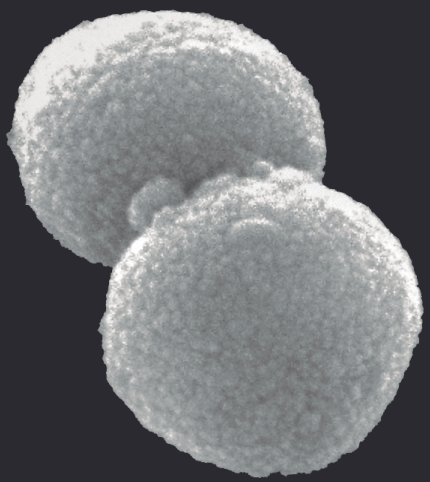
Tesis doctoral
Doctoral Thesis

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular I

Caracterización molecular del regulador transcripcional Mga*Spn* de *Streptococcus pneumoniae*



María Virtudes Solano Collado

Madrid, 2014

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I

TESIS DOCTORAL

CARACTERIZACIÓN MOLECULAR DEL REGULADOR TRANSCRIPCIONAL Mga*Spn* DE *Streptococcus pneumoniae*

María Virtudes Solano Collado

Memoria presentada para optar al grado de Doctor con Mención Europea por la Universidad Complutense de Madrid

Bajo la dirección de la doctora

Alicia Bravo García

Centro de Investigaciones Biológicas
CSIC

Madrid, 2014

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I

DOCTORAL THESIS

MOLECULAR CHARACTERIZATION OF THE Mga*Spn* TRANSCRIPTIONAL REGULATOR OF *Streptococcus pneumoniae*

María Virtudes Solano Collado

Dissertation submitted for the Degree of Doctor of Philosophy by the Universidad Complutense de Madrid

Supervised by

Alicia Bravo García

Centro de Investigaciones Biológicas
CSIC

Madrid, 2014

El trabajo recogido en la presente memoria ha sido realizado por María Virtudes Solano Collado bajo la dirección de la Dra. Alicia Bravo García, en el Departamento de Microbiología Molecular y Biología de las Infecciones del Centro de Investigaciones Biológicas (CIB) del Consejo Superior de Investigaciones Científicas (CSIC) con financiación concedida por el Ministerio de Educación y Ciencia (Proyecto BFU2006-08487 y Beca/Contrato FPI BES-2007-17086), la Comunidad Autónoma de Madrid/Consejo Superior de Investigaciones Científicas (CCG08-CSIC/SAL-3694), y el Ministerio de Ciencia e Innovación (CSD2008-00013-INTERMODS, BFU2009-11868).

The work presented in this Dissertation (PhD) has been done by María Virtudes Solano Collado and supervised by Dr. Alicia Bravo García. The work was performed in the Department of Molecular Microbiology and Infection Biology at the Centro de Investigaciones Biológicas (CIB-CSIC) and it was supported by grants from the Spanish Ministry of Education and Science (BFU2006-08487 and fellowship FPI BES-2007-17086), the Community of Madrid/Spanish National Research Council (CCG08-CSIC/SAL-3694) and the Spanish Ministry of Science and Innovation (CSD2008-00013-INTERMODS, BFU2009-11868).

Opta al grado de Doctor **VºBº Directora de Tesis**

María Virtudes Solano Collado **Alicia Bravo García**

A mi hermano, mi madre y mi abuela.

"Siempre parece imposible...hasta que se hace"

Nelson Mandela

Agradecimientos

"Muchas veces, a lo largo de un mismo día, me doy cuenta que mi propia vida y sus logros se han construido gracias al trabajo de las personas que me rodean. También comprendo, con cuanta seriedad debo esforzarme para darles, en correspondencia, tanto como he recibido. Albert Einstein*."*

Al llegar el final de una etapa, no puedes dejar de agradecer y recordar a todos aquellos que la hicieron posible, a los que estuvieron animándome y apoyándome durante todo el trayecto. Todas y cada una de las personas que forman o han formado parte de mi vida han contribuido de alguna manera a que sea lo que soy, a que terminase estudiando Biología y haciendo esta tesis. Por todo lo que habéis hecho por mí, quisiera dedicar este trabajo:

A Alicia y Manolo, por confiar en mi desde el principio y darme la oportunidad de formar parte de un grupo de investigación tan especial. Nunca olvidaré aquella primera visita y la ilusión con la que salí del CIB. Alicia, gracias por tu paciencia, por sentarte conmigo y guiarme en los primeros pasos que he dado en el mundo de la ciencia, por tu capacidad para animarme cuando las cosas no salían como quería, por no permitirme tirar nunca la toalla y por transmitirme tu pasión por este mundo cada vez más complicado. Manolo, gracias por todos tus consejos e implicación con este trabajo, por retarme y crear en mi la necesidad de superarme cada día un poco más, por todo lo que me has enseñado a lo largo de estos años. Sobre todo, gracias a los dos por ser dos grandes personas, por cuidarme tanto y por haberme hecho sentir especial. Gracias por haber hecho posible esta tesis. Gracias por haberme enseñado a pensar.

A los grandes amigos que he encontrado en el CIB: a Fabi, por haberme ayudado siempre con una sonrisa, porque aprender a tu lado siempre fue fácil, por todos los consejos que me han ahorrado muchos calentamientos de cabeza, por elevar el compañerismo al máximo nivel y por tu buen humor aceptando nuestras locuras y participando en cada fiesta de primavera; a Gloria, Jose y Marta, por todo vuestro apoyo y por estar siempre dispuestos a echarme una mano; a Ramón, por las grandes conversaciones que me regalas (no sólo de ciencia) y por todo el cariño que me demuestras; a Wai Ting, porque tu estancia en el labo me ha enseñado muchas cosas. Lore, aunque no hayamos podido terminar juntas esta etapa quiero darte las gracias por haberme ayudado tanto, porque hacías que purificar proteína o plásmidos fuese algo divertido. Concha, esa persona a la que todos recurrimos cuando tenemos cualquier duda, gracias por ser profe y amiga, por tus buenos consejos tanto dentro como fuera del laboratorio y por tu buena disposición siempre para todo; a Nieves, por haber compartido conmigo los primeros años de esta etapa y haberme ayudado siempre que lo necesité; Eli y Benet, gracias por todos los momentos que hemos pasado juntos (todos buenos, por cierto) y por los sitios que hemos compartido, espero que sigamos viéndonos por el mundo; Dami, Laura, Montse, Nuria, Cristina, Miguel y Srdja, gracias por hacerme reír y crear una atmósfera de confianza en la que es muy fácil sentirse a gusto. Es imposible no cogeros cariño.

No puedo dejar de hacer un párrafo especial para dar las gracias esas personas que han logrado hacer de cada día un día especial, que hacen que te apetezca ir a trabajar cada mañana, las que hacen del CIB un lugar mágico para mí: Ana, gracias por haber estado ahí desde el principio y haber sido testigo de mis primeros logros y mis primeras decepciones, por hacer entrañables los viajes en metro por las mañanas, por enseñarme todo lo que sabes (los footprints dan fe de ello) y estar ahí siempre que necesito una consulta, por estar siempre dispuesta a ayudarme y por cuidarme con tanto esmero, tanto dentro como fuera del CIB; a Cris, Sofi, Inma y Lore B., por ser las mejores compañeras de fatigas que se pueden tener, por vuestras constantes muestras de cariño, por las risas (muchas risas) en los buenos momentos y la ayuda que siempre me ofrecéis en los no tan buenos, por la ilusión que ponéis en todo lo que hacéis y de la que me contagiáis, por vuestras grandes ideas que le arrancan una sonrisa hasta al día más gris. Cada día aprendo algo de alguna de vosotras (siempre bueno). Gracias por haber entrado a formar parte de mi vida, por ser compañeras y amigas.

Durante la realización de este trabajo se han establecido colaboraciones que han sido fundamentales para la elaboración de parte de esta tesis. I am grateful to Dr. Rudi Lurz who gave me the possibility to learn how an electron microscope works. Thanks for your patience teaching me and for your invaluable help to make an important part of this work possible. Thanks for making my stay in Berlin an extraordinary experience. Gracias también al Dr. Juan Antonio López, que ha realizado los ensayos de proteómica en el Servicio de Proteómica del CNIC.

Gracias las todas personas que trabajan en los servicios internos del CIB así como a las chicas de la limpieza y material. Sin todos ellos sería imposible el trabajo diario.

A todos los amigos de los que me separé cuando decidí embarcarme en esta aventura, en especial a Marta A., Marta G., Alberto y Kiki. Aunque la despedida fue difícil, gracias por haberme animado y apoyado desde el principio, por haberme acompañado hasta Madrid cuando vine a hacer la entrevista, por estar ahí cada día aunque cada vez nos separen más distancias. Gracias por todos los buenos ratos que pasamos juntos. A Carmen, por seguir a mi lado desde hace mucho mucho tiempo y apoyarme incondicionalmente.

A la pequeña familia con la que compartí piso los primeros años y con los que inicié la aventura de empezar de cero en un sitio nuevo: Juanfran y Benito, gracias por haberme cuidado tanto y por seguir siendo parte de mi vida; siempre recordaré de una manera muy especial el tiempo que pasamos juntos en Madrid. Creo que siempre tendré idealizado el piso de Moratines por lo a gusto que estuve allí (incluida aquella bañera rosa y la puerta ataúd). A toda la gente a la que tuve el placer de conocer y que compartió conmigo esos primeros años. A Max, por haber sido motivo de felicidad durante el tiempo que estuviste con nosotros. Mil gracias.

A Aaron, por ser un gran profesor de inglés (aunque a veces seas un poco pesao), por haberme enseñado tanto de tantas cosas y por tu valiosa ayuda en la corrección del idioma.

A Edu y Natalia, por ser dos grandes personas y ejemplos de superación constante. Gracias por haber hecho posible que descubriese una nueva pasión que me ha ayudado mucho a liberar la tensión que se acumula en el día a día. Vuestro entusiasmo se contagia y conseguís que durante el rato de las clases no piense en nada más.

Quiero dedicar también este trabajo a Juan Carlos Argüelles, profesor de la Universidad de Murcia, quien hizo que me enamorase del mundo de la microbiología, consiguió que cambiasen mis planes y me transmitió su pasión por el mundo de la investigación. Sin duda, el máximo responsable del camino que elegí.

GRACIAS a mi madre, mi hermano y mi abuela. Gracias por todo lo que hacéis por mí, porque sin vuestro apoyo nunca hubiera llegado hasta aquí. Gracias por la paciencia que tenéis conmigo, por levantarme siempre que caigo. Gracias por respetar mis decisiones, aunque muchas veces me equivoque. Gracias por ser como sois y por hacerme mejor persona. A mi tía Paqui, por estar a mi lado y enseñarme a mantenerme siempre firme con mis resoluciones. A mi primo Álvaro, por haber sido siempre como un hermano. A las personas que han ido entrando en la familia durante esta etapa: Pepe y Laura, gracias por tratarme tan bien y por hacerme sentir tan cómoda a vuestro lado. A Bruno, por todas las veces que me ha salvado. Finalmente quiero dar las gracias a mi padre. Sin él no sería ni la mitad de lo que hoy soy.

GRACIAS POR INSPIRARME.

Table of Contents

List of Tables

List of Figures

Abbreviations

aa	amino acid (s)
ATP	adenosine triphosphate
bp	base pair (s)
BSA	bovine serum albumin
CBPs	choline binding proteins
CD	circular dichroism
Ci	curies
Cm	chloramphenicol
cpm	counts per minute
CSP-1	competence stimulating peptide-1
Da	dalton
DNase I	desoxirribonuclease I
dNTP	deoxynucleotide triphosphate
DOC	deoxycholate
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDTA	ethylenediamine-tetra-acetic acid
EMSA	electrophoretic mobility shift assay (s)
EtBr	ethidium bromide
FPLC	fast protein liquid chromatography
G+	Gram positive
GFP	green fluorescent protein
h	hour (s)
HTH	helix-turn-helix
HK	histidine kinase
IPTG	isopropyl-β-D-thiogalactosidase
Kav	partition coefficient
kb	kilobase (s)
K_d	apparent dissociation constant
kDa	kilo Dalton (s)
Km	kanamycin
MCS	multi-cloning site

min	minute (s)
mRNA	messenger RNA
Mw,a	average molecular mass
nm	nanometer (s)
nt	nucleotide (s)
OD	optical density
OH•	hydroxyl radical
ORF	open reading frame
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEI	polyethylenimine
PRD	phosphotransferase system regulation domain
PSA	ammonium persulfate
PTS	phosphotransferase system
RT	room temperature
RNAP	RNA polymerase
rNTP	ribonucleotide triphosphate
rpm	revolutions per minute
RR	response regulator
s	second (s)
$S_{20,w}$	standardized sedimentation coefficient
SD	Shine-Dalgarno
SDS	sodium dodecyl sulfate
ssDNA	single-stranded DNA
T4-PNK	polynucleotide kinase of T4-bacteriophage
Tc	tetracycline
TCSs	two-component signal transduction systems
TEMED	N,N,N',N'-tetra-methylethylenediamine
Tm	melting temperature
Tris	tris-hydroxymethyl-aminomethane
UV	ultraviolet
V	volts
W	watts

Resumen en castellano

Introducción

Regulación global de la expresión de genes de virulencia en bacterias patógenas

La patogenicidad de determinadas bacterias puede entenderse como una respuesta de adaptación rápida a los cambios producidos en las condiciones del entorno que las rodea. La habilidad para detectar y responder a esos cambios es lo que les conferirá cierta ventaja a la hora de colonizar nuevos nichos así como evadir el sistema inmune del organismo infectado. En el caso concreto de las bacterias patógenas, estas respuestas van unidas a cambios en la expresión de determinados genes que codifican factores implicados en virulencia. Normalmente, las bacterias utilizan los sistemas de transducción de señales de dos componentes (TCSs) para conectar los estímulos ambientales con una respuesta adaptativa concreta. Estos sistemas, que están implicados en diversos procesos celulares, se componen de dos proteínas, una histidina quinasa (HK) y un regulador de respuesta (RR). La HK es generalmente una proteína integral de membrana, encargada de captar y responder a un determinado estímulo modificando el estado de fosforilación del RR citosólico. Esta fosforilación provocará cambios conformacionales en el regulador el cual, ahora, podrá actuar como un factor transcripcional activando o reprimiendo la expresión de determinados genes (para una revisión ver Perry *et al*., 2011). Además de estos sistemas, varios reguladores de respuesta denominados *stand-alone* han sido implicados en la regulación global de la expresión de genes de virulencia. En general, el término *stand-alone* hace referencia a *i)* no están asociados a una HK unida a membrana, *ii)* su actividad y/o concentración intracelular varía en respuesta a estímulos externos y *iii)* los mecanismos implicados en la transducción de dichos estímulos no han sido totalmente definidos (McIver, 2009). Dentro de este grupo de reguladores se encuentra la proteína Mga*Spn* de *Streptococcus pneumoniae*, cuya caracterización molecular ha sido el objetivo principal de este trabajo. En este momento, y como consecuencia de nuestra investigación, se considera que Mga*Spn* es un miembro de la familia Mga/AtxA de reguladores de respuesta global, que incluye a los reguladores de virulencia Mga (*S. pyogenes*) y AtxA (*Bacillus anthracis*).

El regulador global Mga de *Streptococcus pyogenes*

S. pyogenes (estreptococo del grupo A; GAS) es una bacteria Gram positiva (G+) que provoca un gran número de enfermedades en humanos, algunas de ellas muy

graves (Cunningham, 2000). La proteína Mga fue el primer regulador global de virulencia del tipo *stand-alone* descrito en GAS. Este regulador controla la expresión de aproximadamente un 10% del genoma bacteriano durante las primeras etapas de la infección, afectando directamente la expresión de genes importantes en adherencia, internalización y evasión del sistema inmune del hospedador, así como la expresión de su propio gen (McIver *et al*., 1999; Ribardo and McIver, 2006). Además, Mga también regula, probablemente de forma indirecta, genes involucrados en el transporte y utilización de carbohidratos (Hondorp and McIver, 2007).

Mga es una proteína de 62 kDa que se une a secuencias localizadas *upstream* de sus promotores diana a través de la región N-terminal, donde se han identificado tres motivos implicados en el reconocimiento y unión a DNA: CMD-1, HTH-3 and HTH-4 (McIver *et al*., 1995; McIver and Myles, 2002; Vahling and McIver, 2006). Sin embargo, no se ha establecido una secuencia consenso que sea reconocida por la proteína Mga. En los estudios de interacción proteína-DNA publicados hasta el momento se ha utilizado Mga fusionada a la proteína de unión a maltosa (43 kDa) o Mga fusionada a una cola de histidinas. En la región central de Mga se han identificado dos dominios que presentan homología con los dominios reguladores denominados PRDs (**P**hosphotransferase **R**egulation **D**omains). Generalmente, estos dominios contienen residuos de histidina que son fosforilados por componentes del sistema de fosfotransferasa (PTS; **P**hospho**t**ransferase **S**ystem) dependiente de fosfoenolpiruvato. Se ha sugerido que el sistema de fosfotransferasa, normalmente implicado en el transporte de azúcares, modularía la actividad del regulador Mga vía fosforilación (Hondorp *et al*., 2013). Además, en la región C-terminal de Mga se ha identificado un motivo que presenta similitud con una región del componente EIIB de los sistemas PTS. Dicho motivo parece estar implicado en la oligomerización de Mga (Hondorp *et al*., 2012). No obstante, y a pesar de estos hallazgos, el mecanismo de regulación transcripcional mediado por Mga sigue siendo una incógnita.

El regulador global AtxA de *Bacillus anthracis*

La proteína AtxA es un regulador global de *B. anthracis* (bacteria G+; agente causal del ántrax) que activa la expresión de los genes que codifican la toxina del ántrax. Estos genes están localizados en el plásmido pXO1. Además, AtxA regula la expresión de genes localizados en el cromosoma bacteriano y en el plásmido pXO2 (Fouet, 2010).

AtxA es una proteína de 56 kDa que presenta dos motivos posiblemente implicados en unión a DNA (región N-terminal), dos dominios PRD (región central) y un motivo tipo

EIIB (región C-terminal). Se ha demostrado, *in silico* e *in vitro*, que las regiones promotoras de los genes de la toxina del ántrax poseen curvatura intrínseca como característica estructural común (Hadjifrangiskou and Koehler, 2008). Además, se ha propuesto que sucesos de fosforilación/desfosforilación en residuos de histidina localizados en los dominios PRD modularían la actividad de AtxA (Tsvetanova *et al*., 2007). Por otro lado, la región C-terminal parece estar implicada en interacciones AtxA-AtxA y, por tanto, en la formación de homo-dímeros y oligómeros de orden superior (Hammerstrom *et al*., 2011). Sin embargo, hasta ahora, no hay estudios publicados sobre la interacción de AtxA con DNA.

Streptococcus pneumoniae y factores de virulencia

S. pneumoniae (el neumococo), es una bacteria G+ componente de la flora normal de la nasofaringe, donde reside como comensal de forma asintomática coexistiendo con otros microorganismos. Sin embargo, cuando el sistema inmune se debilita, es capaz de colonizar diferentes partes del cuerpo humano produciendo diversas enfermedades tales como neumonía, sepsis, meningitis, otitis media y procesos invasivos severos (Kadioglu *et al*., 2008). Esta bacteria tiene un gran interés clínico por ser una de las mayores causas de morbilidad y mortalidad en todo el mundo debido, principalmente, a la aparición de estirpes resistentes a múltiples antibióticos y al escaso conocimiento que existe sobre los mecanismos implicados en la regulación de los factores de virulencia. Datos recientes de la organización mundial de la salud estiman que la neumonía mata anualmente a unos 1,2 millones de niños menores de cinco años, más que el SIDA, la malaria y el sarampión juntos, siendo la neumonía causada por neumococo la más común (www.who.int/mediacentre/factsheets/fs331/es/).

S. pneumoniae produce diferentes factores de virulencia que le permiten colonizar la nasofaringe e invadir otros nichos, así como protegerse del sistema inmune del huésped. Entre los factores de virulencia más estudiados caben destacar la cápsula polisacarídica, enzimas líticas como la autolisina (LytA), la neumolisina o hemolisina (PLY), y proteínas ancladas a la superficie celular bien a través del motivo LPXTG o mediante interacciones no covalentes con la colina (proteínas de unión a colina; CBPs). También han sido implicadas en virulencia algunas proteínas relacionadas con el transporte de diferentes compuestos, como por ejemplo hierro o manganeso (revisado en Mitchell and Mitchell, 2010).

En esta Tesis hemos trabajado fundamentalmente con la estirpe no capsulada R6, que deriva del aislado clínico D39 (serotipo 2). Los genomas de ambas estirpes han sido secuenciados totalmente (Hoskins *et al*., 2001; Lanie *et al*., 2007).

El regulador Mga*Spn* de *S. pneumoniae*

En el 2002, se identificó una serie de posibles reguladores de virulencia en la estirpe TIGR4 (serotipo 4) de *S. pneumoniae* (Hava and Camilli, 2002). Uno de ellos fue la proteína MgrA (**Mg**a-like **r**epressor **A**), que tiene similitud de secuencia con el regulador global Mga de GAS (Hemsley *et al*., 2003). En modelos murinos, MgrA es importante tanto en las etapas de colonización de la nasofaringe como en el desarrollo de infección pulmonar. Además, se ha descrito que MgrA actúa, directa o indirectamente, como represor transcripcional de genes de virulencia localizados en la isla de patogenicidad *rlrA* (Hava *et al*., 2003).

El genoma de la estirpe R6, así como el de su parental D39, carece de la isla de patogenicidad *rlrA* pero posee un gen (*spr1622* o *mgaSpn*) equivalente al gen *mgrA* de TIGR4. Dicho gen codifica una proteína de 493 aminoácidos, denominada por nosotros Mga*Spn*, que difiere de MgrA en dos aminoácidos y que presenta homología de secuencia con los reguladores globales Mga (42,6% de similitud) y AtxA (39,9% de similitud). Debido a esta homología y a la información disponible sobre el papel de MgrA en virulencia, iniciamos este trabajo proponiendo que Mga*Spn* podría ser un miembro de la familia Mga/AtxA de reguladores globales y que, como tal, podría regular la expresión de múltiples genes de virulencia en respuesta a señales extracelulares específicas.

Importancia de este trabajo

S. pneumoniae continúa siendo una de las principales causas de morbilidad y mortalidad a nivel mundial. Para alcanzar un mayor entendimiento de la patogénesis de este microorganismo es esencial llegar a desentrañar los mecanismos moleculares que controlan la expresión de genes de virulencia en función de estímulos ambientales. En los últimos años, se han identificado nuevos reguladores transcripcionales que podrían tener un papel importante en la regulación de la virulencia de *S. pneumoniae*, como es el caso de la proteína Mga*Spn*. A pesar de que se conoce bien el papel de sus homólogos Mga y AtxA en la patogénesis de GAS y *B. anthracis*, respectivamente, los estudios a nivel molecular de ambos reguladores son todavía muy limitados. Cuando iniciamos este trabajo, se tenía un conocimiento muy escaso de la proteína Mga*Spn*. Se sabía que *(i)* estaba involucrada en la colonización de la nasofaringe y en el desarrollo

de neumonía en modelos de infección murinos, y *(ii)* era un represor, directo o indirecto, de la isla de patogenicidad *rlrA*. Por tanto, esta Tesis se ha enfocado en la caracterización molecular de la proteína Mga*Spn,* lo que ha servido para ampliar el conocimiento que se tenía de la misma como regulador transcripcional de genes asociados a la virulencia de neumococo.

Objetivos

Mediante búsquedas realizadas en las bases de datos, encontramos que el genoma de *S. pneumoniae* codificaba un potencial regulador global de la familia Mga/Atxa, el cual había sido relacionado previamente con la virulencia de neumococo. La investigación realizada durante esta Tesis Doctoral ha estado centrada en la caracterización molecular de dicho regulador (denominado Mga*Spn* por nosotros). Con este fin, hemos trabajado en los objetivos siguientes:

1. Identificación del promotor del gen *mgaSpn.*

2. Desarrollo de un procedimiento para la purificación de la proteína Mga*Spn* a gran escala.

3. Análisis del estado de oligomerización de Mga*Spn* en solución y determinación del contenido de estructura secundaria.

4. Identificación de los genes diana de Mga*Spn*: análisis del efecto de Mga*Spn* en la expresión del operón *spr1623-spr1626*.

5. Estudio de la interacción de Mga*Spn* con DNAs lineales de cadena doble.

CAPÍTULO 1. Expresión del gen *mgaSpn*

En esta Tesis hemos identificado el promotor del gen *mgaSpn* (*Pmga*). Mediante ensayos de RT-PCR, hemos demostrado que hay transcripción del gen *mgaSpn* cuando las bacterias crecen en las condiciones estándar de laboratorio, aunque los niveles de expresión génica son bajos, como hemos confirmado por ensayos de *Western-blot*. Como estrategia para aumentar los niveles del RNA mensajero sintetizado a partir del promotor *Pmga*, clonamos dicho promotor en el plásmido pAS (Ruiz-Cruz *et al*., 2010), justo *upstream* de un gen *gfp* reportero que carece de su propio promotor. Utilizando dicha fusión transcripcional y mediante ensayos de *primer extension*, hemos identificado el sitio de inicio de la transcripción del gen *mgaSpn*, que está localizado 39 nucleótidos *upstream* del codón de inicio de la traducción.

