissen, P. P-Type ATPases. *Annu. Rev. Biophys.* **40**, 243–266 (2011).
17. Rudolph, H. K. *et al*. The yeast secretory pathway is perturbed by mutations in PMR1, a member of a Ca^{2+} ATPase family. *Cell* **58**, 133–145 (1989).
18. Sorin, A., Rosas, G. & Rao, R. PMR1, a Ca^{2+} -ATPase in yeast Golgi, has properties distinct from sarco/endoplasmic reticulum and plasma membrane calcium pumps. *J. Biol. Chem.* **272**, 9895–9901 (1997).
19. Van Baelen, K., Vanoevelen, J., Missiaen, L., Raeymaekers, L. & Wuytack, F. The Golgi PMR1 P-type ATPase of *Caenorhabditis elegans*: Identification of the gene and demonstration of calcium and manganse transport. *J. Biol. Chem.* **276**, 10683–10691 (2001).
20. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
21. Raeymaekers, L., Wuytack, E. Y., Willems, I., Michiels, C. W. & Wuytack, F. Expression of a P-type Ca^{2+}-transport ATPase in *Bacillus subtilis* during sporulation. *Cell Calcium* **32**, 93–103 (2002).
22. Rosch, J. W., Sublett, J., Gao, G., Wang, Y.-D. & Tuomanen, E. I. Calcium efflux is essential for bacterial survival in the eukaryotic host. *Mol. Microbiol.* **70**, 435–444 (2008).
23. Faxén, K. *et al*. Characterization of a *Listeria monocytogenes* Ca^{2+} pump: a SERCA-type ATPase with only one Ca^{2+}-binding site. *J. Biol. Chem.* **286**, 1609–1617 (2011).
24. Geisler, M., Richter, J. & Schumann, J. Molecular cloning of a P-type ATPase gene from the cyanobacterium *Synechocystis* sp. PCC 6803. Homology to eukaryotic Ca^{2+} -ATPases. *J. Mol. Biol.* **234**, 1284–1289 (1993).
25. Hein, K. L., Nissen, P. & Morth, J. P. Purification, crystallization and preliminary crystallographic studies of a PacL homologue from *Listeria monocytogenes*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68**, 424–427 (2012).
26. Vlahovicek, K., Kaján, L. & Pongor, S. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* **31**, 3686–3687 (2003).
27. Rodionov, D. A. *et al*. A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **191**, 42–51 (2009).
28. Slotboom, D. J. Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. *Nat. Rev. Microbiol.* **12**, 79–87 (2014).
29. Majsnerowska, M., Ter Beek, J., Stanek, W. K., Duurkens, R. H. & Slotboom, D. J. Competition between different S-components for the shared energy coupling factor module in energy coupling factor transporters. *Biochemistry* **54**, 4763–4766 (2015).
30. Domínguez, D. C., Guragain, M. & Patrauchan, M. Calcium binding proteins and calcium signaling in prokaryotes. *Cell Calcium* **57**, 151–165 (2015).
31. Hutinet, G., Swarjo, M. A. & de Crécy-Lagard, V. Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. *RNA Biol.* **14**, 1175–1184 (2017).
32. Durand, J. M. B., Dagberg, B., Uhlin, B. E. & Björk, G. R. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. *Mol. Microbiol.* **35**, 924–935 (2000).
33. Noguchi, S., Nishimura, Y., Hirota, Y. & Nishimura, S. Isolation and characterization of an *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J. Biol. Chem.* **257**, 6544–6550 (1982).
34. Thibessard, A. *et al*. Identification of *Streptococcus thermophilus* CNRZ368 genes involved in defense against superoxide stress. *Appl. Environ. Microbiol.* **70**, 2220–2229 (2004).
35. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
36. Siggers, T. & Gordân, R. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res.* **42**, 2099–2111 (2014).
37. Hause, L. L. & McIver, K. S. Nucleotides critical for the interaction of the *Streptococcus pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. *J. Bacteriol.* **194**, 4904–4919 (2012).
38. McIver, K. S., Heath, A. S., Green, B. D. & Scott, J. R. Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. *J. Bacteriol.* **177**, 6619–6624 (1995).
39. Hadjifrangiskou, M. & Koehler, T. M. Intrinsic curvature associated with the coordinately regulated anthrax toxin gene promoters. *Microbiology* **154**, 2501–2512 (2008).
40. Ruiz-Cruz, S., Solano-Collado, V., Espinosa, M. & Bravo, A. Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. *J. Microbiol. Methods* **83**, 156–163 (2010).
41. Shepard, B. D. & Gilmore, M. S. Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. *Methods Mol. Biol.* **47**, 217–226 (1995).

Acknowledgements

Thanks are due to Dr. Virtu Solano-Collado and Daniel García-Rincón for providing the pASTT plasmid, and to Verónica Navarro for her excellent technical assistance. This work was supported by grants BIO2016-76412-C2-2-R (AEI/FEDER, UE) and BIO2015-69085-REDC from the Spanish Ministry of Economy and Competitiveness.

Author Contributions

S.R.-C. and A.M.-B. performed experiments. S.R.-C., M.E. and A.B. designed the study and wrote the manuscript. All authors read and approved the final manuscript.

"No sé para dónde voy

y aunque sí de donde vengo...

me gusta lo que estoy viendo,

me quedo a vivir en hoy,

sólo puedo ser quien soy

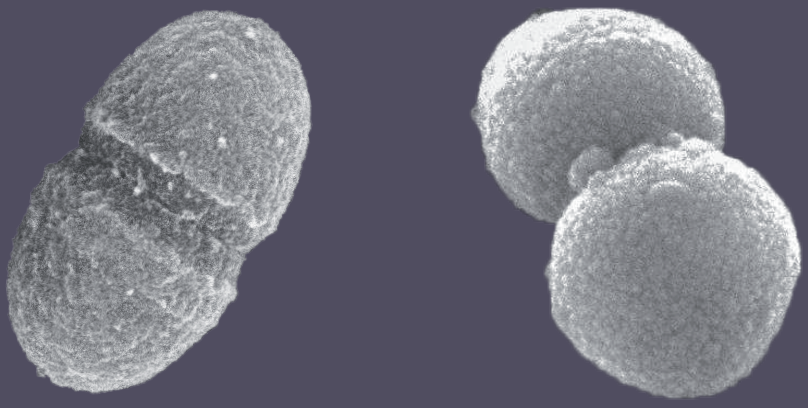
por donde ande abierto el cuento.

Ayer me fui con el viento,

mañana ya habré cambiado,

soy un bicho enamorado

de este pequeño momento"



CSIC

Consejo Superior de Investigaciones Científicas

CIB Margarita Salas