Puesto que la RNA polimerasa (RNAP) de *S. pneumoniae* no es comercial, como primera aproximación para caracterizar el promotor *Pmga* utilizamos la holoenzima comercial de *Escherichia coli.* Los resultados obtenidos en experimentos de transcripción *in vitro* han demostrado que esta RNAP puede iniciar la transcripción no sólo a partir de la coordenada 1598309 (promotor *Pmga*) sino también en la coordenada 1598369 (promotor *P2*). En el laboratorio hemos desarrollado un método para purificar la subunidad σ^{43} de la RNAP de neumococo. Esta subunidad es homóloga a la subunidad σ^{70} de *E. coli* y se utilizó para reconstituir una RNAP funcional utilizando el *core* comercial de la RNAP de *E. coli*. Experimentos de transcripción *in vitro* demostraron que la RNAP reconstituida reconoce el promotor *Pmga* pero no el promotor *P2*.

CAPÍTULO 2. Caracterización biofísica del regulador transcripcional Mga*Spn*

El primer paso para poder llevar a cabo una caracterización en profundidad del regulador Mga*Spn* fue el desarrollo de un protocolo que nos permitiese su purificación en estado nativo. Además, purificamos una versión de Mga*Spn* que tiene seis histidinas adicionales en el extremo C-terminal (Mga*Spn*-His). En ambos casos, clonamos el gen *mgaSpn* en el vector de expresión pET24b, con lo que obtuvimos estirpes de *E. coli* sobre-productoras de la proteína Mga*Spn* en sus dos versiones. Para la purificación de la proteína Mga*Spn*-His, empleamos columnas de afinidad de níquel y cromatografía de filtración en gel. Además, hemos obtenido anticuerpos policlonales contra esta proteína. El procedimiento desarrollado para purificar la proteína Mga*Spn* en su estado nativo consta, esencialmente, de tres etapas: *(i)* precipitación del DNA y Mga*Spn* (presumiblemente unida al DNA) con polietilenimina (PEI) a baja fuerza iónica, *(ii)* elución de la proteína Mga*Spn* del precipitado de PEI aumentando la fuerza iónica del tampón y *(iii)* cromatografía de afinidad en columnas de heparina. Además, mediante electroforesis en geles de poliacrilamida-SDS (SDS-PAGE), y sobrecargando el gel (tinción con Coomassie blue), estimamos que la pureza de la preparación obtenida era superior al 95%.

Mediante ensayos de filtración en gel y ultracentrifugación analítica (equilibrio de sedimentación y velocidad de sedimentación), hemos demostrado que, en las condiciones ensayadas, la proteína Mga*Spn* forma principalmente dímeros en solución, aunque tiene tendencia a formar especies de mayor masa molecular en función de la concentración de proteína. El valor del coeficiente friccional (f/f_0) obtenido con estos experimentos (1,45) indica que Mga*Spn* tendría una forma elipsoidal, su comportamiento hidrodinámico difiere del de una partícula esférica rígida (f/f_0=1).

Utilizando programas informáticos (Pfam y Phyre2), hemos definido la organización de los posibles dominios funcionales de Mga*Spn*. De acuerdo con estos análisis, Mga*Spn* presenta en su región N-terminal dos posibles motivos HTH de unión a DNA, seguidos de dos posibles dominios PRD (región central) y de un motivo tipo EIIB (región C-terminal). Mediante ensayos de dicroísmo circular, hemos determinado que la proteína Mga*Spn* posee un alto contenido en α–hélices (55,3%), datos que correlacionan con los obtenidos mediante análisis bioinformáticos (programas SABLE, PSIPred, JPred, NPS@, PredictProtein).

CAPÍTULO 3. Papel activador del regulador transcripcional Mga*Spn*

En este capítulo presentamos un análisis transcripcional detallado del operón *spr1623-spr1626* de la estirpe R6 de *S. pneumoniae* y demostramos que Mga*Spn* controla positivamente la expresión de dicho operón. Los resultados de esta investigación, publicados en el año 2012 (ver Related publications), pusieron de manifiesto, por primera vez, el papel activador del regulador Mga*Spn*.

Durante el desarrollo de esta Tesis, hemos construido una estirpe mutante de deleción (estirpe R6Δ*mga*) en la que el gen *mgaSpn* ha sido sustituido por el gen *cat* (resistencia a cloranfenicol) del plásmido pC194 (Horinouchi and Weisblum, 1982). Por otro lado, hemos construido una estirpe de neumococo sobre-productora de Mga*Spn*. Para ello, clonamos el gen *mgaSpn* bajo el promotor del gen *sulA* (dihidropteroato sintasa) en el plásmido pDL287 (LeBlanc *et al*., 1993) que confiere resistencia a kanamicina (estirpe R6Δ*mga*/pDL*Psul::mga*). Mediante ensayos de *Western-blot*, hemos detectado Mga*Spn* en extractos celulares de *S. pneumoniae*, tanto de la estirpe silvestre R6 como de la estirpe sobre-productora (niveles de Mga*Spn* ocho veces más altos que en la estirpe silvestre). Además, hemos confirmado que la estirpe mutante de deleción no sintetiza Mga*Spn*.

Para estudiar el efecto de Mga*Spn* sobre la expresión génica global, realizamos estudios preliminares de proteómica en colaboración con el Dr. J. A. López (Servicio de Proteómica del CNIC) comparando la estirpe silvestre (R6) con la estirpe mutante de deleción (R6Δ*mga*). Estos estudios dieron como resultado la identificación de 10 posibles genes diana que estarían regulados directa o indirectamente por Mga*Spn*. Entre ellos cabe resaltar el gen *spr1625* (proteína Gls24; posible proteína general de estrés), en cuya validación hemos trabajado (ver a continuación). El producto del gen *spr1625* presenta homología con la proteína de respuesta a estrés Gls24 de *Enterococcus faecalis*, que ha sido implicada en virulencia y resistencia a sales biliares (Teng *et al*., 2005).

Mediante RT-PCR y utilizando RNA total aislado de *S. pneumoniae* R6, hemos detectado expresión del gen *spr1623* en condiciones estándar de crecimiento bacteriano. Además, hemos demostrado que los genes *spr1623*, *spr1624*, *spr1625* y *spr1626*, de función desconocida, constituyen un operón. Este operón está localizado *upstream* del gen *mgaSpn* y en la cadena complementaria (transcripción divergente). Mediante ensayos de *primer extension*, hemos identificado dos promotores (*P1623A* y

P1623B) del operón *spr1623-spr1626.* Para estudiar en más detalle el efecto del regulador Mga*Spn* sobre la actividad de los promotores *P1623A y P1623B,* clonamos un fragmento de DNA que incluye ambos promotores en el vector plasmídico pAST (Ruiz-Cruz *et al*., 2010), justo *upstream* de un gen *gfp* (proteína fluorescente verde) reportero que carece de promotor. Mediante ensayos de fluorescencia, comprobamos que dicho fragmento tenía actividad promotora*.* Cuando introdujimos esta construcción en la estirpe R6 silvestre (niveles basales de Mga*Spn*) y en la estirpe R6Δ*mga* (ausencia de Mga*Spn*), observamos que Mga*Spn* activaba uno o ambos promotores. Estos experimentos se complementaron con ensayos de *primer extension* utilizando la estirpe R6/pDL*Psul::mga*, lo que nos permitió observar que Mga*Spn* codificada en plásmido (niveles altos) activaba ambos promotores cromosómicos, *P1623A* y *P1623B*, aunque el efecto activador fue más marcado sobre el promotor *P1623B.* El mismo tipo de ensayo utilizando la estirpe R6Δ*mga*/pDL*Psul::mga* nos permitió concluir que la estirpe R6Δ*mga* carece no sólo del gen *mgaSpn* y su promotor, sino también de una región necesaria para la activación del promotor *P1623B* por la proteína Mga*Spn*. La identificación de dicha región se realizó mediante construcción de fusiones transcripcionales (gen *gfp*) en el plásmido pAST, lo que nos permitió evaluar la actividad promotora de diferentes regiones cromosómicas en distintos fondos genéticos (R6 versus R6Δ*mga*). Combinando ensayos de fluorescencia y *primer extension*, demostramos que Mga*Spn* activa el promotor *P1623B in vivo* y que esta activación requiere la presencia de una región de 70-pb que hemos denominado "región de activación *PB*" y que está localizada entre los promotores divergentes *P1623B* y *Pmga*, exactamente 50 nucleótidos *upstream* del sitio de inicio de transcripción correspondiente al promotor *P1623B*.

Para estudiar la posible implicación de Mga*Spn* en autorregulación, construimos una fusión transcripcional entre el promotor *Pmga* y el gen *gfp* en el vector pAST2 (Ruiz-Cruz *et al*., 2010). Mediante ensayos de fluorescencia, observamos que la actividad del promotor *Pmga* era similar en las estirpes R6 silvestre (niveles bajos de Mga*Spn*), R6Δ*mga* (ausencia de Mga*Spn*) y R6Δ*mga*/pDL*Psul::mga* (niveles altos de Mga*Spn*), por lo que concluimos que, en las condiciones ensayadas, Mga*Spn* no influye la actividad de su propio promotor.

CAPÍTULO 4. Propiedades de unión a DNA del regulador Mga*Spn*

Durante el desarrollo de esta Tesis, una de las cuestiones que nos planteamos fue analizar en detalle cómo era la interacción del regulador Mga*Spn* con sus DNAs diana. Los resultados de este estudio, publicados en el año 2013 (ver Related publications), demostraron que Mga*Spn* genera complejos multiméricos al interaccionar con DNAs lineales de cadena doble, una característica que hasta entonces no había sido descrita en otros reguladores globales de la familia Mga/AtxA.

Mediante ensayos de retraso en gel (EMSA), demostramos que Mga*Spn* se une a DNA lineal de cadena doble y determinamos que dicha reacción alcanza el equilibrio rápidamente (en menos de 1 min). Además, la formación de complejos Mga*Spn*-DNA no se vio afectada por la concentración de NaCl, al menos en el rango de 20 a 300 mM. La interacción de Mga*Spn* con DNA lineal parece no requerir el reconocimiento de una secuencia nucleotídica específica, puesto que la proteína fue capaz de unirse a fragmentos de DNA de distinta procedencia. En todos los casos, observamos un patrón de complejos proteína-DNA compatible con la formación de complejos multiméricos, en los que múltiples unidades de Mga*Spn* se unirían ordenadamente sobre la misma molécula de DNA. La capacidad que tiene Mga*Spn* de extenderse a lo largo del DNA fue comprobada posteriormente mediante ensayos de protección a la digestión con DNasa I. Además, mediante EMSA y utilizando DNA de timo de ternera como DNA competidor, determinamos que las múltiples unidades de Mga*Spn* interaccionan con el DNA de manera no cooperativa, al menos en la formación de los primeros cuatro complejos.

Utilizando un fragmento de DNA (222-pb) que contenía la "región de activación *PB*" (70-pb; ver Capítulo 3) en posición interna y mediante ensayos de protección a la digestión con DNasa I, encontramos que Mga*Spn* reconoce preferentemente dicha región. Estos resultados fueron confirmados posteriormente mediante ensayos de protección frente a corte por radical hidroxilo. Concretamente, Mga*Spn* interacciona con la región comprendida entre las posiciones -60 y -99 del promotor *P1623B* (sitio primario; 40-pb). Es decir, la "región de activación *PB",* que es necesaria para que Mga*Spn* active el promotor *P1623B in vivo*, es reconocida por Mga*Spn in vitro,* lo que demuestra que Mga*Spn* activa directamente dicho promotor. Sin embargo, inesperadamente, observamos que la "región de activación *PB"* no es reconocida por Mga*Spn* cuando está localizada en un extremo de la molécula de DNA (fragmento de 224-bp que contiene el promotor *Pmga* en posición interna). En este caso, y mediante ensayos de protección frente a rotura por radical hidroxilo, encontramos que Mga*Spn*

interacciona preferentemente con la región comprendida entre las posiciones -23 y +21 del promotor *Pmga* (sitio primario; 44-pb). Curiosamente, mediante microscopía electrónica y utilizando fragmentos de DNA más largos (640-1458 pb) que contenían ambos sitios primarios ("región de activación *PB"* y promotor *Pmga*) en posición interna, demostramos que Mga*Spn* tiene preferencia por la "región de activación *PB"*. Los ensayos de microscopía electrónica fueron realizados en el Max-Planck Institut für Molekulare Genetik (Berlín) y en colaboración con el Dr. R. Lurz. Además, mediante microscopía electrónica y utilizando relaciones molares de Mga*Spn* frente a DNA más altas, observamos moléculas de DNA cubiertas parcial o completamente de proteína, sin que esta interacción modificase la longitud del DNA. Estos resultados confirman que Mga*Spn* es capaz de reconocer un sitio concreto de la molécula de DNA para, seguidamente, extenderse a lo largo de la misma.

Mediante EMSA y utilizando fragmentos de DNA pequeños (entre 40 y 20-pb), delimitamos la longitud mínima de DNA necesaria para la unión de Mga*Spn*. Esta longitud está comprendida entre 20 y 26-pb.

Hemos analizado la secuencia nucleotídica de los dos sitios reconocidos preferentemente por Mga*Spn* ("región de activación *PB"* y promotor *Pmga*) y hemos visto que comparten una baja identidad de secuencia: **GGT**(A/T)(A/T)**AATT** y **GA**(A/T)**AATT**. Esta baja identidad de secuencia, junto con el hecho de que Mga*Spn* tiene preferencia por la "región de activación *PB"* cuando está presente en posición interna pero no cuando está en el extremo del DNA, nos llevó a concluir que Mga*Spn* requeriría algo más que una secuencia de bases para reconocer específicamente una posición en un DNA determinado. El análisis de los fragmentos de DNA de 222- y 224-pb utilizando el programa bend.it (Vlahovicek *et al*., 2003) mostró que los dos sitios primarios de Mga*Spn* se caracterizan por tener una curvatura intrínseca potencial, mientras que las regiones adyacentes a los mismos presentan un alto grado de bendabilidad. Para analizar si Mga*Spn* mostraba preferencia por fragmentos de DNA curvado, realizamos ensayos de EMSA con el fragmento de 222-pb ("región de activación *PB"* en posición interna) utilizando dos tipos de DNAs como competidores: el fragmento C DNA (321-pb; 72,3% de A+T) o el fragmento NC DNA (322-pb; 71.1% de A+T). De acuerdo con predicciones de curvatura intrínseca, el grado de curvatura del fragmento C DNA es mucho mayor que el del NC DNA. De hecho, el fragmento C DNA tiene una movilidad electroforética anómala en geles de poliacrilamida nativos, una característica descrita en DNAs curvados (Diekmann, 1987). Además, calculamos la afinidad de Mga*Spn* por cada uno de estos fragmentos de DNA. Los resultados

obtenidos en ambos ensayos mostraron que Mga*Spn* tiene preferencia por DNA curvado. Esto nos lleva a concluir que conformaciones locales en el DNA podrían contribuir al reconocimiento específico de una determinada región del DNA por Mga*Spn*.

Conclusiones

Este estudio ha contribuido al conocimiento molecular del regulador transcripcional Mga*Spn* implicado en la virulencia de neumococo. Las conclusiones más importantes obtenidas durante la realización de esta Tesis son las siguientes:

1. El gen *mgaSpn* se transcribe *in vivo* a partir del promotor *Pmga*, que es reconocido *in vitro* por el factor σ^{43} de *S. pneumoniae*. La transcripción del gen *mgaSpn* comienza 39 nucleótidos *upstream* del codón de inicio de la traducción.

2. El operón *spr1623-spr1626,* adyacente al gen *mgaSpn*, se transcribe *in vivo* a partir de dos promotores, *P1623A* y *P1623B.* Estos promotores y el promotor *Pmga* son divergentes.

3. Mga*Spn* actúa, de forma directa, como activador del promotor *P1623B in vivo* y, por tanto, activa la transcripción del operón *spr1623-spr1626*. Esta activación requiere la interacción de Mga*Spn* con un sitio localizado *upstream* del promotor *P1623B* (posiciones -60 a -99).

4. El regulador Mga*Spn* está altamente conservado en las estirpes de neumococo cuyos genomas han sido parcial o totalmente secuenciados. La organización de dominios funcionales predicha para Mga*Spn* es similar a la de los reguladores de respuesta global Mga y AtxA.

5. Mga*Spn* tiene un alto contenido en α-hélices (alrededor de un 55%). En solución y en las condiciones experimentales utilizadas, Mga*Spn* se encuentra, mayoritariamente, formando dímeros. Sin embargo, tiende a formar especies de mayor masa molecular en función de la concentración de proteína.

6. Mga*Spn* interacciona con DNA lineal de cadena doble con alta afinidad pero con baja especificidad de secuencia.

7. Conformaciones locales en el DNA (por ejemplo, curvatura intrínseca) podrían contribuir a la especificidad de unión de Mga*Spn* a una determinada región del DNA.

8. Tras unirse al sitio primario, Mga*Spn* es capaz de extenderse (probablemente mediante oligomerización) a lo largo de las regiones de DNA adyacentes, generando complejos proteína-DNA multiméricos.

Summary

INTRODUCTION: Bacteria usually live in habitats of changing conditions. During infection, pathogenic bacteria must be able to survive in different environments encountered as the pathogen progresses through its host. This adaptation requires sensing the relevant extracellular signals and linking them to a coordinate change in the expression of genes, which encode factors appropriate to the given situation. Global transcriptional regulators that respond to specific environmental signals are key elements in such regulatory networks. Bacteria often use classical two-component signal transduction systems (TCSs) to link the environmental signals to adaptive responses (Stock *et al*., 2000). Moreover, in addition to TCSs, *stand-alone* response regulators have been implicated in the global regulation of virulence gene expression. The term *stand-alone* has been used to define global transcriptional regulators that *(i)* are not associated to a membrane-bound sensor histidine kinase, *(ii)* their activity and/or intracellular concentration changes in response to specific external stimuli and *(iii)* their signal transduction components have yet to be fully defined (McIver, 2009).

The Gram-positive (G+) bacterium *Streptococcus pneumoniae*, commonly called the pneumococcus, is a member of the normal human nasopharyngeal flora, where it exists asymptomatically as a commensal. However, when the immune system weakens, it can cause serious diseases such as sinusitis, conjunctivitis, otitis media, meningitis and bacteremia (Kadioglu *et al*., 2008; van der Poll and Opal, 2009). *S. pneumoniae* remains as a main cause of morbidity and mortality worldwide as a result of its increasing resistance to antibiotics. Recent data estimate that pneumococcal pneumonia kills annually around 1.2 million children younger than five years, more than AIDS, malaria and tuberculosis combined (www.who.int/mediacentre/factsheets/fs331/en/index.html). Understanding the molecular mechanisms that control the expression of pneumococcal virulence genes in response to environmental stimuli will offer new insights into the pathogenesis of this bacterium.

Searching for homologies we found that the genome of the pneumococcal R6 strain (Hoskins *et al*., 2001), which derives from the D39 clinical isolate (serotype 2), encodes a protein (named Mga*Spn* by us), which is highly conserved in the pneumococcal strains whose genomes have been totally or partially sequenced. At present, Mga*Spn* is thought to be a member of the Mga/AtxA family of global response regulators, which includes the Mga and the AtxA virulence regulators encoded by the G+ pathogens *S. pyogenes* (the Group A *Streptococcus*, GAS) and *Bacillus anthracis*, respectively. Mga*Spn* shares 42.6% of similarity and 21.4% of identity with Mga and 39.9% of similarity and 20.7% of identity with AtxA. Regarding the Mga regulator, it

controls the expression of approximately 10% of the GAS genome during the exponential growth phase (Ribardo and McIver, 2006). Mga activates directly the transcription of several virulence genes, which encode factors important for adherence and internalization into non-phagocytic cells, as well as factors that enable the bacterium to evade the host immune responses. Mga also activates the expression of its own gene (McIver *et al*., 1999; McIver, 2009). *In vitro* studies using a His-tagged Mga showed that it binds to regions located upstream of the target promoters with low sequence identity. Its ability to interact with DNA resides at the N-terminal region, where it contains two helix-turn-helix (HTH) motifs (HTH-3 and HTH-4) that have been shown to be involved in DNA binding and transcriptional activation (Vahling and McIver, 2006). The His-tagged Mga protein is able to form oligomers in solution and this ability correlates with transcriptional activation (Hondorp *et al*., 2012). Nevertheless, the transcriptional regulation mechanism mediated by Mga remains unknown. Concerning the AtxA protein, it regulates the expression of numerous genes (chromosomal or plasmid-encoded genes), including the three anthrax toxin genes located in plasmid pXO1 (for a review see Fouet, 2010). However, studies on the interaction of AtxA with DNA are not available. It has been shown that AtxA activity is modulated by the phosphotransferase system (PTS) (Tsvetanova *et al*., 2007). Moreover, its C-terminal region seems to be involved in AtxA-AtxA interactions (Hammerstrom *et al*., 2011). Although the role in pathogenesis of both regulators (Mga and AtxA) is well known, studies at a molecular level are still very limited.

When this work was started, we knew almost nothing regarding the pneumococcal Mga*Spn* regulator except that *(i)* it was involved in nasopharyngeal colonization and development of pneumonia in murine infection models (Hava and Camilli, 2002), and *(ii)* it was a repressor of the *rlrA* pathogenicity islet (Hemsley *et al*., 2003). Therefore, this Thesis has been focussed on the molecular characterization of the Mga*Spn* transcriptional regulator.

PRINCIPAL FINDINGS: In this study, we have identified the *mgaSpn* gene promoter (*Pmga*). By RT-PCR assays we have demonstrated that *mgaSpn* is transcribed under our laboratory conditions. Primer extension experiments were carried out to identify the transcription initiation site of the *mgaSpn* gene, which was found to be located 39 nucleotides upstream of the translation start codon. The pneumococal σ^{43} factor (gene *rpoD*) is homologous to the housekeeping σ^{70} factor of *Escherichia coli*. We used the σ^{43} factor (purified in our laboratory) to reconstitute a functional RNA polymerase (RNAP)

with the commercial *E. coli* RNAP core enzyme. *In vitro* transcription assays using the reconstituted RNAP showed that the *Pmga* promoter is recognized by the σ^{43} factor.

A remarkable achievement of this work has been the development of a procedure to purify the untagged Mga*Spn* protein. As far as we know, it is the first case within the Mga/AtxA family of global regulators. This method involves, essentially, three steps: *i)* precipitation of DNA and Mga*Spn* with polyethyleneimine (PEI) at low ionic strength; *ii)* elution of Mga*Spn* from the PEI pellet with higher ionic strength; and *iii)* chromatography on heparin columns. Gel filtration chromatography and analytical ultracentrifugation assays (sedimentation velocity and sedimentation equilibrium) showed that Mga*Spn* forms dimers in solution. Moreover, it is able to form higher-order oligomers as its concentration increases. The frictional ratio (f/f_0) calculated was 1.45, indicating that the hydrodynamic behaviour of Mga*Spn* deviates from that of a rigid spherical particle with a frictional ratio value of 1.0. Thus, Mga*Spn* can be expected to have an ellipsoidal shape. The secondary structure content of the Mga*Spn* regulator determined by far-ultraviolet (far-UV) circular dichroism (CD) spectroscopic analyses correlates with the predicted by computational methods and revealed that Mga*Spn* has a high content of α-helices (around 55%). *In silico* analyses using the Pfam protein families database (Punta *et al*., 2012) and the protein structure prediction server Phyre2 (Kelley and Sternberg, 2009) have indicated that Mga*Spn* exhibits similarity to Mga and AtxA in the domain organization. It presents: two putative DNA-binding motifs (HTH-Mga and Mga, respectively) within the N-terminal region, two putative PTS regulatory domains (PRDs) at the central region and an EIIB-like motif at the C-terminal region. This organization of functional domains suggests that the N-terminal region of Mga*Spn* might participate in recognition and binding to DNA whereas the capacity to establish protein-protein interactions might reside at the C-terminal region.

To study the effect of Mga*Spn* on gene expression, we constructed two derivatives of the pneumococcal R6 strain: a deletion mutant strain (R6Δ*mga*) which does not synthesize Mga*Spn* and a strain designed to overproduce Mga*Spn* (R6Δ*mga* /pDL*PsulA::mga*). The latter strain carries the *mgaSpn* gene cloned into a plasmid. In order to identify the target genes of the Mga*Spn* regulator, we analysed the effect of the absence of Mga*Spn* on the pattern of global gene expression by proteomics (strain R6 versus strain R6Δ*mga*). The results obtained showed changes in 10 candidates. One of the proteins whose levels decreased (2.3-fold) in the absence of Mga*Spn* was the putative general stress protein 24, encoded by the *spr1625* gene. In this work we have demonstrated by RT-PCR assays that the *spr1625* gene is the third of four genes

(*spr1623, spr1624, spr1625 and spr1626*) of unknown function that are transcribed into a polycistronic mRNA molecule. The operon is located upstream of the *mgaSpn* gene and on the complementary strand. Moreover, it is conserved in other pneumococcal strains whose genomes have been totally sequenced and several observations suggest that it might play a role in virulence. Additionally, primer extension experiments have shown that the *spr1623-spr1626* operon is transcribed from two promoters named *P1623A* and *P1623B*. To study in detail the effect of Mga*Spn* on the activity of the *P1623A* and *P1623B* promoters, a DNA fragment that contained both promoters was cloned into the pAST plasmid (Ruiz-Cruz *et al*., 2010), just upstream of a promoter-less *gfp* gene. Fluorescence assays combined with primer extension experiments (R6, R6Δ*mga*, R6/pDL*PsulA::mga* and R6Δ*mga*/pDL*PsulA::mga* as genetic backgrounds) allowed us to conclude that Mga*Spn* activates *in vivo* the *P1623B* promoter. Such activation requires a region of 70-bp (named *PB* activation region) located between the *Pmga* and the *P1623B* divergent promoters. Furthermore, we have shown that Mga*Spn* does not influence the activity of its own promoter under the experimental conditions tested. This part of the Thesis was published in 2012 (Solano-Collado *et al*., 2012).

We have performed an in-depth analysis of the DNA-binding properties of the Mga*Spn* transcriptional regulator by different techniques (electrophoretic mobility shift assays (EMSA), footprinting and electron microscopy). This part of the Thesis was published in 2013 (Solano-Collado *et al*., 2013). By EMSA we have demonstrated that Mga*Spn* binds to linear double-stranded DNA. The binding reaction reaches the equilibrium very fast (less than 1 min) and is not affected by the salt concentration used in the experiments (20-300 mM NaCl). Using DNA fragments from different sources, we found that Mga*Spn* is able to bind to any tested DNA generating a pattern of complexes compatible with the formation of multimeric complexes, in which multiple protein units bind orderly on the same DNA molecule. Thus, Mga*Spn* binds DNA with high affinity, but with low sequence specificity. The ability of Mga*Spn* to spread along the DNA molecule was confirmed by DNase I footprinting experiments. Also, EMSA experiments suggested that multiple Mga*Spn* units bind to DNA in a non-cooperative manner.

By DNase I footprinting assays, we analysed the binding of Mga*Spn* to a 222-bp DNA, which contained the *PB* activation region at internal position as well as the *Pmga* and the *P1623B* divergent promoters. This study showed that Mga*Spn* interacts with sequences located between the -52 and -102 positions with respect to the *P1623B* transcription start site. Such a region is included within the *PB* activation region. Therefore, the Mga*Spn* regulator activates directly the *P1623B* promoter. These results

were further confirmed by hydroxyl radical footprinting experiments. Specifically, Mga*Spn* interacts with a region between the -60 and -99 positions relative to the *P1623B* promoter. Additionally, we used a 224-bp DNA fragment that lacks the *P1623B* promoter but contains the *Pmga* promoter and the *PB* activation region. The latter placed at one DNA end. Unexpectedly, in such a fragment, Mga*Spn* does not recognize the *PB* activation region as its primary site. Instead, it recognizes a region between the -23 and +21 positions relative to the *Pmga* promoter. In both DNA fragments (222 and 224-bp), upon binding to its primary site, multiple Mga*Spn* units bind orderly along the DNA fragment.

By electron microscopy experiments, we have analysed the binding of Mga*Spn* to long DNA molecules (640-1485 bp). Such fragments contained both primary sites (the *PB* activation region and the *Pmga* promoter) at internal positions. The results obtained revealed that Mga*Spn* binds to the *PB* activation region rather than to the *Pmga* promoter when both sites are located at internal positions on the same DNA. Moreover, they supported that Mga*Spn* is able to spread along the DNA upon binding to a particular site.

Sequence analysis of the two sites recognised preferentially by Mga*Spn* showed a low sequence identity: **GGT**(A/T)(A/T)**AATT** and **GA**(A/T)**AATT.** Using the bend.it server (Vlahovicek *et al*., 2003) we have shown that both primary binding sites of Mga*Spn* contain a potential intrinsic curvature, which is surrounded by regions with the capacity of being easily bent. By EMSA experiments using an intrinsically curved DNA and a non-curved DNA we have demonstrated that Mga*Spn* has a higher affinity for the naturally occurring curved DNA. Our study suggests that local DNA conformations might contribute to the DNA-binding specificity of Mga*Spn.*

Introduction

1. Global regulation of virulence gene expression in pathogenic bacteria

Bacteria usually live in habitats of frequently changing conditions and must be able to survive in the different environmental and physiological situations encountered during their life cycle. The genes necessary to adapt the cell physiology or metabolism to the new state are usually organized in a network of interconnected regulons, which are expressed differentially (activated and/or repressed) in response to the surrounding conditions. The ability to respond to these stimuli involves sensing the relevant signals and linking them to a coordinated expression of genes that encode factors suitable for the given situation. Virulence of pathogenic bacteria can be understood as a rapid adaptive response to different environments encountered as the pathogen progresses through its host. Global transcriptional regulators that respond to specific environmental signals are key elements in such regulatory networks. Bacteria often use classical two-component signal transduction systems (TCSs), which are absent in mammals, to link environmental stimuli to adaptive responses (Stock *et al*., 2000). For instance, 13 TCSs have been identified in *Streptococcus pneumoniae* and some of them are known to contribute to its virulence. However, such a contribution has been shown to vary significantly depending on the pneumococcal strain and/or the infection model used (Beier and Gross, 2006; Paterson *et al*., 2006). A prototypical TCS comprises a sensor protein, which is anchored to the bacterial membrane, and a cytoplasmic regulatory protein (Figure 1A). Sensor proteins are histidine kinases (HKs) that monitor the external stimuli and transmit the signal to their cognate regulatory protein by a phosphorylation event. HKs typically present a modular structure composed of a diverse sensing domain (N-terminal region) and a highly conserved cytoplasmic kinase domain or kinase core (C-terminal region), both of them connected by a transmembrane linker region. External changes are sensed by the N-terminal domain of the HK and transmitted to the kinase core, which is activated to autophosphorylate a conserved histidine residue (Figure 1A). HKs usually function as homodimers in which one HK monomer catalyses the phosphorylation of the conserved histidine in the second monomer. The phosphorylated kinase transfers the phosphoryl group to a conserved aspartate residue of its cognate regulatory protein (response regulator; RR). The phosphorylation event promotes conformational changes in the RR that allow it to regulate gene expression either at transcriptional level (in most of the reported cases) or at post-transcriptional level (for a review see Perry *et al*., 2011).

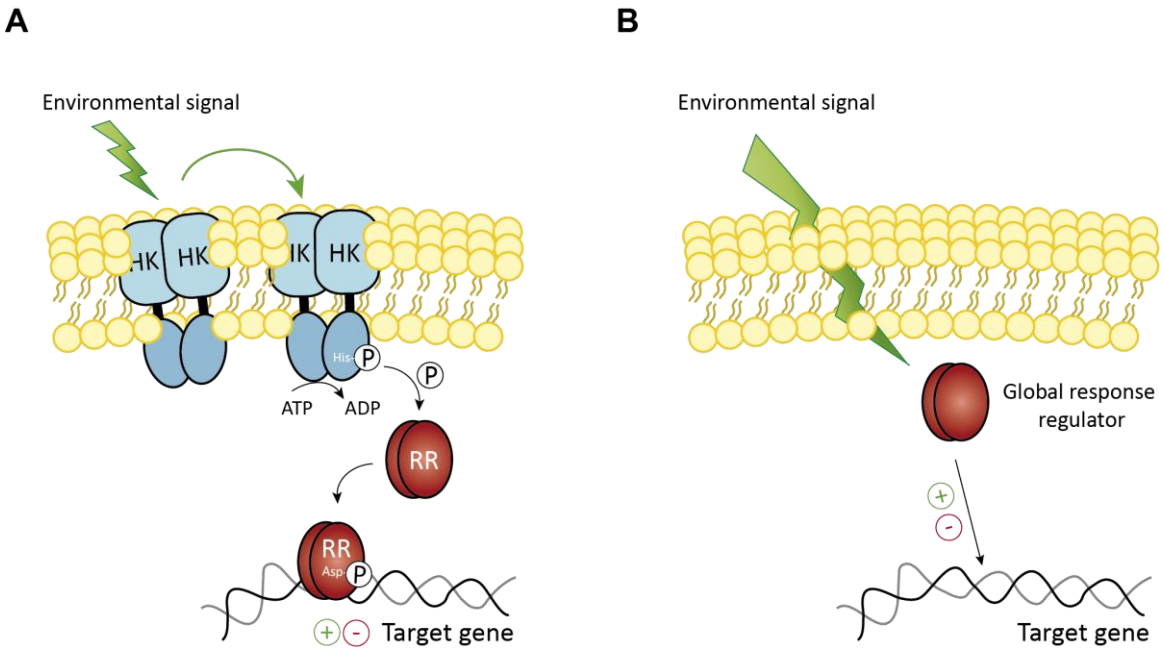


Figure 1. Adaptation to environmental changes. (**A**) A typical two-component signal transduction system (TCS). After detection of the environmental signal by the N-terminal domain of the histidine kinase (HK), it uses ATP to autophosphorylate a conserved histidine residue within the cytoplasmic domain. The phosphoryl group is then transferred to an aspartate residue on its cognate cytoplasmic response regulator (RR). The phosphorylated RR is now able to exert its regulatory function by interaction with promoter regions located upstream of the target genes. (**B**) *Stand-alone* global response regulators. In general, the signal transduction mechanism remains to be defined. The regulator influences positively or negatively the expression of its target genes in response to the environment.

In addition to TCSs, various *stand-alone* response regulators have been implicated in the global regulation of virulence gene expression (Figure 1B). In general, the term *stand-alone* response regulator has been used in the literature to define global transcriptional regulators that *(i)* are not associated to a membrane-bound sensor HK, *(ii)* their activity and/or intracellular concentration changes in response to specific external stimuli and *(iii)* their signal transduction components have yet to be fully defined (McIver, 2009). To this class of global transcriptional regulators belongs the pneumococcal Mga*Spn* protein, whose molecular characterization has been the main goal of this work. At present, Mga*Spn* is thought to be a member of the Mga/AtxA family of global response regulators, which includes the Mga and the AtxA virulence regulators encoded by the Gram-positive (G+) pathogens *S. pyogenes* (the Group A *Streptococcus*, GAS) and *Bacillus anthracis*, respectively.

2. The Group A *Streptococcus* Mga virulence regulator

Unlike many other pathogenic bacteria, GAS does not appear to use alternative sigma factors to regulate virulence gene expression. Instead, it depends on global response regulators, both TCSs and *stand-alone* response regulators (Kreikemeyer *et al*., 2003; McIver, 2009). GAS is a strict human pathogen that causes a broad spectrum of diseases ranging from benign, self-limiting infections to life-threatening invasive disorders (Cunningham, 2000). Mga (**m**ultiple **g**ene regulator of G**A**S) was the first *stand-alone* response regulator described in GAS. This protein controls the expression of approximately 10% of the GAS genome during the exponential growth phase (Ribardo and McIver, 2006), when expression of the *mga* gene is maximum. Expression of *mga* also increases at elevated levels of CO_2, normal body temperature, and increased levels of iron or metabolizable sugars. Mga activates the transcription of several virulence genes which encode factors important for adherence and internalization into non-phagocytic cells (M protein: *emm*, *arp*; M-family proteins: *mrp/fcrA, enn*; fibronectin-binding proteins: *fba, sof/sfbX*; and collagen-like protein: *sclA*) and factors that enable the bacterium to evade the host immune responses (M protein: *emm*; M-like proteins: *mrp*, *arp*, *enn*; C5a peptidase: *scpA*; and complement inhibitors: *sic*, *fba*). Mga is also able to activate the expression of its own gene providing a mechanism to amplify the Mga regulon (McIver *et al*., 1999). In addition, Mga activates or represses, likely in an indirect way, the expression of genes involved in the transport and utilization of carbohydrates, iron, amino acids, and other metabolites. Thus, Mga is able to regulate not only virulence genes but also genes important for the metabolic homeostasis of GAS (Hondorp and McIver, 2007; McIver, 2009). Moreover, Mga is able to regulate not only genes located adjacent to its own gene but also distant genes (Figure 2).

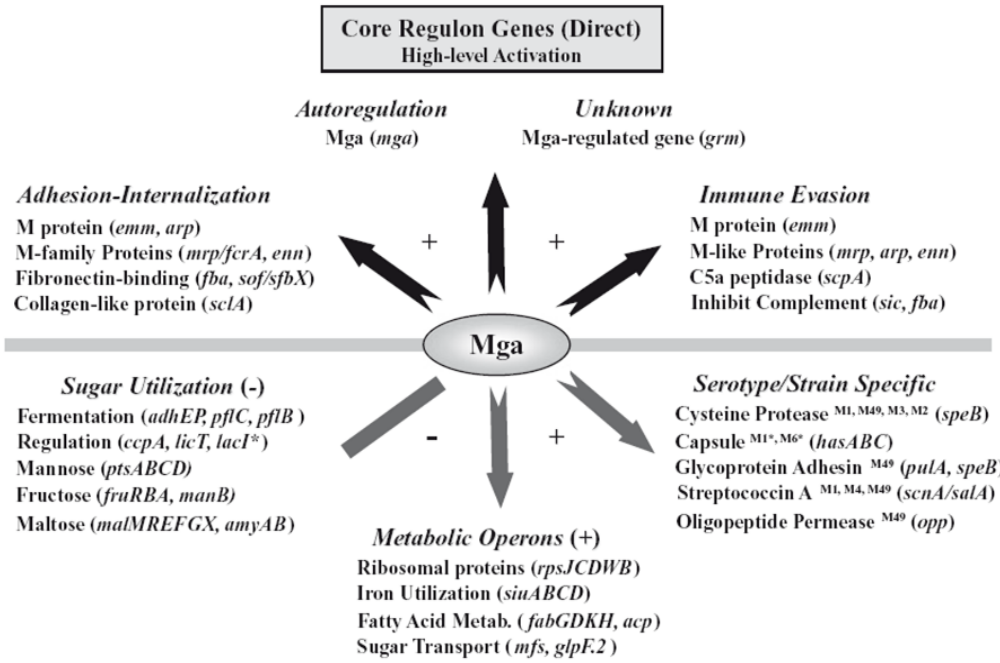


Figure 2. Overview of the Mga regulon in GAS. A compilation of Mga-regulated genes and their products based on their known or predicted function in GAS. An arrow indicates transcriptional activation, while repression is indicated with a bar. Genes that are activated directly by Mga (Core Regulon Genes) are listed at the top. Genes that are activated or repressed by Mga likely indirectly (low-level regulation) are shown at the bottom (Figure from Hondorp and McIver, 2007).

Mga is a large protein (530 amino acids; 62 kDa) with capability to bind directly to regions located upstream of its target genes, as was shown by electrophoretic mobility shift assays (EMSAs) and DNase I footprinting experiments using Mga fusion proteins (either Mga fused to the maltose binding protein (43 kDa) or Mga fused to a His-tag) (McIver *et al*., 1995; McIver *et al*., 1999). However, a consensus site for Mga binding has not been defined (Hondorp and McIver, 2007; Hause and McIver, 2012). The N-terminal region of Mga contains two motifs that were shown to be involved in DNA binding and transcriptional activation: a classical HTH motif (HTH-3, residues 53-72) and a winged HTH motif (HTH-4, residues 107-126). Unlike the HTH-3, the HTH-4 motif is essential for Mga binding to all the promoters tested so far. In addition, a small conserved domain CMD-1 (**c**onserved **M**ga **d**omain-1, residues 10-15) also seems to contribute to DNA binding and activation (Vahling and McIver, 2006). Two central regions of Mga were predicted to have homology to the dual **P**hosphotransferase System (**P**TS) **R**egulation **D**omains (PRDs): PRD-1 (residues 170-287) and PRD-2 (residues 288-390) (Figure 3) (Hondorp and McIver, 2007). PRDs are structural domains found generally in transcriptional activators and antiterminators involved in the regulation of sugar

metabolism (reviewed in Deutscher *et al*., 2006). The activity of such regulators is modulated by phosphorylation of conserved histidine residues within the PRDs in response to the utilization of different carbon sources. In a recent study, Hondorp *et al*. (2013) indicated that Mga activity might be modulated (both positively and negatively) by PTS-mediated phosphorylation, thereby linking Mga-mediated regulation to the sugar status of the cell. Also recently, an EIIB-like domain has been identified at the C-terminal region of Mga (residues 407-490). Truncation analyses suggested that the C-terminal region of Mga is important for oligomerization in solution, and that the formation of oligomers appears to correlate with transcriptional activation (Hondorp *et al*., 2012). Nevertheless, and despite the mentioned findings, the transcriptional regulation mechanism mediated by Mga remains unknown.

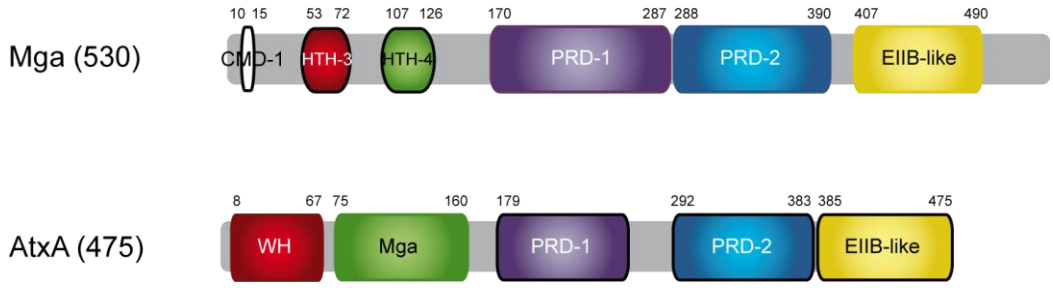


Figure 3. Domain organization of Mga and AtxA. The Mga regulator contains three N-terminal motifs involved in DNA binding (CMD-1, HTH-3 and HTH-4), two central regions with homology to PTS regulation domains (PRD-1 and PRD-2) and an EIIB-like domain at the C-terminal region (Vahling and McIver, 2006; Hondorp and McIver, 2007; Hondorp *et al*., 2012). AtxA contains two N-terminal DNA binding motifs (WH and Mga), two central PRDs and a C-terminal EIIB-like domain involved in AtxA-AtxA interactions (Tsvetanova *et al*., 2007; Hadjifrangiskou and Koehler, 2008; Hammerstrom *et al*., 2011).

3. The *Bacillus anthracis* AtxA virulence regulator

The G+ bacterium *B. anthracis* is the causative agent of anthrax. It carries two plasmids that are essential for its virulence: pXO1 and pXO2. Plasmid pXO1 encodes the AtxA (**A**ntrax **t**o**x**in **A**ctivator) global transcriptional regulator, which activates directly the expression of the three anthrax toxin genes located in plasmid pXO1: *pagA* (Protective Antigen, PA), *cya* (Edema Factor, EF) and *lef* (Lethal Factor, LF). Moreover, AtxA activates indirectly, through the AcpA and AcpB regulators, the expression of the capsule biosynthetic operon (*capBCADE*), which is located in plasmid pXO2. In addition to plasmid-encoded genes, AtxA is able to activate or repress the expression of multiple genes located on the bacterial chromosome (reviewed in Fouet, 2010). Expression of the

atxA gene is affected by temperature, carbohydrate availability, growth phase and redox conditions. Two independent promoters were shown to govern the expression of *atxA* (Bongiorni *et al*., 2008). However, the mechanism(s) regulating the activity of these promoters are largely unknown. On the other hand, AtxA activity is modulated by phosphorylation within the PRDs (presumably by components of the PTS) (see below). In addition, the regulator CodY controls indirectly the intracellular levels of AtxA. Schaik *et al.* (2009) have suggested that either CodY represses an unidentified protease that degrades AtxA or CodY activates the synthesis of a factor (e.g. a chaperone) that influences AtxA stability. The key host-related signal associated with AtxA-regulated gene expression is CO_2/bicarbonate, which affects AtxA function (Dai and Koehler, 1997; Hammerstrom *et al*., 2011).

AtxA is a 56 kDa protein (475 amino acids) whose domain organization is similar to that mentioned before for the Mga virulence regulator (Figure 3). Its N-terminal region contains two DNA binding motifs: a winged-helix (WH) motif (residues 8-67) and a HTH motif (residues 75-160). Nevertheless, studies on the interaction of AtxA with DNA are not available. It was shown that the promoter regions of several AtxA-regulated genes do not exhibit sequence similarities, but *in silico* and *in vitro* analyses revealed that the anthrax toxin promoter regions are characterized by intrinsic curvature (Hadjifrangiskou and Koehler, 2008). The central region of the AtxA regulator contains two PRDs. It was reported that the activity of AtxA is influenced by phosphorylation/dephosphorylation events within the PRDs. In the current model of AtxA activity, phosphorylation of a conserved histidine (H199) in PRD1 results in activation of the regulator whereas phosphorylation of a conserved histidine (H379) in PRD2 inhibits the transcription of the toxin genes (Tsvetanova *et al*., 2007). Moreover, the C-terminal region of AtxA, which contains an EIIB-like motif (residues 385-475), is involved in AtxA-AtxA interactions and, therefore, in the formation of homodimers and higher-order oligomers (Hammerstrom *et al*., 2011).

4. Virulence factors of *Streptococcus pneumoniae*

The G+ bacterium *S. pneumoniae*, commonly called the pneumococcus, is a member of the normal human nasopharyngeal flora, where it exists asymptomatically as a commensal. The rate of pneumococcal carriage varies upon age, socioeconomic status, environmental factors and geographic area (Bogaert *et al*., 2004; Weiss-Salz and Yagupsky, 2010). *S. pneumoniae* generally occurs as characteristic diplococci and produces α-haemolysis. This bacterium is genetically closely related to other commensal

streptococci such as *S. mitis* and *S. oralis*. However, *S. pneumoniae* is a well-known human pathogen capable of causing a wide spectrum of diseases. From the nasopharynx, the pneumococcus can spread through the airway to the lower respiratory tract producing pneumonia. Also, it can invade other places causing diseases such as sinusitis, conjunctivitis, otitis media, meningitis and bacteremia (Kadioglu *et al*., 2008; van der Poll and Opal, 2009). The increasing resistance of *S. pneumoniae* to antibiotics has resulted in the pneumococcus remaining as a major cause of morbidity and mortality worldwide. Recent data estimate that pneumococcal pneumonia kills annually around 1.2 million children younger than five years, more than AIDS, malaria and tuberculosis combined (http://www.who.int/mediacentre/factsheets/fs331/en/index.html).

Pneumococci are transmitted from person to person through airborne droplets from a cough or sneeze, and in Europe and the United States it is the most common cause of community-acquired bacterial pneumonia in adults. In fact, the annual incidence of invasive pneumococcal disease ranges from 10 to 100 cases per 100,000 population (http://www.who.int/ith/diseases/pneumococcal/en). Genetically, *S. pneumoniae* is very versatile because of its ability to take up DNA from the environment and incorporate it into the genome. Pneumococci can obtain exogenous DNA from different sources, such as the extracellular matrix of pneumococcal biofilms (Hall-Stoodley *et al*., 2008) or via microbial fratricide (Havarstein *et al*., 2006), generating a large variety of genetic differences among pneumococcal strains. *S. pneumoniae* is a facultative anaerobe but it is able to adapt to conditions of high oxygen tension. This ability is crucial to protect itself from the harmful effects of oxygen, as the generation of H_2O_2 during bacterial metabolism. In this study, we have worked with *S. pneumoniae* R6, a strain that derives from the D39 clinical isolate (serotype 2) (Lanie *et al*., 2007). R6 is avirulent due to the lack of a polysaccharide capsule and its genome has been totally sequenced (Hoskins *et al*., 2001).

As indicated above, the pneumococcus resides asymptomatically and as a commensal in the nasopharynx of healthy individuals. However, when the immune system weakens, it can cause serious diseases. The pneumococcus produces virulence factors that enable the organism to colonize and invade other host niches as well as to escape from the host immune response (Figure 4). The role of several pneumococcal virulence factors is described in this section. Among them, the polysaccharide capsule is probably the most important virulence determinant in pneumococci. Over ninety different antigenic types of *S. pneumoniae* have been identified so far. (Bentley *et al*., 2006; Yother, 2011). Briles *et al*. (1992) demonstrated that exists a relationship between capsular serotype and virulence. Loss of capsule reduces *S. pneumoniae* virulence in

animal models (Morona *et al*., 2004). In systemic infections such as pneumonia and bacteremia the capsule protects pneumococci from the host immune system by impeding opsonization and subsequent phagocytosis (van der Poll and Opal, 2009; Hyams *et al*., 2010). The capsule also confers advantages to the pathogen by preventing its physical removal by mucus (Nelson *et al*., 2007). In addition, it restricts autolysis and reduces the exposure to antibiotics (Kadioglu *et al*., 2008).

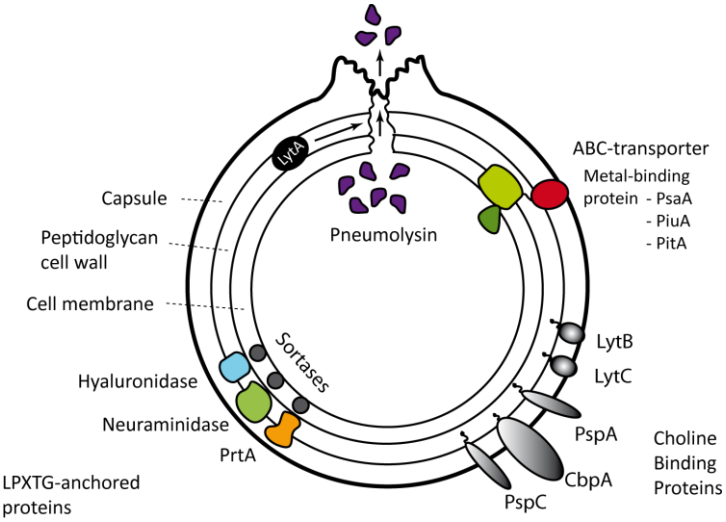


Figure 4. Main pneumococcal virulence factors. *S. pneumoniae* resides as a commensal in the upper respiratory tract of healthy people. However, it can cause several diseases. Important virulence factors of the pneumococcus include the polysaccharide capsule, choline-binding proteins, LPXTG-anchored proteins, ABC-transporters and both, the autolysin (LytA) and the pneumolysin (PLY; haemolysin) (modified from Kadioglu *et al.*, 2008).

The toxin pneumolysin (PLY; haemolysin) has been found in almost all pneumococcal clinical isolates. It is produced as a soluble protein that oligomerizes in the membrane of target cells forming a transmembrane pore, which finally provokes cell lysis. In addition to its citolytic activity, PLY is able to activate the classical complement pathway, producing a reduction of serum opsonic activity (Paton *et al*., 1983). However, how PLY is able to prevent complement deposition on the bacterial surface is unknown. Its role in diseases such as bacteremia and pneumonia was reported (Hirst *et al*., 2004; reviewed in Mitchell and Mitchell, 2010).

On the surface of the pneumococcus there are a variety of proteins anchored either by the G+ LPXTG attachment motif or by non-covalent interactions with the choline present in both the cell wall teichoic acid or in the membrane-associated lipoteichoic acid (choline-binding proteins; CBPs). The group of LPXTG-anchored surface proteins includes the enzymes hyaluronidase, neuraminidase, and the serine protease PrtA. Hyaluronidase degrades hyaluronic acid, a component of the mammalian connective tissue, thus acting as a spreading factor during pneumococcal infection and facilitating bacterial colonization of new host niches (Hynes and Walton, 2000). The enzyme

neuraminidase cleaves N-acetylneuraminic acid from glycolipids, glycoproteins, lipoproteins and oligosaccharides, causing a direct damage to the host. Also, it probably favors colonization as a result of its action on glycans. Among all pneumococcal enzymes with neuraminidase activity, neuraminidase A (NanA) has been shown to have a role in colonization and development of otitis media in animal models (Tong *et al*., 2000). The cell wall-associated serine protease PrtA has been reported to be involved in virulence in a mouse-peritonitis infection model, since mice infected with a PrtA-deficient mutant lived longer than mice infected with the wild-type strain (Bethe *et al*., 2001). Although the exact role of PrtA is unknown, proteases are generally involved in virulence as a result of its capability to cleave host proteins such as immunoglobulins, complement compounds and proteins of the extracellular matrix. The LPXTG-containing proteins are covalently attached to the cell surface by the action of sortase enzymes. In the pneumococcus, inactivation of sortase A (SrtA) causes the release of neuraminidase and hyaluronidase into the surrounding medium and reduces the adherence of the non-capsular R6 pneumococcal strain to human cells (Kharat and Tomasz, 2003; Paterson and Mitchell, 2006). Regarding the CBPs, it is known that they contribute to pneumococcal virulence in different ways. This protein family includes the hydrolytic enzymes LytA (autolysin), LytB, LytC, CbpA and CbpE among others. These factors are involved in nasopharyngeal colonization and adhesion to host tissues. In addition, LytA releases other virulence factors such as the cytoplasmic PLY and cell wall degradation components that can provoke inflammatory responses of the host. Moreover, there are CBPs that protect the pneumococcus against the host complement system such as PspA and PspC (reviewed in Mitchell and Mitchell, 2010).

Many other proteins related to the transport of different compounds like manganese (Mn^{2+}) and iron (Fe^{2+}) have been also involved in pneumococcal virulence. Focused on the pneumococcal infection process, Mn^{2+} is available in the human nasopharynx but it is more restricted at internal sites. It has been shown that pneumococci require Mn^{2+} to grow in the presence of Fe^{2+} under aerobic conditions (Johnston *et al*., 2004) and it is needed for the activity of CpsB, which is involved in the regulation of capsule production (Bender and Yother, 2001). Because of its importance and its relative low accessibility within the human body, the pneumococcus requires an efficient system of transport and utilization of Mn^{2+}. The Mn^{2+} transport system is an ABC-type permease encoded by the *psaBCA* genes. PsaA is a lipoprotein, which initially binds the metal on the cell surface. PsaB (ATP-binding protein) and PsaC (hydrophobic membrane protein) are also components of the transport permease. A gene for thiol peroxidase (PsaD) is both transcribed separately and also co-transcribed within the *psaBCAD* transcript. Although

S. pneumoniae lacks catalase, the presence of PsaD protects bacteria against oxidative damage caused by hydrogen peroxide generated as a product of pyruvate oxidation during growth in the presence of oxygen (Tseng *et al*., 2002; McAllister *et al*., 2004; Hajaj *et al*., 2012). Fe^{2+} is an essential element in the metabolism of most bacteria since it is used as a cofactor by many proteins. *S. pneumoniae* has a limited demand for iron because it lacks a respiratory chain and it does not possess cytochromes. Nevertheless, Fe^{2+} is required for enzymes such as ribonucleotide reductase, which catalyses the formation of deoxyribonucleotides from ribonucleotides and, therefore, is involved in the synthesis of DNA. However, the availability of iron in the host is very restricted because of its chelation by host iron-binding proteins such as lactoferrin, haemoglobin and transferrin (Wooldridge and Williams, 1993). Its low availability constitutes a nutritional barrier to infection. In *S. pneumoniae*, three iron-uptake ABC transporters have been identified, termed Pit (*pitADBC*), Pia (*piaABCD*) and Piu (*piuBCDA)*. Each one comprises four proteins: one iron carrier-binding protein, two heterodimer-forming permeases and one ABC ATPase (Brown *et al*., 2001; Brown *et al*., 2002). Mutation of both *pia* and *piu* leads to attenuation of *S. pneumoniae* virulence in a murine model of septicemia and pneumonia (Brown *et al*., 2001; Jomaa *et al*., 2005).

5. The pneumococcal Mga*Spn* regulator

The genomic sequence of the *S. pneumoniae* TIGR4 strain (a serotype 4 clinical isolate) showed that about 5% of its genome is composed of insertion sequences that may contribute to genome rearrangements through uptake of foreign DNA (Tettelin *et al*., 2001). Signature-tagged mutagenesis experiments in TIGR4 led to the identification of several TCSs and other putative transcriptional regulators that may play a role in the ability of the bacterium to adapt to changing environments (Hava and Camilli, 2002). One of such putative regulators was named MgrA (**Mg**a-like **r**epressor **A**) due to its homology to the multiple gene regulator Mga of GAS (51% of similarity and 25% of identity) (Hemsley *et al*., 2003). This protein is encoded by the *sp1800* gene (*mgrA*), which is highly conserved in the pneumococcal genomes of those that have been partially or totally sequenced. Using murine infection models, MgrA was shown to be required for both nasopharyngeal colonization and lung infection, but not for the ability of pneumococci to cause bacteremia. Microarray experiments carried out comparing the transcriptional profile of two mutant strains, one lacking the *mgrA* gene and another overexpressing *mgrA*, indicated that MgrA acts as a repressor of genes located within the *rlrA* pathogenicity islet (Hemsley *et al*., 2003). Some of these genes had been characterized previously as virulence factors (Hava *et al*., 2003). However, in contrast to

the *mgrA* gene, the *rlrA* pathogenicity islet has been found only in a small number of pneumococcal strains. These findings suggested that the islet might not be the main target of MgrA and, therefore, novel MgrA-regulated genes could be identified in other pneumococcal strains and/or under different bacterial growth conditions. In fact, some transcriptional regulators can alter the transcriptional profile in a different manner depending on the bacterial strain and/or serotype (Hendriksen *et al*., 2009; Hendriksen *et al*., 2007).

In 2001, the genome of the pneumococcal R6 strain, which derives from the D39 clinical isolate (serotype 2), was totally sequenced (Hoskins *et al*., 2001). Unlike the TIGR4 strain, both R6 and D39 lack the *rlrA* pathogenicity islet (Hoskins *et al*., 2001; Lanie *et al*., 2007). In 2008, Paterson *et al.* reported that the R6 strain contains a putative *mga*-like gene (*spr1404*). It is adjacent to a gene (*spr1403*) that encodes a collagen-like protein (PclA). Both genes are located in the called R6-specific cluster (Brückner *et al*., 2004), a region of 9.6 kb that is absent in TIGR4. The *spr1404* gene product has homology (40% of similarity) to the Mga virulence regulator of GAS. However, single-deletion mutants lacking either *spr1404* or *spr1403* were not attenuated in a mouse model of invasive pneumonia (Paterson *et al*., 2008). Thus, as pointed out by the authors, further work is required to elucidate whether the *spr1404* gene has a significant role in pathogenesis.

Searching for homologies, we found another putative *mga*-like gene in strain R6 (gene *spr1622*, from now on named *mgaSpn*), which is equivalent to the *sp1800* (*mgrA*) gene of the TIGR4 strain. Mga*Spn* (493 residues) differs from MgrA in two amino acid residues. EMBOSS needle global sequence alignment (Rice *et al*., 2000) of the Mga*Spn* protein and the Mga virulence regulator (530 amino acids; encoded by the M6_Spy1720 gene of *S. pyogenes* MGAS10394) revealed 42.6% of similarity and 21.4% of identity. Similar alignment analyses of the Mga*Spn* protein and the AtxA virulence regulator (475 amino acids; encoded by the *atxA* gene of plasmid pXO1 of *B. anthracis*) revealed 39.9% of similarity and 20.7% of identity. Because of its similarity with both regulators, we proposed at the beginning of this work that the pneumococcal Mga*Spn* protein could be considered as a member of the Mga/AtxA family of transcriptional regulators, and it might act as a global response regulator activating and/or repressing the expression of virulence genes.

6. Significance of this work

The pneumococcus remains a major cause of morbidity and mortality worldwide. Understanding the molecular mechanisms that control the expression of pneumococcal virulence genes in response to environmental stimuli offers new insights into the pathogenesis of this bacterium, and gives new visions to improve interventions. Due to their role in viability and virulence of bacterial pathogens as well as its absence in vertebrates, TCSs have received attention as potential antimicrobial targets (Barrett and Hoch, 1998). However, novel pneumococcal transcriptional regulators that are emerging, as it is the case of the *stand-alone* regulator Mga*Spn,* might have important roles in the virulence of *S. pneumoniae*. Therefore, the molecular characterization of Mga*Spn* may reveal new mechanisms used by the pneumococcus to control virulence gene expression in response to external stimuli. Although the role in pathogenesis of other members of the Mga/AtxA family of regulators (specifically Mga and AtxA) is well known, studies of both regulators at a molecular level are still very limited. When this work was started, we knew almost nothing regarding the pneumococcal Mga-like regulator except that *(i)* it was involved in nasopharyngeal colonization and development of pneumonia in murine infection models (Hava and Camilli, 2002), and *(ii)* it was a repressor of the *rlrA* pathogenicity islet (Hemsley *et al*., 2003). Therefore, this Doctoral Thesis has been focussed on the molecular characterization of the Mga*Spn* transcriptional regulator. Our findings have contributed to increase the knowledge of this pneumococcal virulence regulator.

Objectives

Searching for homologies, we found that the genome of *S. pneumoniae* encoded a potential global response regulator of the Mga/AtxA family. Such a regulator (here named Mga*Spn*) had been previously associated with virulence. The molecular characterization of Mga*Spn* has been the main aim of this Thesis because of its likely contribution to fully understand the pathogenic mechanisms of *S. pneumoniae*. To this end, we have worked on the following objectives:

1. Identification of the *mgaSpn* gene promoter.

2. Development of a procedure for large-scale purification of the untagged Mga*Spn* protein.

3. Analysis of the oligomerization state of Mga*Spn* in solution and determination of its secondary structure content.

4. Identification of Mg*Spn*-regulated genes: analysis of the effect of Mga*Spn* on the expression of the *spr1623-spr1626* operon.

5. Study of the interaction of Mga*Spn* with linear double-stranded DNAs.

Materials and Methods

Materials

1. Bacterial strains

Table 1. Bacterial strains used in this work

Strain	Features	Source
S. pneumoniae 708	*end*-1, *exo*-1, *trt*-1, *hex*-4, *malM*594	(Lacks and Greenberg, 1977)
S. pneumoniae R6	Nonencapsulated strain derived from the serotype 2 clinical isolate D39	(Lacks *et al*., 1986)
S. pneumoniae R6Δ*mga*	Strain derived from R6. It lacks the *mgaSpn* gene	This work
E. coli BL21 (DE3)	λ DE3 (*lac*I *lac*UV5-T7 gene 1 *ind*1 *sam*7 *nin*5) F- *dcm omp*T *hsd*S(*r*B -*m*B+) *gal*	(Studier and Moffatt, 1986)

2. Culture media

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. In the case of *S. pneumoniae* R6Δ*mga* strain, chloramphenicol (Cm) was added to a final concentration of 1.5 µg/ml. Plates for bacterial growth in solid medium were freshly prepared as is reported in Methods, Section 1. Pneumococcal cells harbouring plasmids were grown in media supplemented with tetracycline (Tc; 1 µg/ml) and/or kanamycin (Km; 30-50 µg/ml). Competent cells were grown in AGCH medium supplemented with 0.2% sucrose and 70 µM $CaCl_2$. *E. coli* cells were grown in tryptone-yeast extract (TY) medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) (Maniatis *et al*., 1982). In the case of plasmid-harbouring cells, the medium was supplemented with Km (30-50 µg/ml). Plates were prepared with TY and 1.5% agar. 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 reactives

Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase and bovine serum albumin (BSA) were acquired from New England BioLabs. DNase I RNase-free, isopropyl-β-D-thiogalactopyranoside (IPTG), proteases inhibitor cocktail EDTA-free complete, proteases inhibitor cocktail complete, High Pure Plasmid Isolation kit, rNTPs and dNTPs were provided by Roche Applied Science. Phusion High-Fidelity DNA Polymerase (Finnzymes) was used. ThermoScript reverse transcriptase, RNase inhibitor Superase•In and protein pre-stained standard SeeBlue Plus were purchased to Invitrogen. HyperLadder DNA molecular weight marker was from Bioline. Proteinase K, RNase A, lysozyme, tetracyclin (Tc), polyethylenimine (PEI), imidazole, His-select nickel affinity gel, and acrylamide:bis-acrylamide 40% solution (29:1 ratio) were from Sigma Aldrich. Ethanol absolute, hydrochloric acid, chloroform, isopropanol, aminoacids, vitamins, carbohydrates (sucrose and glucose) and magnesium were purchased to Merck. The saturated phenol was from AppliChem. Culture media components were from Pronadisa, Merck, Sigma, BD and Difco. Low molecular weight protein marker and the HiLoad Superdex 200 column were from Amersham Biosciences. DNA sequencing kit (Sequenase Version 2.0) was from USB Corporation. Radioactive nucleotides were purchased to Perkin-Elmer, Amersham Biosciences or Hartmann. Agarose, acrylamide:bis-acrylamide 30% solution (37.5:1 ratio), ammonium persulfate (PSA),β-mercaptoethanol, sodium dodecyl sulfate (SDS), Triton-X100, Bio-Safe Coomassie Stain, RNA isolation kit (Aurum Total RNA Mini kit), Immun-blot PVDF membranes, the Affi-Gel heparin gel for protein purification and the Econo-Column chromatography columns were from Bio-Rad. Acrylamide:bis-acrylamide 40% solution (19:1 ratio) was from National Diagnostics. The HiTrap Heparin HP and HisTrap HP columns and Illustra MicroSpin^{TM} G-25 columns were from GE Healthcare. For gel extraction or cleanup of DNA from enzymatic reactions (PCR or enzymatic digestion) the QIAquick Gel Extraction kit from QIAGEN was used. Dialysis membranes were from Spectrum. Autoradiography films were acquired from Kodak (X-Omat S). Cronex Lightning Plus amplifying x-ray screens were from Dupont. Imaging plates to visualize radioactive labelling using the Fujifilm Image Analyzer FLA-3000 were from Fuji. Centrifugal filters for protein concentration were from Pall. 96-well plates for fluorescence measurements were from Millipore.

4. Nucleic acids

4.1. Plasmids

Table 2. Plasmids used in this work.

Plasmid	Size (bp)	Description	Source
pMV158	5,540	Isolated from *S. agalactiae*: mobilizable by pAMβ1; Tc^R	(Burdett, 1980)
pLS1	4,408	Derivative of pMV158 that lacks the 1,132-bp *Eco*RI restriction fragment; non-mobilizable; Tc^R	(Stassi *et al*., 1981)
pET-24b	5,309	*E. coli* expression vector based on the Φ10 promoter of phage T7; Km^R	Novagen
pET-*mgaSpn*	6,821	Derivative of pET-24b that encodes an untagged Mga*Spn* protein; Km^R	This work
pET-mga*Spn*-His	6,790	Derivative of pET-24b that encodes a His-tagged Mga*Spn* protein; Km^R	This work
pAS	5,210	Terminator-probe vector. Derivative of pLS1 that carries the *gfp* reporter cassette of the pGreenTIR plasmid; Tc^R	(Ruiz-Cruz *et al*., 2010)
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)
pAST2 (pAS-*T2T1rrnB*)	5,456	Derivative of pAS that carries the transcriptional termination sites *T1T2* of the *E. coli rrnB* ribosomal RNA operon in the opposite orientation (compared to the pAST vector); Tc^R	(Ruiz-Cruz *et al*., 2010)
pAS-*Pmga*	5,352	Derivative of pAS that carries the promoter region of the *mgaSpn* gene (159440-1598305); Tc^R	This work
pAST2-*Pmga*	5,757	Derivative of pAST2 that carries the promoter region of the *mgaSpn* gene (1598600-1598304); Tc^R	This work
pAST-*PAB*	5,757	Derivative of pAST that carries the promoter region of the *spr1623-1626* operon (1598304-1598600); Tc^R	This work
pAST-*PAB*Δ84	5,672	Derivative of pAST that carries the promoter region of the *spr1623-1626* operon (1598388-1598600); Tc^R	This work
pAST-*PAB*Δ153	5,602	Derivative of pAST that carries the promoter region of the *spr1623-1626* operon (1598457-1598600); Tc^R	This work
pDL287	5,740	Derivative of pVA380-1; Km^R	(LeBlanc *et al*., 1993)
pDL*PsulA*::*mga*	7,529	Derivative of pDL287 containing the *PsulA::mga PsulA-mga* fusion gene into the *Cla*I site; Km^R	This work

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

4.2. Oligonucleotides

The oligonucleotides used in this work were all synthesized at the CIB-Protein Chemistry Facility (Applied Biosystems 3400 synthesizer), purified by HPLC and resuspended in water to a final concentration of 100 µM. Oligonucleotides used to obtain dsDNA fragments for electrophoretic mobility shifts assays were also purified from polyacrylamide denaturing gels (15-20% PAA, 8 M urea).

Table 3. Oligonucleotides

Name	Bases	Sequence (5´-3´)	Applications
1622Nde	27	GAGAGAAAGATA**CATATG**AGAGATTTA	Mga*Spn* purification
1622Xho	26	GGTACAGTTCAAA**CTCGAG**ATAGCGT	Mga*Spn* purification
1622XhoHis	30	TTTTGTTATTTT**CTCGAG**CTCATCTAATCG	Mga*Spn*-His purification
C1622D	25	CTAAAAAAAGTCATAGGCAATTAGA	RT-PCR
1622A	20	AGTTCCTGATTGTATTCCCT	RT-PCR. EMSA. EM
1622C	20	GATTCTGTATTCACGCCCTC	RT-PCR. EMSA. Footprinting
1622D	26	TTCTAATTGCCTATGACTTTTTTTAG	RT-PCR. EMSA. Footprinting
1622B	20	CACAACACTGCCTACCCTCC	EMSA. EM
1622E	21	TAGATGAAGAAGTTGTTTGCC	EMSA
pUC-Rev	21	TTGTGAGCGGATAACAATTTC	EMSA

Name	Bases	Sequence (5´-3´)	Applications
pUC-A	24	GGCTGCGCAACTGTTGGGAAGGGC	EMSA
PrSp1	26	ATAAATTATC**GGATCC**AACCTCTTGC	Construction of pAS-*Pmga* plasmid
PrSp2	26	GAATTTGATTCT**GGATCC**ACGCCCTC	Construction of pAS-*Pmga* plasmid
UpSph	29	CCGTCTATTGAGGGCGT**GCATGC**AGAATC	Construction of R6Δ*mga*
UpCla	29	TTTGACATATACA**ATCGAT**TCGATTTAAC	Construction of R6Δ*mga*
DwSal	33	GTCTCACTCATATACTT**GTCGAC**TGCCATGATG	Construction of R6Δ*mga*
DwCla	34	CTATTCTTTTTCATAC**ATCGAT**CATAATTATCAG	Construction of R6Δ*mga*
CmSph	32	CTACAGAAAGTAAA**GCATGC**AAAGAGTAATGC	Construction of R6Δ*mga*
CmSal	34	GCGAAAAAGGAGAA**GTCGAC**TCAGAAAAAGAAGG	Construction of R6Δ*mga*
91G_2	23	GGCTATTTTGATGCACATATCTG	EMSA
92A_2	21	CCCGCCTTCCTTCCCTTGCTC	EMSA
INTgfp	22	CATCACCATCTAATTCAACAAG	Primer Extension
Oligo1	40	TATATTGTCTCCGTAGTGTTATTATACGAAATAAAAGATT	EMSA
Oligo1C	40	AATCTTTTCTTTCGTATAATAACACTACGGAGACAATATA	EMSA

Name	Bases	Sequence (5´-3´)	Applications
Oligo4	20	CTCCGTAGTGTTATTATACG	EMSA
Oligo4C	20	CGTATAATAACACTACGGAG	EMSA
Oligo2	20	TATATTGTCTCCGTAGTGTT	EMSA
Oligo2C	20	AACACTACGGAGACAATATA	EMSA
Oligo3	20	ATTATACGAAATAAAAGATT	EMSA
Oligo3C	20	AATCTTTTATTTCGTATAAT	EMSA
Oligo 40p	40	TATATCATGCTATACCTATTCTTTGTGGTATAATTGCAAG	EMSA
Oligo 40pC	40	CTTGCAATTATACCACAAAGAATAGGTATAGCATGATATA	EMSA
Oligo 32A	32	TTCTTTGTGGTATAATTGCAAGAGGTTTAATC	EMSA
Oligo 32B	32	GATTAAACCTCTTGCAATTATACCACAAAGAA	EMSA
Oligo 26A	26	TTCTTTGTGGTATAATTGCAAGAGGT	EMSA
Oligo 26B	26	ACCTCTTGCAATTATACCACAAAGAA	EMSA
1622F	23	CGATGAAACCAACGTTTATGTTC	*In vitro* transcription
1623A	20	GAGGGCGTGAATACAGAATC	RT-PCR

Name	Bases	Sequence (5´-3´)	Applications
1623B	24	CGTAAATTTACATGAACAGTTGGG	RT-PCR
1623C	23	GGAGGGTAGGCAGTGTTGTGATC	RT-PCR
1626A	24	GCACCTTCTACAGCGTCTTTAGCG	RT-PCR
PsulNde	26	CAAGGATTTTCAT**CATATG**ATTTTTC	Construction of pDL*PsulA::mga* plasmid
PsulCla	28	ACTGATTGTTA**ATCGAT**TTGCTTTCTGT	Construction of pDL*PsulA::mga* plasmid
mgaNde	31	TGCAAGAGGTTT**CATATG**ATAATTTATAAAG	Construction of pDL*PsulA::mga* plasmid
mgaCla	33	GTACATTTTTCTTA**ATCGAT**TGAAGGTCTTTTC	Construction of pDL*PsulA::mga* plasmid
EM1	25	AGTTGAATGTTTAAAGAAATGATGG	EM in combination with 1622B
EM5	27	CAATACAAATATTGTTTTGAAGAAGCC	EM in combination with 1622F
PmgaSac	30	CTTTATAAATTAT**GAGCTC**AAACCTCTTGC	Construction of pAST2-*Pmga* plasmid
PABSac	31	ATATCAAAAAATC**GAGCTC**TTTGATTATTAC	Construction of pAST-*PAB* plamid
PABΔ84Sac	32	ATTTCGTATAA**GAGCTC**TACGGAGACAATATA	Construction of pAST-*PAB*Δ84 plasmid
PABΔ153Sac	28	GAATACAGAATC**GAGCTC**AAGTCTAAAG	Construction of pAST-*PAB*Δ153 plasmid
PDA	23	GTGATTTTACCTGCCAAGAGACC	Primer Extension

Name	Bases	Sequence (5´-3´)	Applications
PDB	22	GAAAAGTCAATTATTTCGATTG	Primer Extension
PErpoE	23	GCCCAGCAAATACTTCTAATTCC	Primer Extension (internal control)
ASTtetL	23	GAGGGCAGACGTAGTTTATAGGG	Primer Extension (internal control)
1622H	26	CGGATTAAACCTCTTGCAATTATACC	EMSA. Footprinting
1622I	24	CAAATTCTTTAATTGTTGCTATTA	EMSA. Footprinting

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

4.3. Linear double-stranded DNA fragments

All dsDNA fragments were synthesized during the development of this work.

Table 4. Linear dsDNA fragments

Name	Size (bp)	Coordinates	Features	Applications
1622CF	265	1598188-1598452	Includes *Pmga* promoter	*In vitro* transcription
1622CD	224	1598229-1598452	Includes *Pmga* promoter	EMSA. Footprinting. *In vitro* transcription.
1622HI	222	1598298-1598519	Includes *Pmga* and *PB* activation site	EMSA. Footprinting.

Name	Size (bp)	Coordinates	Features	Applications
1622AE	282	1597232-1597513	From coding region of *mgaSpn*	EMSA
1622BD	421	1598229-1598649	Includes *Pmga* promoter	EMSA
91G-92A2 (C)	321	94488-94808	DNA fragment intrinsically curved	EMSA
EM1-26A (NC)	322	1598010-1598331	Control without curvature	EMSA
20	20	1598376-1598395	Small dsDNA	EMSA
26	26	1598306- 1598331	Small dsDNA	EMSA
32	32	1598300-1598331	Small dsDNA	EMSA
40	40	1598310-1598349	Small dsDNA	EMSA
EM1-1622B	640	1598010-1598649	Includes *Pmga* and *PB* activation region	EM
1622BA	1418	1597232-1598649	Includes *Pmga* and *PB* activation region	EM
EM5-1622F	1458	1598188-1599645	Includes *Pmga* and *PB* activation region	EM

5. Acrylamide solutions

For protein electrophoresis, an acrylamide:bis-acrlylamide 40% solution (40:1) or 30% solution (37.5:1) was used. For nucleic acids, an acrylamide:bis-acrlylamide 30% solution (30:0.8) or 40% solution (29:1) for non-denaturing gels (native gels) and an acrylamide:bis-acrlylamide 38% solution (38:2) or 40% solution (19:1) for denaturing gels, were used.

6. Buffer solutions

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

Table 5. Buffers

Buffer	Composition	Application
TE	10 mM Tris-HCl, pH 8.0 1mM EDTA	Storage of 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
TG	50 mM Tris-HCl, pH 8.3 300 mM Glycine 0.1% SDS 2 mM EDTA	Protein electrophoresis: SDS-PAGE (Tris-Glycine)
Buffer gel	3 M Tris-HCl, pH 8.45 0.3% SDS	Protein electrophoresis: SDS-PAGE (Tris-Tricine)
Cathode buffer	100 mM Tris-HCl, pH 8.25 100 mM Tricine 0.1% SDS	Protein electrophoresis: SDS-PAGE (Tris-Tricine)
Anode buffer	200 mM Tris-HCl, pH 8.9	Protein electrophoresis: SDS-PAGE (Tris-Tricine)
SLB 5X	250 mM Tris-HCl, pH 7.2 10% SDS	Loading-dye for protein electrophoresis (SDS-PAGE)

Buffer	Composition	Application
	3.5 M β-mercaptoethanol 50% Glycerol 0.5% Bromophenol blue	
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
V-His	10 mM Tris-HCl, pH 7.6 5% Glycerol 300 mM NaCl 5 mM β-mercaptoethanol	Purification of Mga*Spn*-His protein. Buffer V-His was supplemented with imidazole (10 or 250 mM)
P	20 mM Tris-HCl, pH 7.6 5% Glycerol 250 mM NaCl 1 mM EDTA 1 mM DTT	Storage of Mga*Spn*-His
VL	50 mM Tris-HCl, pH 7.6 5% Glycerol 1 mM DTT 1 mM EDTA	Purification and storage of Mga*Spn* protein. Buffer VL was supplemented with different concentrations of NaCl
FXBE	80% Formamide 0.1% Xylene cyanol 0.1% Bromophenol blue 10 mM EDTA	Loading-dye for RNA electrophoresis
AU	20 mM Tris-HCl, pH 7.6 1 mM EDTA 0.1 mM DTT 5% Glycerol 250 mM NaCl	Protein analysis by analytical ultracentrifugation
EB	200 mM NaCl 20 mM Tris-HCl, pH8 2 mM EDTA	Elution of DNA from native PAA (5%) gels
OEB	10 mM Magnesium acetate 200 mM Sodium chloride 0.1% SDS	Elution of oligonucleotides from denaturing PAA gels

Buffer	Composition	Application
EM	20 mM Tris-HCl, pH7.5 $MgCl_2$ 10 mM	Mica adsorption in electron microscopy
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 PAA gels
TE (2:0.2)+NaCl	2 mM Tris-HCl, pH 8.0 0.2 mM EDTA 50 mM NaCl	Oligonucleotides annealing buffer
STOP DNase I	2 M Ammonium acetate 0.8 mM Sodium acetate 0.15 M EDTA	Stop solution for DNase I digestion
Transcription Buffer 5X	200 mM Tris-HCl, pH 7.5 750 mM KCl 50 mM $MgCl_2$ 0.05% Triton X-100	*In vitro* transcription
STOP-T	2% SDS 100 mM EDTA	Stop solution for *in vitro* transcription
LB	50 mM Tris-HCl, pH 7.6 1 mM EDTA 50 mM NaCl 0.1% Deoxycholate	Lysis buffer for RNA isolation and for total extract preparations from *S. pneumoniae*
LBP	20 mM Tris-HCl, pH 8.5 10 mM DTT	Lysis of pneumococcal cells for Proteomics
PBS	10 mM Na_2HPO_4 2 mM KH_2PO_4 2.7 mM KCl 137 mM NaCl pH 7.4	Resuspension of pneumococcal cells for fluorescence measurements. Washing buffer for Proteomics. Washing buffer in Western-bots
CD	10 mM Potassium phosphate 100 mM Ammonium sulphate pH 7.6	Circular dichroism analysis
TB	25 mM Tris	Western-blot Transfer Buffer

Buffer	Composition	Application
	192 mM Glycine 20% Methanol	
WB	PBS 0.05% Tween20	Western-blot Washing Buffer
SB	PBS 0.05% Tween20 0.2% Casein	Western-blot Saturation Buffer

7. Bioinformatics tools

Table 6. Bioinformatics tools

Applications	Program	Company/Webpage
Homologies finder	BLAST	blast.ncbi.nlm.nih.gov/Blast.cgi
Promoter finder	BPROM	linux1.softberry.com/berry.phtml
Open Reading Frame Finder	ORF Finder	ncbi.nlm.nih.gov/projects/gorf
Analysis oligonucleotides	OligoAnalyzer 3.1 Tm calculator	eu.idtdna.com finnzymes.fi/tm_determination.html
Restriction maps	ApE	biology.utah.edu/jorgensen/wayned/ape
Analysis DNA sequences	Chromas ApE	mb.mahidol.ac.th/pub/chromas biology.utah.edu/jorgensen/wayned/ape
Sequence alignments	ClustalW Clustal Omega	ebi.ac.uk/Tools/clustalw2 ebi.ac.uk/Tools/msa/clustalo
Primary protein structure analysis	ProtParam	web.expasy.org/protparam
Secondary protein structure prediction	PsiPred Jpred SABLE PredictProtein NPS@	bioinf.cs.ucl.ac.uk/psipred compbio.dundee.ac.uk/www-jpred sable.cchmc.org predictprotein.org npsa-pbil.ibcp.fr
Deconvolution of CD spectra	SELCON3	dichroweb.cryst.bbk.ac.uk

Applications	Program	Company/Webpage
	CONTINLL	
	CDSSTR	
	K2D	
Prediction of functional domains	Pfam	pfam.sanger.ac.uk
	Phyre2	www.sbg.bio.ic.ac.uk/phyre2
Sedimentation equilibrium	HeteroAnalysis	biotech.uconn.edu
Sedimentation velocity	SEDFIT	analyticalultracentrifugation.com
	SEDNTERP	jphilo.mailway.com
3D-Strucutres Visualization	PyMol	pymol.org
Radiolabelled DNA visualization	Image-reader	Fuji
	MultiGauge	Fuji
	Quantity One	Bio-Rad
Radiolabelled DNA visualization	Image-reader	Fuji
	MultiGauge	Fuji
	Quantity One	Bio-Rad
Non-labelled DNA visualization	Quantity One	Bio-Rad
CD Analysis	SpectraManager	Jasco
Data analysis and graphing	SigmaPlot	SigmaPlot
DNA bending analysis	Bend.it	hydra.icgeb.trieste.it/dna

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 (Dupon).

Methods

1. Bacterial growth conditions

In liquid medium, *E. coli* cells were grown at 37°C with rotary shaking in Erlenmeyer flasks. The flask volume was 5-times greater than the culture volume in order to maintain a constant aeration. In solid medium, cells were spread uniformly over TY-agar. In the case of *S. pneumoniae*, cells were grown under low aeration conditions at 37°C. In liquid medium, cells were grown in a static bath in conditions in which the volume of the flask was twice the volume of the culture. In solid medium, 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. In all cases, bacterial growth in liquid medium was followed by turbidity at 600 nm for *E. coli* and at 650 nm for *S. pneumoniae* using a Braush & Lomb (Spectronic 20D+) spectrophotometer.

For preservation of bacterial strains, cells were grown to an optical density at 650 nm (OD_{650}) of 0.3 (*S. pneumoniae*) or to an OD_{600} of 0.5 (*E. coli*), which correspond with the exponential phase of growth. 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.

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_{600} 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. Cell pellet was washed several times with cool sterile water. Finally, cells were resuspended in 10% glycerol, and aliquots (50 µl) were stored at -80°C. *S. pneumoniae* competent cells were prepared as reported previously (Lacks, 1966). Mainly, cells were grown in ACGH medium supplemented with sucrose (0.3%) and antibiotic (if required) at 37ºC to an OD_{650} of 0.3 (exponential phase). The culture was then diluted 1:40 with pre-warmed 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 electrocompetent cells were transformed as described previously (Dower *et al*., 1988). DNA in water (5-40µl) was mixed with electrocompetent cells (50 µl) in a 0.2 cm electroporation cuvette (Bio-Rad) pre-cooled on ice. The pulse was generated with a MicroPulser (Bio-Rad) (2.50 kV and 5 ms). After the electric pulse, cells were transferred to 0.8 ml SOB medium supplemented with glucose (0.4%) and incubated at 37°C with rotary shaking for 1 hour. Finally, cells were spread over TY-agar plates with the appropriate antibiotic for selection of transformants.

2.2.2. Natural transformation

Foreign DNA goes into competent pneumococcal cells by a process of natural transformation, which was previously reported (Espinosa *et al*., 1982). Briefly, cells were inoculated in AGCH medium supplemented with sucrose (0.2%) and $CaCl_2$ (70 µM) and incubated at 30°C for 20 min before addition of DNA. When strain R6 or derivatives were used, 25 ng of CSP-1 (**C**ompetence **S**timulating **P**eptide-1) was added at the same time as the DNA. After addition of DNA, cultures were incubated at 30°C for 40 min. In general, to enable the phenotypic expression (resistance to Tc or Km), cultures were incubated at 37°C for 90 min. However, to select transformants resistant to Cm, cultures were first incubated at 37ºC for 70 min. Then, Cm was added to a final concentration of 0.05-0.5 µg/ml and cultures were incubated at 37ºC for 20 min. Transformants were selected in AGCH plates (described in Section 1) supplemented with the appropriate antibiotic.

3. Construction of bacterial strains

3.1. Construction of *S. pneumoniae* R6Δ*mga* strain

For the construction of the *S. pneumoniae* R6Δ*mga* strain, gene replacement by homologous recombination was performed. Briefly, a 1,165-bp DNA fragment (Cm) that contained the pC194 *cat* gene, which confers Cm resistance (Horinouchi and Weisblum, 1982), was PCR-amplified using the CmSph and the CmSal primers. The PCR product was then digested with *Sph*I and *Sal*I. In addition, a 543-bp region (Up) and a 605-bp region (Down) that flank the *mgaSpn* gene (promoter plus coding sequence) were amplified using the UpSph and UpCla primers and the DwSal and DwCla primers, respectively. Then, the PCR products were digested with the corresponding restriction enzymes (*Sph*I and *Cla*I were used to digest the Up fragment and *Sal*I and *Cla*I were

used to digest the Down fragment). Once all the PCR products had been digested, the Up and Down fragments were ligated to the Cm fragment, generating the *Up::Cm::Down* fusion fragment, in which the regions flanking the *mgaSpn* gene are flanking the *cat* gene. The cassette generated *in vitro* was then used to transform competent *S. pneumoniae* R6 cells. Selection of transformants resistant to Cm (1.5 µg/ml) led to the isolation of the *S. pneumoniae* R6Δ*mga* mutant strain, which lacks the *mgaSpn* gene (including its promoter) (coordinates 1596826-1568431), as confirmed by dye terminator sequencing at Secugen (CIB, Madrid).

4. Fluorescence measurements

Fluorescence assays were carried out as described previously (Ruiz-Cruz *et al*., 2010). Briefly, pneumococcal cells harbouring a plasmid that carries the *gfp* reporter gene were grown to an OD_{650} of 0.3. For these analyses, different volumes of the culture (from 25 µl to 1 ml) were used. Cells were harvested by centrifugation at 4ºC and resuspended in 200 µl of PBS buffer (see Table 5). 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, three independent cultures were analysed. The fluorescence corresponding to 200 µl of PBS buffer without cells was also measured (values around 0.03 arbitrary units).

5. DNA preparations

5.1. Plasmid DNA isolation

For small-scale plasmid preparations, the High Pure Plasmid Isolation kit (Roche Applied Science) was used for both, *E. coli* and *S. pneumoniae*. In the case of *S. pneumoniae*, the kit specifications were modified at two steps: 1) *Suspension buffer* was supplemented with 50 mM glucose and 0.1% deoxycholate (DOC) for cell lysis, 2) *Lysis buffer* was prepared containing 0.170 M NaOH, 1% SDS for chromosomal DNA denaturation.

5.2. Genomic DNA extraction

To isolate chromosomal DNA from *S. pneumoniae*, cells were grown in 40 ml AGCH supplemented with 0.3% sucrose and 0.2% yeast extract to an OD_{650} of 0.8. Cultures were centrifuged, and cells were resuspended in 1 ml of TE buffer supplemented with 0.6 mg proteinase K and 0.6% SDS and incubated at 37ºC for 30 min with gentle shaking. Samples were phenol treated, centrifuged and the supernatant was

dialyzed against TE buffer. Then, 40 µg RNase A was added and the mixture was incubated at 37ºC for 30 min. Finally, samples were phenol treated, ethanol precipitated and resuspended in TE. For small-scale genomic DNA isolation, the Bacterial Genomic DNA Isolation kit (Norgen) was used as specified by the supplier.

5.3. Preparation of linear double-stranded DNA fragments

Linear dsDNA fragments were obtained by digestion with restriction enzymes, by Polymerase Chain Reaction (PCR) or by annealing of complementary oligonucleotides.

5.3.1. 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, the enzyme was inactivated at 65ºC for 10 min.

5.3.2. Polymerase chain reaction

All PCR reactions were carried out in an iCycler Thermo Cycler (Bio-Rad). Taq DNA polymerase (Invitrogen) was used for analysis of transformants (both *E. coli* and *S. pneumoniae*) by colony PCR. Essentially, a single colony was inoculated into 100 µl of growth medium. Then, 10 µl were added to 90 µl of water and 1 µl of this mixture was used as template for PCR. Reactions (25 µl) contained 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 2 mM $MgCl_2$, 20 pmol of each primer, 200 µM of each dNTP and 1 unit of Taq DNA polymerase. Phusion High-Fidelity DNA polymerase (Finnzymes) was used for the rest of the PCR applications. Reaction mixtures (50 µl) contained 5-30 ng of template DNA, 20 pmol of each primer, 200 µM of each dNTP and 1 unit of DNA polymerase. PCR conditions were: initial denaturation step at 94ºC for 3 min (Taq DNA polymerase) or at 98ºC for 1 min (Phusion DNA polymerase). Then, it was followed by 30 cycles including the next steps: denaturation at 94ºC for 45 s (Taq DNA polymerase) or at 98ºC for 10 s (Phusion DNA polymerase), annealing of the primers to the DNA template at around 55ºC (depending on the primer Tm) for 20-30 s followed by an extension at 72ºC for 20-40 s (depending on the amplicon length). A final extension step was performed at 72ºC for 10 min. The QIAquick PCR Purification kit (QIAGEN) was used to purify DNA from both restriction endonuclease digestion and PCR.

5.3.3. Annealing of complementary oligonucleotides

Complementary oligonucleotides were annealed in buffer TE (2:0.2) containing 50 mM NaCl. For non-radioactive oligonucleotides, equimolar amounts of each oligonucleotide were used. The reaction mixtures (150 µl) were incubated at 95ºC for 10 min and then cooled down slowly to 37ºC. Then, they were kept at this temperature for 10 min and on ice for 10 min. When preparing radiolabelled dsDNA, one of the complementary oligonucleotides was 5´-labelled using [γ-^{32}P]-ATP and T4 PNK. In this case, the concentration of the radioactively labelled oligonucleotide in the annealing reaction was twice the non-labelled oligonucleotide.

5.4. DNA purification

5.4.1. Oligonucleotides

Oligonucleotides used to obtain small dsDNA fragments (20-bp to 40-bp) were purified from PAA (15-20%) denaturing (8 M urea) gels (Maniatis *et al*., 1982). The oligonucleotide solution was mixed with deionised formamide (50% final concentration) and heated at 55ºC for 5 min. The sample was cooled down on ice before being loaded onto the sequencing gel. After electrophoresis, the gel was placed on a piece of saran wrap and then on a Fluor-coated TLC plate (Ambion) and shined with a short wavelength UV light. The DNA absorbs the UV light and casts a shadow against the fluorescent background allowing the DNA to be visualized. The area of the gel containing the full-length oligonucleotide was excised. To elute the oligonucleotide, the gel slice was incubated in buffer OEB (see Table 5) overnight at 42ºC with continuous shaking. The remains of gel in the eluted DNA solution were removed using a Spin-X column (Costar). Finally, the eluted oligonucleotide was ethanol precipitated, dissolved in 50 µl of distilled water and loaded onto a MicroSpin™ G-25 column (GE Healthcare) to eliminate the remains of salt.

5.4.2. Linear double-stranded DNAs

Linear dsDNA fragments obtained either by digestion with restriction enzymes or by PCR, were purified (when required) from preparative agarose gels or non-denaturing polyacrylamide gels. When the DNA fragments were purified from agarose gels, the QIAquick gel extraction kit (QIAGEN) was used, following the specifications of the supplier. To elute DNA fragments from polyacrylamide, gels were stained with ethidium bromide (EtBr) and the band was excised with a clean scalpel. Then, the gel slice was incubated overnight in elution buffer (see Table 5) with continuous shaking (450 rpm) at

42ºC using a Thermo-shaker (TS-100, Biosan). The eluted DNA was finally precipitated with ethanol. When the quality of the DNA preparations was good enough, it was purified with the QIAquick PCR purification kit (QIAGEN) without the gel purification step.

5.5. Ligation

In general, for ligation of linear DNA fragments, the molar ratio vector to insert was between 1:5 and 1:10. The reaction mixture was incubated at 22ºC for 20 min or at 16ºC overnight for DNAs with sticky ends or at 20ºC for 2 h for DNAs with blunt ends. In all cases, 400 units of T4 DNA ligase (New England Biolabs) were added to a final reaction volume of 20-40 µl.

5.6. Construction of recombinant plasmids

5.6.1. Construction of pET24b-mgaSpn and pET24b-mgaSpn-His plasmids

The pET24b-*mgaSpn* and pET24b-*mgaSpn*-His plasmids were used to overproduce and purify the Mga*Spn* and the Mga*Spn*-His proteins, respectively (see Section 9). In both cases, the pET24b expression vector (Novagen) was used. For overproduction of the Mga*Spn* native protein, a 1,540-bp region of the R6 genome containing the *mgaSpn* gene was amplified by PCR using the 1622Nde and 1622Xho primers. These primers contained a single restriction site for *Nde*I and *Xho*I, respectively. The amplified product was digested with both enzymes, and the 1,512-bp digestion product was cloned into the pET24b expression vector. For overproduction of the Mga*Spn*-His protein, the *mgaSpn* gene was engineered to encode a His-tagged Mga*Spn* protein. Specifically, a 1,512-bp region of the R6 genome was amplified by PCR using the 1622Nde and 1622Xho-His primers, generating target sites for *Nde*I and *Xho*I restriction enzymes, respectively. Then, the amplified DNA fragment was digested with both enzymes, and the 1,481-bp digestion product was cloned into the pET24b expression vector, which enables a C-terminal His_6-tag fusion.

5.6.2. Construction of pAS-Pmga plasmid

The pAS-*Pmga* recombinant plasmid was used to identify the transcription start site of the *mgaSpn* gene. For its construction, a 170-bp region of the R6 chromosome containing the *Pmga* promoter was amplified by PCR using the PrSp1 and PrSp2 primers, which contained a single restriction site for *Bam*HI. The amplified fragment was digested with *Bam*HI, and the 142-bp digestion product was cloned into the *Bam*HI site of plasmid pAS (Ruiz-Cruz *et al*., 2010). In this plasmid, expression of the *gfp* reporter gene remains under the control of the inserted promoter.

5.6.3. Construction of the pAST-PAB, pAST2-Pmga pAST-PABΔ84 and pAST-PABΔ153 plasmids

To construct pAST-*PAB* and pAST2-*Pmga*, a 333-bp region of the R6 chromosome was amplified with the PmgaSac and PABSac primers, generating restriction sites for *Sac*I. After digestion, the 301-bp restriction fragment, which included the *P1623A, P1623B* and *Pmga* promoters, was cloned into the *Sac*I site of plasmid pAST (Ruiz-Cruz *et al*., 2010) (pAST-*PAB*, in which *gfp* expression is under control of the *P1623A* and *P1623B* promoters) and pAST2 (Ruiz-Cruz *et al*., 2010) (pAST2-*Pmga*, in which *gfp* expression is under control of the *Pmga* promoter). To construct the pAST-*PAB*Δ84 plasmid, a 246-bp region of the R6 genome (promoters *P1623A* and *P1623B*) was amplified with the PABSac and the PABΔ84Sac primers, digested with *Sac*I, and the 216-bp product was cloned into the pAST plasmid. To construct the pAST-*PAB*Δ153 plasmid, a 177-bp region of the R6 genome (promoters *P1623A* and *P1623B*) was amplified with the PABSac and the PABΔ153Sac primers, digested with *Sac*I, and the 146-bp restriction fragment was cloned into the pAST plasmid. In pAST-*PAB*Δ84 and pAST-*PAB*Δ153, *gfp* gene expression is under control of the *P1623A* and *P1623B* promoters.

5.6.4. Construction of the pDLPsulA::mga plasmid

Plasmid pDL*PsulA::mga* is a derivative of pDL287, which carries a Km resistance gene (LeBlanc *et al*., 1993). The construction of pDL*PsulA::mga* involved several steps. First, a 189-bp region of the R6 genome, which contained the *PsulA* promoter (Lacks *et al*., 1995; Ruiz-Cruz *et al*., 2010), was amplified using the PsulNde and PsulCla primers. At the same time, a 1,650-bp region of the R6 chromosome was amplified using the mgaNde and the mgaCla primers. This fragment contained a promoter-less *mgaSpn* gene. Both PCR fragments were digested with *Nde*I (generating fragments of 172-bp and 1,636-bp, respectively) and ligated with the T4 DNA ligase, generating the *PsulA::mga* fusion gene. This DNA fragment was then amplified by PCR using the PsulCla and mgaCla primers. The amplified product was digested with *Cla*I, generating a 1,777-bp restriction fragment that was cloned into the *Cla*I site of plasmid pDL287 (LeBlanc *et al*., 1993).

5.7. 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)

5.7.1. 5´-end labelling

Oligonucleotides were radioactively labelled at the 5´-end using [γ-^{32}P]-ATP and the T4 PNK. Reactions mixtures (25 µl) contained 25-50 pmol of oligonucleotide, 2.5 µl of 10x kinase buffer (provided by the supplier), 40-80 pmol of [γ-^{32}P]-ATP (3,000 Ci/mmol; 10 µCi/µl) and 10 units of T4 PNK. After incubation at 37ºC for 30 min, additional T4 PNK (10 units) was added and the reaction mixture was incubated at 37ºC for 30 min. Finally, to inactivate the enzyme, reaction mixtures were incubated at 65ºC 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 either by PCR amplification (labelling at the 5´-end of one strand) or by annealing the labelled oligonucleotide to the non-labelled complementary oligonucleotide. The 5´-labelled oligonucleotides were also used in primer extension reactions and for manual sequencing.

5.7.2. Internal labelling

The incorporation of a radiolabelled nucleotide ([α-^{32}P]-dATP, [α-^{32}P]-dCTP, or [α-^{32}P]-UTP 3,000 Ci/mmol, Hartmann) was used for manual sequencing (using the Sequenase v 2.0 kit, following the indications of the supplier), primer extension experiments and *in vitro* transcription assays.

6. Analysis of DNA

6.1. DNA quantification

To quantify non-radiolabelled DNA, agarose or polyacrylamide gels were stained with EtBr (1 µg/ml). Bands were visualized using a Gel-doc system (Bio-Rad) and quantified using a molecular weight marker (HyperLadder I, Bioline; designed for easy size determination) with the Quantity One program. In addition, the concentration of DNA samples was determined using a NanoDrop ND-1000 Spectrophotometer (Bio-Rad). For 5´-labelled DNA, the ionizing radiation was measure with a scintillation counter (Wallac 1450 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.

6.2. DNA electrophoresis

6.2.1. Agarose gels

Horizontal agarose gel electrophoresis in TAE buffer was used to analyse chromosomal DNA, plasmid DNA and linear dsDNA fragments larger than 100 bp. DNA samples were mixed with BXGE buffer (see Table 5) and loaded onto the gel. The agarose concentration used (generally 0.8-1%) depended on the size of the DNA analysed. After electrophoresis, gels were soaked in a solution of EtBr (1 µg/ml in TAE buffer) for 15 min at room temperature. DNA was visualized using a short wavelength UV light (254 nm) for analytical agarose gels or a long wavelength UV light (360 nm) for preparative agarose gels, using a Gel-doc XR system (Bio-Rad). The image obtained was captured with the Quantity One software and quantified when it was necessary.

6.2.2. Polyacrylamide gels

4.2.2.1. Native polyacrylamide gels

Vertical polyacrylamide gels were run using a Mini Protean-III system (Bio-Rad). DNA samples were mixed with BXGE buffer and loaded onto the gels. Gels were run at 100 V and at 4ºC or room temperature in TBE buffer (see Table 5). The concentration of acrylamide used (5-15%) depended on the size of the DNA analysed. After electrophoresis, gels were stained with EtBr and visualized as described before. When the DNA was radioactively labelled, after electrophoresis, gels were fixed with acetic acid (10%), dried using a gel dryer (model 583, Bio-Rad) and the DNA was visualized by autoradiography or using a Fujifilm Image Analyzer FLA-3000.

6.2.2.2. Denaturing polyacrylamide gels

Oligonucleotides were purified (when needed) from denaturing polyacrylamide gels in TBE buffer and 8 M urea. The concentration of acrylamide used depended of the size of the oligonucleotide: 19% for 15-25 nt and 15% for 25-40 nt. After electrophoresis, oligonucleotides were purified as described before (see Section 5.4.1). Sequencing reactions, primer extension products, *in vitro* transcription products as well as footprinting reactions (DNase I and OH•) were run in sequencing gels (6% PAA, 8 M urea), pre-heated at 50ºC. A Sequi-Gen GT sequencing system and 21x50 cm gels (Bio-Rad) were used. The electrophoresis conditions were 50 W, 50ºC, and the time depended on the experiment. After electrophoresis, bands were visualized by autoradiography or using a Fujifilm Image Analyzer FLA-3000.

6.3. *In silico* prediction of intrinsic DNA curvature

The intrinsic curvature of the C and NC DNA fragments was predicted with the bend.it server (http://hydra.icgeb.trieste.it/dna/bend_it.html). It was calculated as degrees per helical turn (10.5º/helical turn = 1º/basepair). The curvature propensity plot was calculated using the consensus scale algorithm (DNase I + nucleosome positioning data) with a windows size of 20-bp.

7. DNA sequencing

7.1. Manual DNA sequencing

7.1.1. Dideoxy chain-termination sequencing method (Sanger method)

In general, the dideoxy chain-termination method was chosen for DNA sequencing (Sanger *et al*., 1977). The Sequenase v2.0 DNA Sequencing kit (USB Corporation) was used and some modifications were included when working with linear dsDNAs. For plasmid DNA sequencing, 1 µg of alkaline denatured plasmid was annealed to a 5´-labelled oligonucleotide (1.5 pmol). Specifically, the reaction mixture was incubated at 65ºC for 5 min and then cooled down slowly to 37ºC. Then, Sequenase DNA polymerase (6.5 units) was added to the mixture. In reactions with non-radiolabelled primers, [α-^{32}P]-dATP or [α-^{32}P]-dCTP (3,000 Ci/mmol; Hartmann) was added at the same time as the DNA polymerase. In both cases, and after 2-5 min at room temperature, the mixture was distributed into four tubes. Each tube contained all four dNTPs and one of the four ddNTPs (ddATP, ddCTP, ddGTP or ddTTP) pre-warmed at 37ºC. Mixtures were then incubated at 37ºC for 5 min. Reactions were stopped by adding the Stop Solution of the kit (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). For linear dsDNA sequencing, 2.5 pmol of a 5´-labelled primer were mixed with 0.25 pmol of dsDNA (obtained by PCR). The mixture was heated at 95ºC for 3 min and placed on ice. Then, the Sequenase DNA polymerase was added and the reactions proceeded as described above.

7.1.2. Maxam and Gilbert DNA sequencing method

When necessary, the Maxam and Gilbert DNA sequencing method for cleavage at purine residues (G+A) was used (Maxam and Gilbert, 1980). Briefly, 1 µg of tRNA and 1µl of 0.2 M formic acid were added to 2 µl of labelled dsDNA (30,000-50,000 cpm). The reaction mixture (4 µl) was incubated at 65ºC for 30 s. Then, DNA was precipitated by adding 30 µl of 1.5 M sodium acetate, pH 7.0, co-precipitant Pellet Paint™ and 150 µl of absolute ethanol. The mixture was kept at -80ºC for 10 min followed by centrifugation at

4ºC for 10 min (12,000 rpm in a Sigma 12124 rotor). The DNA pellet was dissolved in 30 µl of water. Then, 150 µl of absolute ethanol was added, and the mixture was incubated again at -80ºC for 10 min. After centrifugation at the same conditions, the DNA pellet was then washed with 80% ethanol. The DNA pellet was dissolved in 50 µl of water and 5 µl of 10 M Piperidine. The mixture was then heated at 100ºC for 10 min and the DNA was extracted with 1-butanol. Next, DNA was precipitated by mixing with 50 µl of 1% SDS and 500 µl of 1-butanol. Finally, DNA was washed twice with ethanol 80%, air-dried and dissolved in buffer BXF (see Table 5).

7.2. Automated DNA sequencing

All the genetic manipulations performed in chromosomal and extra-chromosomal (plasmid) DNAs were confirmed by dye-terminator sequencing at Secugen (Automated DNA Sequencing Service, CIB).

8. RNA techniques

8.1. Total RNA isolation from *E. coli* and *S. pneumoniae*

To isolate total RNA from either *E. coli* or *S. pneumoniae*, either the Aurum Total RNA Mini kit (Bio-Rad) or the RNeasy Mini kit (QIAGEN) were used as specified by the suppliers. In the case of pneumococcal RNA, cells were concentrated (20X) in LB buffer (see Table 5) and incubated at 37ºC for 5-10 min for cell lysis. In both cases, an additional DNase I digestion was performed to eliminate residual DNA in the RNA preparations. To this end, DNase I recombinant RNase free (Roche) was used and the samples were cleaned up by a second purification using the kit columns. Integrity of the rRNAs was checked by agarose (0.8%) gel electrophoresis. RNA concentration was determined using a NanoDrop ND-1000 Spectrophotometer (Bio-Rad).

8.2. *In vitro* transcription

Blunt-ended dsDNA fragments (obtained by PCR) were used as templates for *in vitro* transcription assays under multiple-round conditions. Two RNA polymerases (RNAPs) were used: *(i)* the commercially available *E. coli* RNAP holoenzyme (Epicentre), and *(ii)* a reconstituted RNAP composed by the *E. coli* RNAP core enzyme (Epicentre) and the pneumococcal σ^{43} factor purified in our laboratory. When the *E. coli* RNAP holoenzyme was used, reactions (50 µl) contained 10 nM of template DNA, 10 µl of 5x Transcription Buffer (see Table 5), 2 mM DTT, 100 µg/ml BSA (New England BioLabs), 10 units of SUPERase•In (RNase inhibitor, Ambion), 250 µM of each rNTP, 10

µCi of [α-^{32}P]-UTP (3,000 Ci/mmol, GE Healthcare) and 1 unit of RNAP holoenzyme. After 20 min at 30ºC, reactions were stopped by addition of 50 µl STOP-T buffer (see Table 5). Non-incorporated nucleotide was removed using MicroSpin^{TM} G-25 columns (GE Healthcare). Then, RNA was ethanol precipitated, dissolved in FXBE buffer, heated at 90ºC for 5 min and subjected to electrophoresis on sequencing gels (6% PAA, 8 M urea). When the reconstituted RNAP was used, reactions (33.5 µl) contained 1 unit of *E. coli* RNAP core enzyme, 250 ng of the pneumococcal σ^{43} factor, 2 mM DTT and 6.7 µl of 5x Transcription Buffer. The mixtures were incubated at 30ºC for 15 min (RNAP reconstitution). Then, the template DNA (10 nM) was added. After 15 min at the same temperature, transcription was initiated by addition of the rNTPs (including the [α-^{32}P]-UTP) and the RNase inhibitor. Reaction mixtures were incubated for 15 min at 30ºC, and then processed as described above. The sizes of the runoff transcripts were estimated by comparison with the sizes of DNA fragments generated by dideoxy-mediated chain termination sequencing reactions. In sequencing gels, RNA runs about 5 to 10% more slowly than DNA of the same size (Maniatis *et al*., 1982).

8.3. Primer extension

Primer extension reactions were carried out using total RNA isolated as described in Section 8.1. The ThermoScript Reverse Transcriptase kit (Invitrogen) was used as specified by the supplier. Specific primers were ^{32}P-labelled at the 5´-end as described in Section 5.7.1. In experiments with non-radiolabelled primers, [α-^{32}P]-dATP or [α-^{32}P]-dCTP was used in the extension reactions. Reaction mixtures (12 µl) contained 1-2 pmol of primer and 2-15 µg of total RNA. The mixture was incubated at 65ºC for 5 min (annealing). Then, 100 µM of each dNTP, 15 units of ThermoScript Reverse Transcriptase, 5 mM DTT and cDNA Synthesis buffer (supplied in the kit) were added. Extension reactions were carried out at 50-58ºC for 45-60 min (depending on the Tm of the primer and the length of the extension product). When non-radiolabelled primers were used, 9.75 µM of dATP or dCTP and 0.25 µM of [α-^{32}P]-dATP or [α-^{32}P]-dCTP, respectively, were added to the extension reactions. Reactions were stopped by heating at 85ºC for 5 min. Finally, samples were ethanol precipitated and dissolved in BXF buffer (see Table 5). cDNA products were analysed by sequencing gel electrophoresis. 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 or by autoradiography. When necessary, the intensity of the bands was quantified using the Quantity One software (Bio-Rad).

8.4. Reverse transcription polymerase chain reaction

RT-PCR assays were carried out using the ThermoScript Reverse Transcriptase kit for the synthesis of cDNA. Specifically, 20 pmol of primer were annealed to 1-2 µg of total RNA (12 µl) at 65ºC for 5 min. Then, the Thermoscript Reverse Transcriptase (15 units), 100 µM of each dNTP, 5 mM DTT and cDNA Synthesis buffer (provided by the supplier) were added. Reaction mixtures (20 µl) were incubated at 55ºC for 45 min. PCRs were then carried out as described in Section 5.3.2 using the cDNA as template (10% of the first-strand reaction). To ensure the absence of DNA in the RNA preparations, the same reactions were performed without adding reverse transcriptase (negative control). As positive control to guarantee the integrity of the primers, PCRs were done using chromosomal DNA as template. PCR products were analysed by agarose (0.8%) gel electrophoresis.

9. Protein purification

In all cases, the commercially available pET24b expression vector (Novagen) was used for overproduction and purification of proteins. In this vector, the gene of interest is expressed under control of the ϕ10 promoter of phage T7 using the *E. coli* BL21 (DE3) strain. This strain contains a chromosomal copy of the *lacI* gene and the T7 RNAP-encoding gene (T7 gene) under control of the *lac*UV5 promoter. In the absence of IPTG, the LacI repressor binds to the operator region of the *lac*UV5 promoter and represses the transcription of the T7 gene. In the presence of IPTG, LacI is blocked and, therefore, there is synthesis of the T7 RNAP, which transcribes the gene of interest from the ϕ10 promoter.

9.1. Purification of Mga*Spn*-His

E. coli BL21 (DE3) cells harbouring the pET24b-*mgaSpn*-His plasmid were grown at 37ºC with rotary shaking in TY broth containing Km (30 µg/ml) to an OD_{600} of 0.45. Expression of the *mgaSpn*-His gene was induced by addition of 1mM IPTG. After 25 min at 37ºC, rifampicin (200 µg/ml), which specifically inhibits bacterial RNAP, was added and the culture was incubated for 60 min. Cells were harvested by centrifugation (9,000 rpm in an SLA-3000 rotor for 20 min at 4ºC), washed twice with buffer V-His and stored at -80ºC. The cell pellet was concentrated (40X) in buffer V-His containing an EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by two passages through a pre-chilled cell-pressure French Press, and the whole-cell extract was clarified by centrifugation (9,000 rpm in an Eppendorf F-34-6-38 rotor for 60 min at 4ºC). Then,

imidazole (10 mM final concentration) was added to the clarified extract, which was subjected to a nickel affinity column (HisTrap HP column, GE Healthcare) pre-equilibrated with buffer V-His containing 10 mM imidazole. After washing with the same buffer, Mga*Spn*-His was eluted with buffer V-His containing 250 mM imidazole. Fractions containing Mga*Spn*-His were identified by Coomassie-stained SDS polyacrylamide (10%) gels, pooled, dialyzed at 4ºC against buffer P (see Table 5) and concentrated by filtering through a 10-kDa-cutoff membrane (Macrosep; Pall). The protein sample was then loaded onto a HiLoad Superdex 200 gel filtration column (Amersham Biosciences) and subjected to fast-pressure liquid chromatography (FPLC; Biologic DuoFlow; Bio-Rad). Fractions were analysed as described above, pooled, concentrated, and stored at -80ºC.

9.2. Purification of Mga*Spn*

E. coli BL21 (DE3) cells carrying the pET24b-*mgaSpn* plasmid were grown at 37ºC with rotary shaking in TY medium containing Km (30 µg/ml) to an OD_{600} of 0.45, followed by induction of *mgaSpn* gene expression with 1 mM of IPTG. After 25 min, rifampicin (200 µg/ml) was added and the culture was incubated for 60 min under the same conditions. Cells were harvested by centrifugation, and washed twice with buffer VL containing 400 mM NaCl. The cell pellet was concentrated (40X) in buffer VL containing 400 mM NaCl and a protease inhibitor cocktail (Roche). Cells were disrupted by two passages through a pre-chilled French pressure cell, and the whole-cell extract was centrifuged 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 4ºC. Under these conditions, Mga*Spn* was recovered in the pellet, which was then washed twice with buffer VL containing 400 mM NaCl to eliminate contaminant proteins. Mga*Spn* was eluted from the pellet with buffer VL containing 700 mM NaCl. Then, proteins in the supernatant were precipitated with 70% saturated ammonium sulphate, which was added slowly. After the addition of ammonium sulphate, the mixture was kept on ice with constant stirring for 60 min. After centrifugation (9,000 rpm in an Eppendorf F-34-6-38 rotor for 20 min at 4ºC), the precipitate was dissolved in buffer VL containing 400 mM NaCl and dialyzed at 4ºC against buffer VL containing 100 mM NaCl. The protein preparation was applied to a heparin affinity column (HiPrep Heparin, GE Healthcare) equilibrated with the same buffer. Bound protein was washed with buffer VL containing 300 mM NaCl and Mga*Spn* was subsequently eluted using a 300-800 mM NaCl gradient. Fractions containing Mga*Spn* were identified by Coomassie-stained SDS-polyacrylamide (10%) gels, pooled and dialyzed at 4ºC against VL buffer containing 100 mM NaCl. The protein preparation

was concentrated by filtering through a 3-kDa-cutoff membrane (Macrosep; Pall) and stored at -80ºC.

10. Protein analysis

10.1. Determination of protein concentration

For measuring the concentration of a protein in solution, a Nanodrop (ND-1000) Spectrophotometer (Bio-Rad) was used. 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. For circular dichroism experiments (see Section 10.6.2), the concentration of Mga*Spn* was calculated by measuring the absorption spectra at 220 and 350 nm in a Shimadzu UV-2401PC Spectrophotometer.

10.2. N-terminal sequencing

N-terminal sequencing of purified proteins was performed by Edman degradation using a Procise 494 Sequencer (Perkin Elmer) (Protein Chemistry Facility; CIB). When necessary, proteins were separated by SDS-PAGE. Pre-stained proteins (SeeBlue Plus 2; Invitrogen) were run in the same gel as molecular weight markers. Then, proteins were transferred electrophoretically to Immun-blot polyvinylidene difluoride (PVDF) membranes (Mini Trans-Blot; Bio-Rad) at 100 V and 4ºC for 90 min. Finally, membranes were stained with a fresh Coomassie solution, de-stained with 50% methanol in water, washed with water and air dried. Bands of interest were excised and used for the analysis.

10.3. Protein electrophoresis

10.3.1. Tris-Glycine SDS-PAGE

Protein samples were analysed by SDS-PAGE using TG buffer (see Table 5) as running buffer and a Mini-Protean III Electrophoresis System (Bio-Rad). The stacking gel contained 4% polyacrylamide in 0.125 M Tris-HCl, pH 6.8, and 0.1% SDS. The resolving or separating gel contained 10-12% polyacrylamide in 0.374 M Tris-HCl, pH 8.8, and 0.1% SDS. TEMED and ammonium persulfate (PSA) were used for polyacrylamide gel polymerization. Protein samples were mixed with SLB buffer (see Table 5) and heated at 95ºC for 5 min prior to electrophoresis. Electrophoresis was performed 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.

10.3.2. Tris-Tricine SDS-PAGE

For the resolution of proteins smaller than 30 kDa, Tris-Tricine SDS-PAGE (Schagger and von Jagow, 1987) was commonly used. Gels were prepared in a Mini-Protean III Electrophoresis System (Bio-Rad). The staking gel (4 ml) contained 4% polyacrylamide, and 1 ml Buffer gel (see Table 5). The resolving gel (5 ml) contained 16% polyacrylamide, 1.65 ml Buffer gel and glycerol (13%). TEMED and PSA were used to catalyse polyacrylamide gel polymerization. In this electrophoretic system, cathode buffer and anode buffer were used (see Table 5). Electrophoresis conditions were as described in Section 10.3.1.

10.4. Analytical ultracentrifugation

Analytical ultracentrifugation experiments were carried out at the CIB-Analytical Ultracentrifugation and Macromolecular Interactions Facility. These experiments provided information about the molecular mass of Mga*Spn* as well as the hydrodynamic behaviour of the protein. All the experiments were done in an Optima XL-A (Beckman-Coulter) analytical ultracentrifuge equipped with an UV-visible optical detection system, using an An60Ti rotor with standard six-channel centrifuge cells (12-mm optical path) and centrepieces of epon charcoal.

10.4.1. Sedimentation velocity

Sedimentation velocity experiments were performed at 43,000 rpm and 12ºC. Mga*Spn* was equilibrated in buffer AU (see Table 5) and two protein concentrations (5 and 10 µM; 400 µl) were analysed. The sedimentation coefficient for Mga*Spn* was estimated using the program SEDFIT (version 11.8) (Schuck and Rossmanith, 2000) applying a direct linear least-squares boundary modelling of the sedimentation velocity data. The sedimentation coefficient was corrected to standard conditions using the program SEDNTERP (Laue *et al*., 1992) to obtain the corresponding $S_{20,w}$ value. The translational frictional coefficient (*f*) of Mga*Spn* was determined from the molecular mass and sedimentation coefficient of the protein (van Holde, 1985), whereas the frictional coefficient of the equivalent hydrated sphere (f_0) was estimated using a hydration of 0.38 g H_2O/g protein (Pessen and Kumosinski, 1985). With these parameters the translational frictional ratio (f/f_0) was calculated, which allows an estimation of the hydrodynamic shape of Mga*Spn.*

10.4.2. Sedimentation equilibrium

Sedimentation equilibrium experiments were performed at 10ºC. Two protein concentrations (5 and 10 µM; 80 µl) were analysed. The samples were centrifuged at three successive speeds (8,000, 9,500 and 12,500 rpm) and absorbance readings were done after the sedimentation equilibrium was reached. The absorbance scans were taken at 250, 280 and 288 nm, depending on the Mga*Spn* protein concentration used. In all cases, the baseline signals were measured after high-speed centrifugation (40,000 rpm). Apparent average molecular masses of Mga*Spn* protein were determined using the program HETEROANALYSIS (Cole, 2004). The partial specific volume of Mga*Spn* was 0.742 ml/g, determined from the amino acid composition with the program SEDNTERP (Laue *et al*., 1992).

10.5. Gel filtration chromatography

Gel filtration chromatography was used to determine the molecular size (Stokes radius) of Mga*Spn*, and was carried out in an Åkta HPLC system (Amersham Biosciences) using a HiLoad Superdex 200 gel-filtration column (120 ml; 16 x 600 mm) (Amersham Biosciences), equilibrated with buffer VL containing either 100 mM or 250 mM NaCl. All chromatographic runs were performed at 4ºC with a flow rate of 0.5 ml/min. Elution positions of the proteins were monitored at 280 nm. The column was pre-calibrated by loading a set of standard proteins of known Stokes radius (molecular size), as the behaviour of non-globular proteins during gel filtration correlates with the molecular size more than with the molecular weight. The standards used were ferritin (F; 61 Å), alcohol dehydrogenase (ADH; 45 Å), ovalbumin (O; 30.5 Å) and carbonic anhydrase (CA; 20.1 Å). Proteins were diluted in buffer VL containing 250 mM NaCl to final concentrations of 0.3 mg/ml (F), 5 mg/ml (ADH), 4 mg/ml (O) and 3 mg/ml (CA). Purified Mga*Spn* protein (1 ml, 25 µM) equilibrated in the same buffer was loaded onto the column and eluted with a constant flow rate of 0.5 ml/min. Fractions of 2 ml were collected, aliquots of the peak fractions were analysed in coomassie-stained SDS-polyacrylamide (10%) gels. The K_{av} parameter was calculated for each protein as follows:

$$K_{av} = (V_e - V_0)/(V_t - V_0)$$

where V_e is the elution volume, V_0 is de void volume (determined by elution of blue dextran) and V_t is the total volume of the packed bed.

10.6. Protein secondary structure analysis

10.6.1. Protein secondary structure prediction

The prediction of the content and distribution of secondary structures derived from the amino acid sequence of the Mga*Spn* protein was carried out with the following bioinformatics programs: SABLE (Adamczak *et al*., 2005), PsiPred (McGuffin *et al*., 2000), Jpred (Cole *et al*., 2008), NPS@ (Combet *et al*., 2000) and PredictProtein (Rost *et al*., 2004), that are accessible through the ExPASy Bioinformatics Resource Portal (proteomics tools) (Artimo *et al*., 2012; www.expasy.org). Search of functional domains in Mga*Spn* was done with the Pfam protein families database (Punta *et al*., 2012) and with the Phyre2 protein fold recognition server (Kelley and Sternberg, 2009).

10.6.2. Circular dichroism analyses

Circular dichroism (CD) spectra of Mga*Spn* were acquired in a Jasco J-810 spectropolarimeter employing a 0.2-mm path length rectangular quartz cuvette. Samples were dialyzed against CD buffer (see Table 5) and diluted in the same buffer to a protein concentration of 0.71 mg/ml. Spectra were recorded over a wavelength range from 185 to 260 nm (far-UV) at 4ºC, with a resolution of 1 nm at a scan speed of 50 nm/min. The CD spectrum was the average of four accumulations or scans. The final spectrum was obtained by subtracting the buffer spectrum measured under identical conditions. The results were expressed as mean residue ellipticity (Θ) at a given wavelength. To obtain structural information, the CD data were analysed using different algorithms: SELCON3 (the spectrum of the protein to be analysed is included in the basis set. The resulting matrix is solved and the initial guess is replaced by the solution. The process is repeated until self-consistency is attained. SELCON3 provides good estimates of the structure of globular proteins, however, it gives poor estimates of turns; Sreerama and Woody, 1993), CONTINLL (compares and fits the CD spectrum of the protein under study with the spectra of a database of proteins with known conformations. This method results in good estimates of helices and sheets; Provencher and Glockner, 1981), CDSSTR (in this method an initial database of standard spectra form proteins with known spectra and secondary structure is created and provides superior fits of the conformation of globular proteins; Compton and Johnson, 1986); and K2D (an artificial intelligence program is used to fin correlations in data. KD2 gives a good estimate of the helical and sheer contents, however does not estimate turns; Andrade *et al*., 1993).

10.6.3. Thermal Stability

CD analyses are also useful as a method to determine the thermal stability of a protein, since changes in the ellipticity as a result of protein denaturation (disruption of secondary structure) can be detected. Thus, temperature-induced changes in the Mga*Spn* protein secondary structure were measured by increasing the temperature from 4ºC to 90ºC at two different rates (20 and 50ºC/h). Samples were dialyzed against CD buffer and diluted in the same buffer to a protein concentration of 0.35 mg/ml. Changes in ellipticity were recorded at 220 nm in a 1-mm optical path length quartz cell. Additionally, CD spectra over a wavelength range from 185 to 260 nm were recorded at different temperatures (4, 15, 25, 37, 50, 60, 70, 90ºC). The temperature was equilibrated for 1 min prior to the acquisition of each spectrum. Finally, the sample was cooled to the initial temperature and a spectrum over the same wavelength range was recorded.

10.7. Western blots

S. pneumoniae R6 cells were grown to an OD_{650} of 0.3 and cells from 3 ml of culture were sedimented by centrifugation. Cell pellets were resuspended in 100 µl of LB buffer (see Table 5) and incubated at 30ºC for 5-10 min. Then, 25 µl of 5x SLB buffer were added, and the samples were heated at 95ºC for 5 min before being loaded onto SDS-polyacrylamide (10%) gels. The SeeBlue Plus 2 pre-stained protein standard (Invitrogen) was loaded in the same gel as a molecular weight marker. After electrophoresis, the gel was equilibrated with TB buffer. Immun-Blot PVDF membranes (Bio-Rad) were used for protein blotting. Methanol was used to pre-wet the membrane prior to equilibration in TB buffer. Proteins were transferred electrophoretically to the membrane using a Mini Trans-Blot (Bio-Rad) and TB buffer. The transfer was carried out at 100 mA at 4ºC for 90 min. Then, the membrane was dried by immersion in methanol for 2 min. Membranes were probed with polyclonal antibodies against His-tagged Mga*Spn*, which were diluted 1:1,000 in SB buffer. After 1 h of incubation, the membrane was rinsed three times with WB buffer and incubated for 1 h with the secondary antibody (anti-rabbit IgG) conjugated with horseradish peroxidase (HRP). The membrane was then washed three times with PBS. Antigen-antibody complexes were detected using the Immun-Star™ HRP substrate kit (Bio-Rad) in which the peroxidase conjugated with the secondary antibody catalyses oxidation of luminol (the substrate) and subsequently enhances chemiluminescence when the luminol recovers its original estate (reduction). The resulting light was detected with a Luminescent Image Analyzer LAS-3000 (Fujifilm

Life Science) or by autoradiography. The intensity of the bands was quantified using the Quantity One software (Bio-Rad).

10.8. Proteomics

Pneumococcal cells were grown in AGCH medium supplemented with sucrose (0.3%), yeast extract (0.2%) and antibiotic (if required) to an OD_{650} of 0.4. Then, cells were sedimented by centrifugation at 8,000 rpm in an Eppendorf F-34-6-38 rotor for 25 min at 4ºC. Cells were washed twice with pre-cooled PBS buffer and the cell pellet was stored at -80ºC. The cell pellet was concentrated (30x) in LBP buffer (Table 5). Bacteria were disrupted by two passages through a pre-chilled French pressure cell. The whole-cell extract was immediately frozen in dry ice and stored at -80ºC. Separation and identification of proteins was carried out at the Proteomics Core Facility of the CNIC using Difference Gel Electrophoresis (DIGE) (pre-labelling of samples with differential fluorochromes and 2-dimensional gel electrophoresis) a MALDI-TOF/TOF and ESI Ion-Trap spectrometers.

11. DNA-protein interactions

11.1. Electrophoretic mobility shift assays

11.1.1. Standard conditions for EMSA

In general, standard binding reactions were performed in a volume of 10-20 µl containing 40 mM Tris-HCl, pH 7.6, 1.4 mM DTT, 0.4 mM EDTA, 2% glycerol, 50 mM NaCl, 10 mM $MgCl_2$, 500 µg/ml BSA, 10 nM of unlabelled DNA or 0.1-2 nM of 5´-labelled DNA and varying concentrations of purified Mga*Spn* (1-480 nM). After 20 min of incubation at room temperature, BXGE buffer was added and free and bound DNA forms were separated on native polyacrylamide (5 or 8%, depending on the length of the DNA used) gels at 100 V and room temperature using TBE buffer. Labelled DNA bands were visualized using a Fujifilm Image Analyzer FLA-3000 and the Quantity One software. DNA fragments used in EMSA experiments are listed in Table 4.

When indicated, non-labelled competitor DNA was added to the binding reaction: calf thymus DNA or linear dsDNA fragments (containing or not intrinsic curvature). In general, ^{32}P-labelled DNA and non-labelled competitor DNA were added simultaneously to the reaction mixture. In dissociation experiments, Mga*Spn* was incubated with the ^{32}P-labelled DNA under standard conditions (formation of protein-DNA complexes). Then,

different amounts of the non-labelled competitor DNA were added and the reaction mixtures were incubated for 5 min (dissociation of protein-DNA complexes).

11.1.2. Determination of the apparent dissociation constant (K_d)

To estimate the affinity of Mga*Spn* for different DNA fragments, EMSA experiments were performed in which the Mga*Spn* concentration was varied from 1 to 150 nM. The concentration of the labelled DNA fragment was 0.1 nM. Free DNA was visualized using a Fujifilm Image Analyzer FLA-3000 and was quantified using the Quantity One software (Bio-Rad). The protein concentration required to bind half the DNA was determined by measuring the decrease in free DNA rather than the increase in complexes (Carey, 1991), which gives an indication of the approximate magnitude of the dissociation constant, K_d. The data were plotted as fraction of free DNA versus protein concentration. The experiments were repeated twice.

11.2. DNase I footprinting assays

The DNase I footprinting assay is based on the enzymatic digestion of the DNA by DNase I in the presence of a DNA binding protein (Galas and Schmitz, 1978). Prior to the footprinting assay, DNase I titration was done to determine the optimum DNase I concentration to be used in the experiments. DNase I (stock solution) was diluted in 50% glycerol to 1:100, 1:250, 1:500 and 1:1000. Reactions (50 µl) contained 2 nM of the DNA fragment, 30 mM Tris-HCl, pH 7.6, 1% glycerol, 1.2 mM DTT, 0.2 mM EDTA, 50 mM NaCl, 1 mM $CaCl_2$, 10 mM $MgCl_2$. The digestion reaction was initiated by addition of 2 µl of the corresponding DNase I dilution. After incubation for 5 min at room temperature, reactions were stopped by adding 25 µl of STOP DNase buffer and 187 µl of absolute ethanol. DNA was precipitated, dissolved in BXF buffer, heated at 95ºC for 5 min and loaded onto sequencing gels (6% PAA, 8 M urea). A good footprint is obtained when approximately 50% of the DNA fragment remains intact. The optimal dilution of the DNase I stock was used in the DNase I footprinting assays.

Binding reactions were performed in a volume of 10-50 µl containing 2-4 nM of 5´-end labelled DNA, 30 mM Tris-HCl, pH 7.6, 1% glycerol, 1.2 mM DTT, 0.2 mM EDTA, 50 mM NaCl, 1 mM $CaCl_2$, 10 mM $MgCl_2$, 500 µg/ml BSA and different concentrations of Mga*Spn*. When necessary, heparin (2 µg/ml) was added to the binding reactions. Reaction mixtures were incubated for 20 min at room temperature. Then, 0.04 units of DNase I was added and the incubation proceeded for 5 min at the same temperature. In reactions of 50 µl, DNase I digestion was stopped by adding 25 µl of STOP DNase

solution. DNA was precipitated with ethanol, dried and dissolved in 5 µl of BXF buffer. In reactions of 10 µl, DNase I digestion was stopped by adding 1 µl of 250 mM EDTA. Then, 4 µl of BXF buffer was added. After heating at 95°C for 5 min, samples were loaded on sequencing gels (6% PAA, 8 M urea).

11.3. Hydroxyl radical (OH•) footprinting assays

In the hydroxyl radical (OH•) footprinting technique, the DNA is chemically cut for the action of hydroxyl radicals generated by reduction of hydrogen peroxide by iron (II) (Tullius and Dombroski, 1986). To generate the hydroxyl radicals, equal volumes of 6% H_2O_2, 20 mM sodium ascorbate and a Fe^{2+}-EDTA solution (equal volumes of 4 mM ammonium iron (II) sulphate hexahydrate freshly prepared and 8 mM EDTA) were mixed immediately before being used ("Three reagents"). First, ^{32}P-labelled DNA (4-8 nM) was incubated with Mga*Spn* (320-640 nM) in 50 µl of buffer containing 30 mM Tris-HCl, pH 7.6, 50 mM NaCl, 10 mM $MgCl_2$, 1.2 mM DTT, 0.2 mM EDTA, 0.02% glycerol, 500 µg/ml BSA, and incubated at room temperature for 20 min. Then, 20 ng of calf thymus DNA was added to enrich the sample in complex C1, and the mixture was incubated for 5 min at the same temperature. The cleavage reaction was started by adding 9 µl of the "Three reagents". The OH• cleavage was allowed to proceed for 5 min at room temperature. Reactions were stopped by the addition of thiourea (9.5 mM). To increase the footprint signal, protein-DNA complexes were separated from free DNA by native polyacrylamide (5%) gel electrophoresis and visualized by autoradiography. Free and bound DNAs were excised from gels and eluted at 42ºC overnight in EB buffer (see Table 5). Finally, DNA was precipitated with ethanol, air-dried and dissolved in 5µl of BXF buffer. Samples were heated at 95°C for 5 min before being loaded onto sequencing gels (6% PAA, 8 M urea).

11.4. Analysis of protein-DNA complexes by electron microscopy

These experiments were carried out using different linear dsDNA fragments (see Table 4). First, binding reactions (40-50 µl) were done in buffer containing 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.2 mM EDTA, 1% glycerol, 50-100 mM NaCl, 10 mM $MgCl_2$, 0.5-1 nM of DNA and Mga*Spn* (2-40 nM). Reactions were incubated at room temperature for 20 min. Then, complexes were adsorbed onto freshly cleaved mica, positively stained with 2% uranyl acetate and rotary shadowed. This is the main contrasting method for nucleic acids and it is produced by evaporation of a metal (Pt/Ir) under high vacuum conditions while the sample is rotating. Finally, the mica sheets were covered with a carbon film, which provides homogeneous low noise films of high stability under electron bombardment (Spiess and Lurz, 1988). Samples were visualized in an

electron microscope (Philips CM100, 100 kV; FEI Company, Hillsboro, Oregon) and micrographs of the carbon film replica were taken with a coupled Fastscan CCD camera. To determine the protein binding position on a given DNA, the contour length of the DNA in the protein-DNA complexes as well as the distance from one end of the DNA to the protein-binding site, was measured on projections of 35-mm negatives using a digitizer LM4 (Brühl, Nüremberg, Germany).

Results

Chapter 1

Expression of the pneumococcal *mgaSpn* gene

As a consequence of living in habitats of changing conditions, bacteria have developed several mechanisms to regulate gene expression. Although genetic regulation occurs at different levels, one of the most important mechanisms of adaptation to environmental changes is the regulation at a transcriptional level. Bacterial transcription is carried out by a multisubunit RNA polymerase (RNAP) holoenzyme. This enzyme is a complex of five subunits that constitute the core enzyme ($\alpha_2\beta\beta'\omega$) and a σ factor, which confers promoter specificity to the RNAP, interacts with transcription activators, participates in promoter DNA opening and influences the early phases of transcription (Saecker *et al*., 2011). In general, bacterial genomes encode different species of σ factors, each of which directs transcription of specific sets of genes (Gruber and Gross, 2003). In exponentially growing bacterial cells, most of the genes expressed are transcribed by an RNAP containing a housekeeping σ factor similar to the *Escherichia coli* σ^{70}. Promoters that are recognised by this holoenzyme are characterized by two main sequence elements: the -35 (5´-**TTGACA**-3´) and -10 (5´-**TATAAT**-3´) hexamers (relative to the transcription initiation site). The optimum spacer length between them is 17 nucleotides (nt). In addition, some of these promoters also contain the extended -10 element, which is more conserved in G+ bacteria (5´-**TRTGN**-3´) than in *E. coli* (5´-TGN-3´), and is located immediately upstream of the -10 element (Figure 5) (Mitchell *et al*., 2003; Voskuil and Chambliss, 1998). The presence of an extension within the -10 region is generally associated with strong promoters that may lack the -35 element both in *S. pneumoniae* (Sabelnikov *et al*., 1995) and *E. coli* (deHaseth *et al*., 1998). It has been shown that sequence-specific interactions between σ^{70} and the promoter, occur at the -10 element, the extended -10 element and the -35 element (reviewed by Haugen *et al*., 2008). Another element recognised by the RNAP is the UP element. It is located upstream of the -35 hexamer and is contacted by the α subunit of the RNAP (Ross *et al*., 1993). UP elements have been found upstream of many bacterial promoters and can be recognised not only by the RNAP σ^{70} holoenzyme, but also by RNAPs containing alternative σ factors (Fredrick *et al*., 1995). This interaction facilitates the initial binding of the RNAP and enhances transcription (Aiyar *et al*., 1998). The UP element consists of two distinct subsites: the proximal subsite (positions -46 to -38, consensus 5´-AAAAAARNR-3´) and the distal subsite (positions -57 to -47, consensus 5´-AWWWWWTTTTT-3´ (Estrem *et al*., 1999). Promoters that contain a single near-consensus region (either proximal or distal) are more common than promoters with a near-consensus full UP element, and the effect on transcription of these upstream sequences generally correlates with the similarity to the consensus (Ross *et al*., 1998).

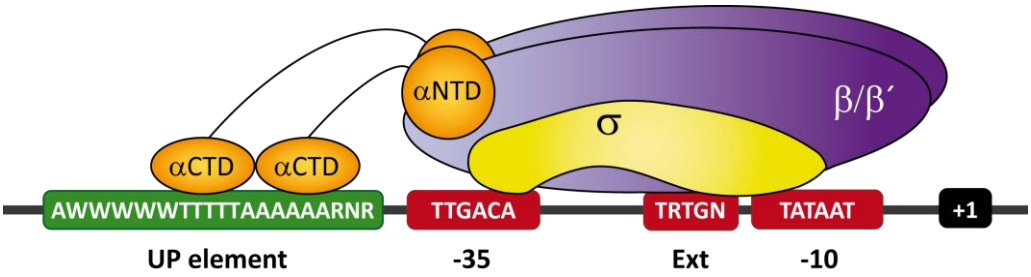


Figure 5. Architecture of bacterial promoters. Schematic representation of the major RNAP holoenzyme and its interaction with the different promoter elements. Consensus sequences of the elements recognized by the σ^{70} and α (αCTD: carboxy-terminal domain; αNTD: amino-terminal domain) subunits are indicated. The optimal UP element (α binding site) consists of alternating A- and T- tracts (-57 to -38; 5´- AWWWWWTTTTTAAAAAARNR-3´; W = A or T; R = A or G; N = any base) (Estrem *et al.*, 1999). Ext refers to the extended -10 element, whose consensus sequence in G+ bacteria is indicated (R = A or G; N = any base). The -35 element, the extended -10 and the -10 hexamer are contacted by the σ^{70} subunit. +1 (black box) is the transcription start site.

Although bioinformatics can predict some promoters correctly, definitive identification of promoters requires the use of several experimental approaches, both *in vivo* and *in vitro* (Ross and Gourse, 2009). In this chapter, we show that the pneumococcal *mgaSpn* regulatory gene is transcribed from the *Pmga* promoter. Moreover, we demonstrate that this promoter is recognized by a reconstituted RNAP holoenzyme containing the *E. coli* core enzyme and the *S. pneumoniae* σ^{43} subunit (homologue of the *E. coli* σ^{70}).

1.1. The *mgaSpn* gene is transcribed from the *Pmga* promoter

In the R6 pneumococcal strain, whose complete genome sequence was published in 2001 (Hoskins *et al.*, 2001) (GenBank AE007317.1), the *spr1622* gene (named *mgaSpn* in this work) encodes an Mga-like regulator. Such a gene is flanked by the *spr1623* gene (unknown function) and the *scrR* gene (*spr1621*), which encodes the sucrose utilization system repressor (Figure 6). The four genes located downstream of the *scrR* gene encode an ABC transporter involved in the uptake and metabolism of sucrose (genes *spr1620*, *spr1619*, *spr1618*) and a sucrose-6-phospate hydrolase (*spr1617* or *sacA*), respectively. According to the National Center for Biotechnology Information (NCBI) Entrez Genome Database, translation of the *mgaSpn* gene starts at coordinate 1598327. However, we think that the ATG codon at coordinate 1598270 is likely the translation start site of the *mgaSpn* gene because it is preceded by a putative Shine-Dalgarno sequence (5´-AAAGAGAGAAAG-3´) (Figure 6) that matches the reported consensus sequence for pneumococcus (Chang Bioscience, San Francisco, CA). Translation from this ATG codon would result in a protein of 493 residues (Mga*Spn*).

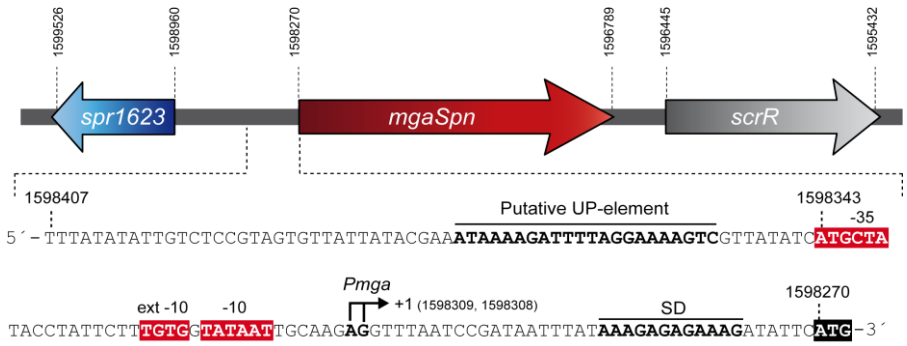


Figure 6. Scheme of the pneumococcal R6 genome spanning coordinates 1595432 to 1599526 (Hoskins *et al.*, 2001). The *spr1622* gene has been named *mgaSpn* in this work. For each open reading frame (ORF), the coordinates of the start and stop codons are indicated. The translation start site (ATG codon) of the *mgaSpn* gene is marked (black box). The putative Shine-Dalgarno sequence (SD) of the *mgaSpn* gene is indicated (bold letters). The main sequence elements of the *Pmga* promoter (-35 hexamer, -10 box, extended –10 element and putative UP-element) as well as the transcription initiation site (+1) are shown.

To study whether the *mgaSpn* gene was transcribed under our bacterial growth conditions, RT-PCR experiments were carried out (Figure 7). Specifically, the 1622A oligonucleotide, which anneals to an internal region of the *mgaSpn* gene, was used as primer for extension on total RNA isolated from R6 cells. It generated a specific cDNA extension product, which was further amplified by PCR using either the 1622A and

C1622D or the 1622A and 1622C primers (Table 3). As controls, PCR reactions were performed using total RNA (negative control) or genomic DNA (positive control) as template. The resulting PCR products were analysed by agarose gel electrophoresis. When primers 1622A and C1622D were used, a PCR product that migrated at the position expected for a 1,023-bp DNA was amplified. This product was also detected in the positive control but not in the negative control. With the 1622A and 1622C primers no PCR products were detected, whereas a PCR product with the mobility expected for a 1,221-bp fragment was synthesized in the positive control. These results indicated that the transcription of the *mgaSpn* gene was initiated at a site located downstream of coordinate 1598452. We then proceeded to analyse the region located between coordinate 1598452 and the translation start codon (coordinate 1598270) of the *mgaSpn* gene using the BPROM promoter recognition program (www.softberry.com). This sequence analysis led to the identification of a putative promoter (herein named *Pmga*) (Figure 6), which contains a consensus -10 hexamer (5´-**TATAAT**-3´), a consensus –10 extension (5´-**TGTG**-3´) and shows a 3/6 match at the –35 hexamer (5´-a**TG**ct**A**-3´). In addition, the distance between the -35 and the -10 elements is 16 nt, and the -10 element is located 45 nt upstream of the translation start codon. These features indicated that the *Pmga* promoter might be recognised by a housekeeping σ factor similar to the *E. coli* σ^{70} (Haugen *et al*., 2008). Moreover, according to the consensus sequence reported for UP elements (Estrem *et al*., 1999), the presence of A+T rich regions (positions -43 to -64) upstream of the -35 element suggests that the *Pmga* promoter might contain an UP element.

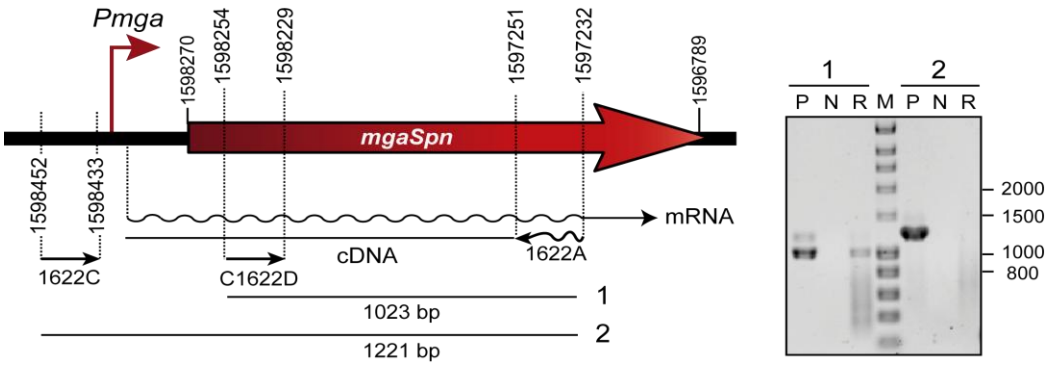


Figure 7. Transcription of the *mgaSpn* gene *in vivo*. RT-PCR assays were performed using total RNA isolated from *S. pneumoniae* R6 strain. The positions of the oligonucleotides used (1622A, 1622C, C1622D) as well as the expected DNA fragments (1 and 2) are shown. RT-PCR reaction products (lane R) were subjected to agarose (0.8%) gel electrophoresis and visualized using ethidium bromide. RT-PCR reactions without reverse transcriptase were performed as negative control (lane N). The size of PCR-amplified DNA fragments (fragments 1 and 2) using genomic DNA as template (lane P, positive control) is indicated. The size (in bp) of DNA fragments (lane M) used as molecular weight markers (HyperLadder I, Bioline) are indicated on the right of the gel.

To identify the transcription initiation site of the *mgaSpn* gene, we carried out primer extension experiments. For these experiments we used total RNA isolated from pneumococcal R6 cells and the 1622D primer, which is complementary to the C1622D oligonucleotide (see Table 3, Figure 7). Nevertheless, we did not detect any cDNA extension product (not shown), which suggested that the amount of *mgaSpn* transcripts in the total RNA preparation was too small. As an alternative to amplify the signal, we cloned a 136-bp fragment (coordinates 1598440 to 1598305 of the R6 genome), which contained the *Pmga* promoter sequence, into the *Bam*HI site of the pAS vector (pAS-*Pmga* recombinant plasmid; Figure 8). Vector pAS (constructed in our laboratory; Ruiz-Cruz *et al*., 2010) has a multiple cloning site (including *Bam*HI) upstream of a promoter-less *gfp* allele, which encodes a green fluorescence protein (GFP) that carries both the F64L mutation (it increases GFP solubility) and the S65T mutation (it increases GFP fluorescence and causes a red shift in the excitation spectrum) (Cormack *et al*., 1996; Heim, 1995). Moreover, the *gfp* allele carries translation initiation signals that are optimal for its expression in prokaryotes (Miller and Lindow, 1997). Therefore, the promoter activity of the 136-bp DNA fragment was evaluated by monitoring *gfp* expression. The intensity of fluorescence (measured at 515 nm; excitation at 488 nm) in cells harbouring the pAS-*Pmga* recombinant plasmid was slightly higher (1.5 fold) than in cells carrying the pAS vector, indicating that the 136-bp region contained a promoter signal. Subsequently, we performed primer extension experiments using total RNA from cells harbouring the pAS-*Pmga* recombinant plasmid. In this case, the INTgfp primer, which anneals to the *gfp* transcript, was used for the extension reactions (Figure 8). Two cDNA extension products of 120 and 121 nt were detected, indicating that transcription of the *gfp* gene started at a site located 7 to 8 nt downstream of the –10 element of the *Pmga* promoter. These results demonstrated that the *Pmga* promoter was functional *in vivo* under our experimental conditions.

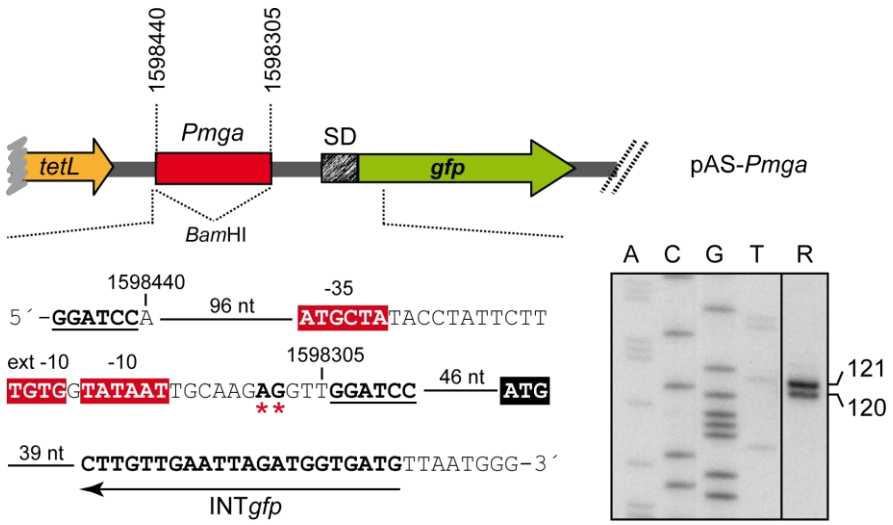


Figure 8. The *Pmga* promoter is functional *in vivo*. Primer extension reactions were carried out on total RNA isolated from pneumococcal cells carrying plasmid pAS-*Pmga*. The *gfp* gene carries translation initiation signals optimized for prokaryotes (SD) (Miller and Lindow, 1997). The *tetL* gene confers resistance to tetracycline. The main sequence elements of the *mgaSpn* gene promoter (red boxes) and the ATG initiation codon of the *gfp* gene (black box) are indicated. *Bam*HI sites are underlined. The asterisks indicate the 3′ -ends of the cDNA products synthesized in the reaction using the INTgfp primer. The sizes of the cDNA products (lane R) are indicated in nucleotides on the right of the gel. A, C, G, T sequence ladders were used as DNA size markers using DNA from pLS1 plasmid (Lacks *et al*., 1986) and the F-pLSI primer (5´-TGCTGGCAGGCACTGGC-3´; coordinates 802 to 818).

1.2. The *Pmga* promoter is recognized by the pneumococcal σ^{43} factor

Unlike the *E. coli* RNAP holoenzyme (containing the σ^{70} subunit), the *S. pneumoniae* RNAP holoenzyme is not commercially available. As a first approach to characterize the *Pmga* promoter, we investigated whether the *E. coli* RNAP holoenzyme was able to initiate transcription from the *Pmga* promoter. To this end, *in vitro* transcription experiments under multiple-round conditions were performed (Figure 9).

Two linear DNA fragments of 224-bp (coordinates 1598452 to 1598229) and 265-bp (coordinates 1598452 to 1598188) were used as templates (see Figure 9A). Transcription from the *Pmga* promoter should generate runoff transcripts of 81 nt or 122 nt using the 224-bp or 265-bp DNA, respectively. The *in vitro* transcription products were resolved on denaturing gels (6% PAA, 8M urea) (Figure 9B), and their sizes were calculated taking into account that an RNA molecule migrates about 5-10% more slowly than a DNA molecule of the same size in denaturing gels (Sambrook *et al*., 1989). When the 224-bp DNA fragment was used as a template (lane 1), transcripts of 80-81nt and 140-141 nt were detected. In the case of the 265-bp DNA fragment (lane 2), transcripts

of 121-122 nt and 182 nt were observed. These results indicated that the *E. coli* RNAP holoenzyme was able to initiate transcription not only at coordinate 1598309 (promoter *Pmga*) but also at coordinate 1598369 (Figure 9A). According to the BPROM prediction program there is a promoter sequence (here named *P2*) just upstream of coordinate 1598369. Such a promoter has a near consensus -10 element (5´-**TAT**t**AT**-3´), a near consensus -10 extension (5´-a**GTG**-3´), and a 3/6 match at the –35 hexamer (5´-**TTG**ttt-3´). Moreover, the -10 and -35 sequence elements are separated by 20 nt.

The pneumococcal σ^{43} factor (gene *rpoD*), which is homologous to the *E. coli* σ^{70} factor, has been purified in our laboratory by Sofía Ruiz-Cruz. We used the purified σ^{43} factor to reconstitute a functional RNAP holoenzyme with the commercial *E. coli* core enzyme, as described in Methods (section 8.2). The reconstituted RNAP was further used for *in vitro* transcription experiments using again the 224-bp and 265-bp DNAs as templates (Figure 9A). As shown in Figure 9C, the reconstituted RNAP was able to initiate transcription from the *Pmga* promoter but not from the *P2* promoter. Therefore, we conclude that the *Pmga* promoter is recognized by the pneumococcal housekeeping σ^{43} factor.

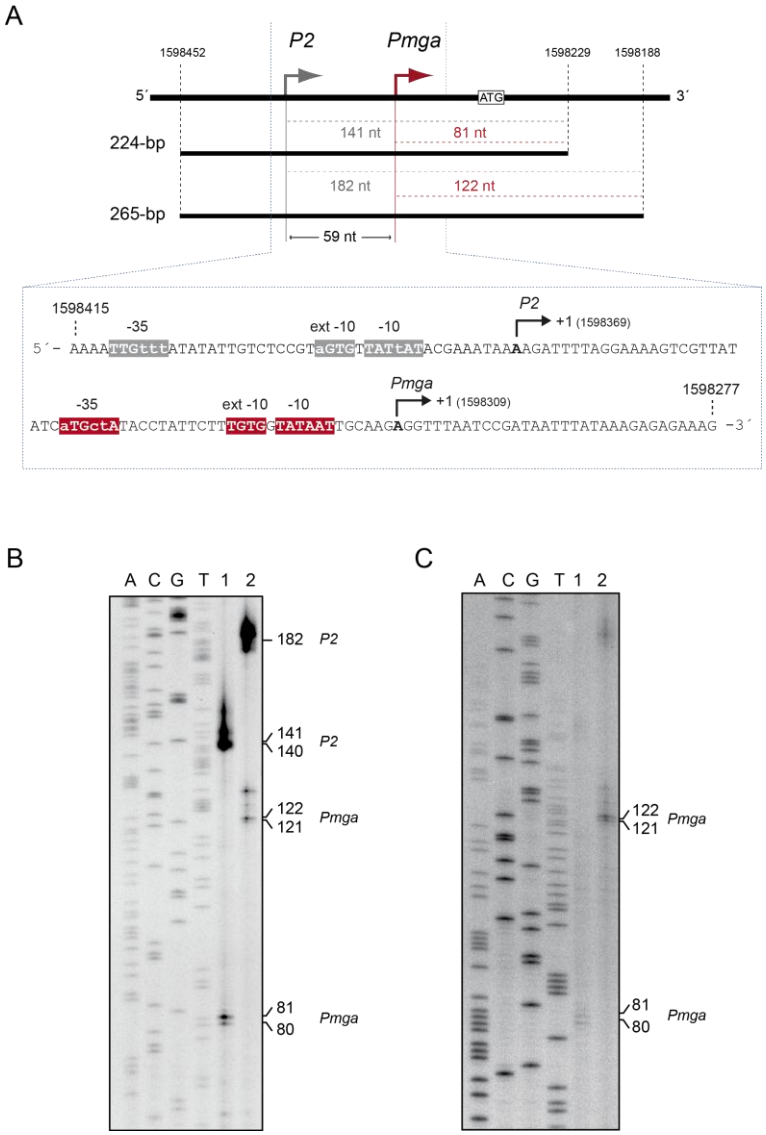


Figure 9. The *Pmga* promoter is recognized by the σ^{43} factor. (**A**) Scheme showing the 224-bp (coordinates 1598452 to 1598229) and 265-bp (coordinates 1598452 to 1598188) linear DNA fragments used as templates for *in vitro* transcription assays. In red, runoff transcripts expected to be synthesized from the *Pmga* promoter using the 224-bp (81 nt) or the 265-bp (122 nt) DNAs. In grey, runoff transcripts generated from the *P2* promoter sequence. The *Pmga* and *P2* promoters are separated by 59 nt. The translation start codon (ATG) of the *mgaSpn* gene is indicated. Nucleotide sequence spanning coordinates 1598415 to 1598277 is shown. The main sequence elements (-35 element, -10 box and extended -10 element) of the *Pmga* (red boxes) and *P2* (grey boxes) promoters as well as the transcription initiation sites (+1) identified in this work are shown. (**B**) *In vitro* transcription experiments using the *E. coli* RNAP holoenzyme (Epicentre). Reactions using the 224-bp (Lane 1) or the 265-bp (Lane 2) DNA were loaded onto a sequencing gel (6% PAA, 8M urea). The sizes (in nucleotides) of the runoff transcripts are indicated on the right of the gel. Sequence ladders were used as DNA size markers (lanes A, C, G and T). They were prepared using a PCR-amplified fragment from the pneumococcal R6 genome (1221-bp; coordinates 1597232 to 1598452) and the 1622D oligonucleotide (see Table 3). (**C**) *In vitro* transcription experiments using a reconstituted RNAP (*E. coli* RNAP core enzyme and the pneumococcal σ^{43} factor). Reactions using the 224-bp (Lane 1) or the 265-bp (Lane 2) DNA were loaded onto a sequencing gel (6% PAA, 8M urea). The sizes (in nucleotides) of the transcripts are indicated on the right of the gel. Dideoxy-mediated chain termination sequencing reactions, using a PCR-amplified fragment from the R6 genome (421-bp; coordinates 1598229 to 1598649) and the 1622C primer, were run in the same gel, and used as DNA size markers (lanes A, C, G and T).

Chapter 2

Biophysical characterization of the Mga*Spn* transcriptional regulator

In this chapter, we describe the protocol developed to overproduce and purify an untagged form of the Mga*Spn* regulator. As far as we know, it is the first case within the Mga/AtxA family of global regulators. We have also undertaken the study of the multimeric state of Mga*Spn* in solution by gel filtration chromatography and analytical ultracentrifugation. The secondary structure content of Mga*Spn* has also been determined by circular dichroism studies.

2.1. Purification of the native Mga*Spn* protein

We have developed a protocol to overproduce and purify an untagged form of the Mga*Spn* protein. This protocol is extensively described in Methods (Section 9.2). Basically, the *mgaSpn* gene was cloned into the *E. coli* inducible expression vector pET24b, which is based on a promoter recognized by the T7 RNAP (Figure 10). The recombinant plasmid pET24b-*mgaSpn* was introduced into the *E. coli* BL21 (DE3) strain, which carries the T7 RNAP-encoding gene (T7 gene) fused to the *lac*UV5 promoter. In this system, expression of the T7 gene and, consequently, expression of *mgaSpn* is induced when IPTG is added to the bacterial culture. Various bacterial growth conditions and IPTG concentrations were assayed to define the optimal *mgaSpn* gene expression conditions (TY medium, 37ºC, 1 mM IPTG). The method used for large-scale purification of Mga*Spn* involved essentially three steps (Figure 11): firstly both DNA and Mga*Spn* (presumably bound to DNA) were precipitated with PEI at a low ionic strength, then Mga*Spn* was eluted from the PEI pellet using a higher ionic strength buffer, and finally samples were subjected to heparin affinity chromatography. Fractions containing Mga*Spn* were identified using Coomasie-stained SDS-polyacrylamide (10%) gels (Figure 11). Then, such fractions were pooled, dialyzed and concentrated prior to storage at -80ºC. The yield of pure protein was 4-5 mg per litre of cell culture. Coomassie-stained overloaded gels showed that the protein preparation was more than 95% pure. Under denaturing conditions, Mga*Spn* migrated between the 45 and 66 kDa bands of the molecular weight marker (Figure 11), which agrees with the molecular weight of the Mga*Spn* monomer calculated from its predicted amino acid sequence (58,723.2 Da; 493 residues). In fact, the quantitative amino acid analysis of Mga*Spn* (Pharmacia-Biochrom 20 Amino Acid Analyzer) was in agreement with the amino acid composition predicted from the nucleotide sequence. Determination of the N-terminal amino acid sequence of Mga*Spn* by Edman degradation showed that the first Met residue was not processed after protein synthesis.

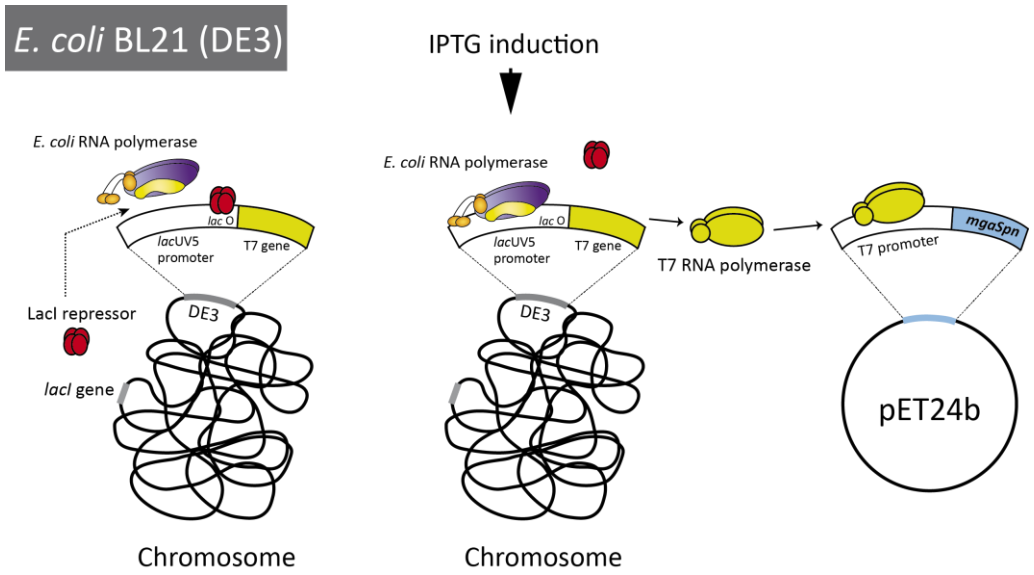


Figure 10. Experimental design to overproduce the Mga*Spn* protein. The *mgaSpn* gene was cloned into the pET24b expression vector under the control of a T7 RNAP promoter (T7 promoter). The recombinant plasmid was introduced into the *E. coli* BL21 (DE3) strain, which carries the T7 RNAP gene (T7 gene) under control of the IPTG-inducible *lac*UV5 promoter. The host strain also carries a chromosomal 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 ϕ10 promoter of the phage T7 (T7 promoter) and the *mgaSpn* gene is transcribed.

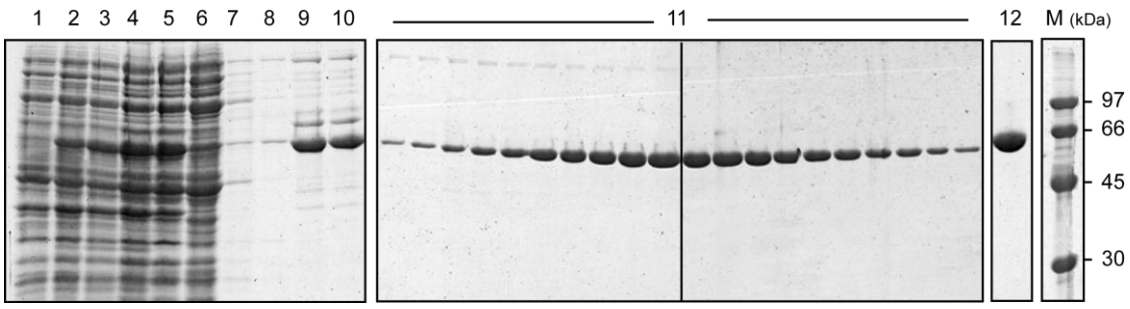


Figure 11. Purification of the native Mga*Spn* protein. Protein fractions of each step were analysed by SDS-polyacrylamide (10%) gel electrophoresis. Lanes 1 to 3: Induction of *mgaSpn* gene expression; (1) without IPTG; (2) with IPTG for 25 minutes; (3) after treatment with rifampicin for 60 minutes. Lanes 4 to 12: Purification steps; (4) supernatant of a whole-cell extract; (5) supernatant of a clear lysate; (6) proteins that remained in the supernatant after precipitation with polyethyleneimine (PEI) at low ionic strength; (7-8) proteins eluted from PEI pellet using a low ionic strength buffer; (9) proteins eluted from the PEI pellet using a higher ionic strength buffer; (10) proteins precipitated with ammonium sulphate; (lane 11) heparin affinity chromatography: elution of the proteins retained in the heparin column with a salt gradient (0.3-0.8 M NaCl); (lane 12) Mga*Spn* preparation after heparin affinity chromatography. M indicates the molecular weight standards (in kDa) (LMW Marker, GE Healthcare).

2.2. Domain organization of Mga*Spn*

The three-dimensional structure of a putative Mga-like transcriptional regulator (EF3013) from *Enterococcus faecalis* has been resolved by X-ray crystallography (Osipiuk *et al*., 2011). Such a structure has been deposited in the Protein Data Bank (PDB 3SQN). With the exception of EF3013, whose function is being characterized in our laboratory, structural data are not available for any member of the Mga/AtxA family of regulators. However, Mga and AtxA were reported to have a similar organization of known or predicted functional domains (Hondorp and McIver, 2007; Tsvetanova *et al*., 2007; Hondorp *et al*., 2013) (see Figure 3). Regarding the Mga*Spn* regulator, our *in silico* analyses indicated that it exhibits similarity to Mga and AtxA in the domain organization. The Pfam protein families database (Punta *et al*., 2012) revealed the presence of two putative DNA-binding motifs within the N-terminal region, the so-called HTH_Mga (residues 6 to 65) and Mga (residues 71 to 158) motifs (Figure 12). Both helix-turn-helix motifs are also present in the Mga (Hondorp and McIver, 2007) and AtxA (Tsvetanova *et al*., 2007) global response regulators (Figure 3). Moreover, in the central region of Mga*Spn*, the Pfam database predicted a PRD domain that includes amino acids 173 to 392. Thus, PTS components might modulate the activity of Mga*Spn* by phosphorylation of histidine residues located within such a domain. Analysis of Mga*Spn* with the protein structure prediction server Phyre2 (Kelley and Sternberg, 2009) confirmed the Pfam predictions and, in addition, revealed structural homology of the C-terminal region (amino acids 399 to 487) to an EIIB-like domain used by the PTS (Figure 12). An EIIB-like domain has also been identified at the C-terminal region of the Mga (Hondorp *et al*., 2013) and AtxA (Hammerstrom *et al*., 2011) regulators.

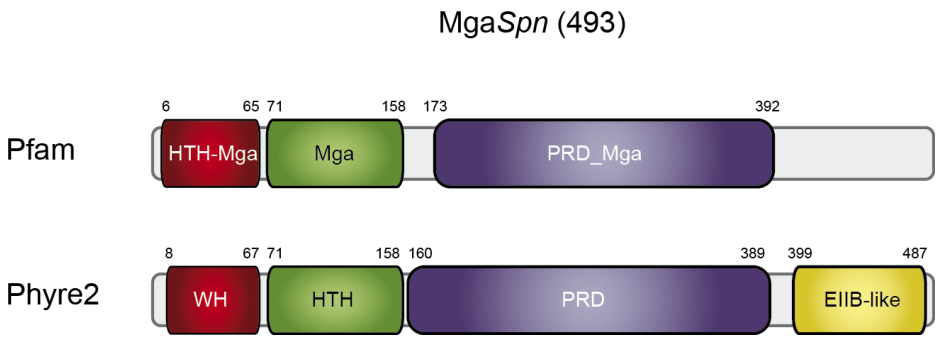


Figure 12. Predicted domains in the Mga*Spn* regulatory protein. At the N-terminal region, two putative DNA binding motifs are indicated: HTH-Mga (Pfam) or WH (Phyre2) and Mga-like HTH motif. A central domain is found to have structural homology to a PTS regulatory domain (PRD). According to Phyre2, the C-terminal region shows structural homology to an EIIB-like domain used by the PTS.

2.3. Oligomerization state of the Mga*Spn* protein

The availability of a method to purify the untagged Mga*Spn* protein to near homogeneity allowed us to study its oligomerization state in solution. To determine the molecular size (Stokes radius) of Mga*Spn*, gel filtration chromatography was performed using a running buffer that contained 250 mM of NaCl. The elution profile is shown in Figure 13A. At a loading concentration of 25 µM (1.45 mg/ml), most of the Mga*Spn* protein eluted as a symmetrical single peak. The elution volume of Mga*Spn* was obtained from experimental data and its corresponding K_{av} value was calculated (see Methods, Section 10.5). Standard proteins of known Stokes radius were loaded at same conditions to obtain a calibration curve. Values of K_{av} were calculated from the elution volume of each protein and plotted against Stokes radii (Figure 13B). The Stokes radius of Mga*Spn* was inferred from the calibration curve and was shown to be 46 Å. This value is very close to the Stokes radius of alcohol dehydrogenase standard protein (45 Å), with a molecular weight of 150 kDa. The same result was obtained when the running buffer contained 100 mM of NaCl. These results indicate that the untagged Mga*Spn* protein appears to be a dimer under our experimental conditions.

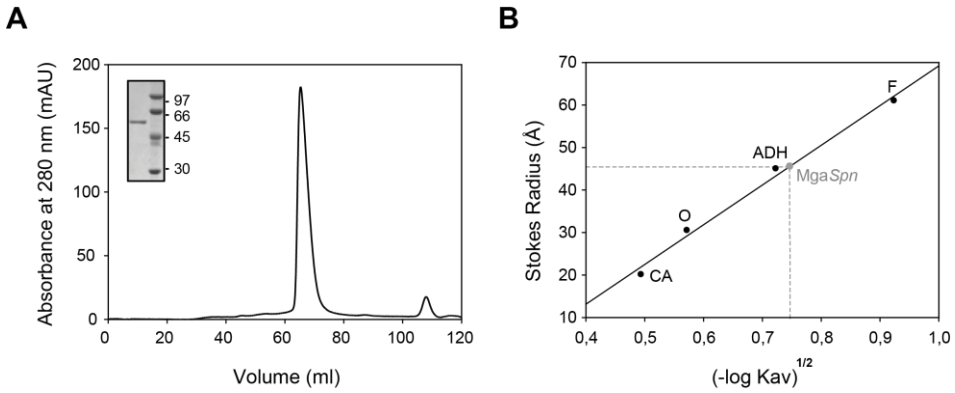


Figure 13. Mga*Spn* exists as a dimer in solution. (**A**) Elution profile of Mga*Spn* on a HiLoad Superdex 200 gel-filtration column. Inset shows analysis of the eluted protein by SDS-polyacrylamide (10%) gel electrophoresis. The molecular weight (in kDa) of proteins used as markers (GE Healthcare) is indicated on the right of the gel. (**B**) Stokes radius of Mga*Spn*. The column was calibrated by loading several standard proteins of known Stokes radius: ferritin (F; 61 Å), alcohol dehydrogenase (ADH; 45 Å), ovalbumin (O; 30.5 Å) and carbonic anhydrase (CA; 20.1 Å). Each protein was prepared at the concentration recommended by the suppliers, in a final volume of 1 ml of buffer VL containing 250 mM NaCl. The flow rate used was 0.5 ml/min.

To further analyse Mga*Spn* oligomerization, analytical ultracentrifugation experiments (sedimentation velocity and sedimentation equilibrium) were carried out at different protein concentrations (5 and 10 µM). At 5 µM, the sedimentation velocity profile showed a major peak (83.6%) with an $S_{20,w}$ value of 5.7 S, and an average molecular mass ($M_{w,a}$) of 113,924 Da, which is compatible with the dimer of the protein. When the

protein concentration was increased up to 10 µM, the same major species was found (77.7%), but species of higher sedimentation coefficient also appeared in the sample (Figure 14A). This indicates that the concentration of Mga*Spn* may affect its multimeric state; nevertheless, the dimeric form appeared to be predominant in the samples. Samples were also analysed by sedimentation equilibrium (Figure 14B). Analysis of the experimental data obtained at a Mga*Spn* concentration of 5 µM indicates an $M_{w,a}$ of 167,719 ± 894 Da, significantly larger than the value expected for the monomer (58723.2 Da) and even for the dimer (117,446.4 Da), suggesting that during the sedimentation equilibrium, the protein tends to form higher-order molecular species. The $M_{w,a}$ of Mga*Spn* at 10 µM (194,648 ± 1,233 Da) was even larger than the theoretical mass of the Mga*Spn* dimer (not shown). Analysis of the data using different models indicated no improvement in the best-fit parameters. Moreover, the frictional ratio (f/f_0) calculated was 1.45, indicating that the hydrodynamic behaviour of Mga*Spn* deviates from that of a rigid spherical particle with a frictional ratio value of 1.0. Thus, Mga*Spn* can be expected to have an ellipsoidal shape. Taking into account the experimental results, in solution and under the conditions tested, Mga*Spn* is able to establish protein-protein interactions, generating molecular species that might be elongated homo-dimers or higher-order oligomers. Moreover, the concentration of the protein may favour the formation of homo-multimers.

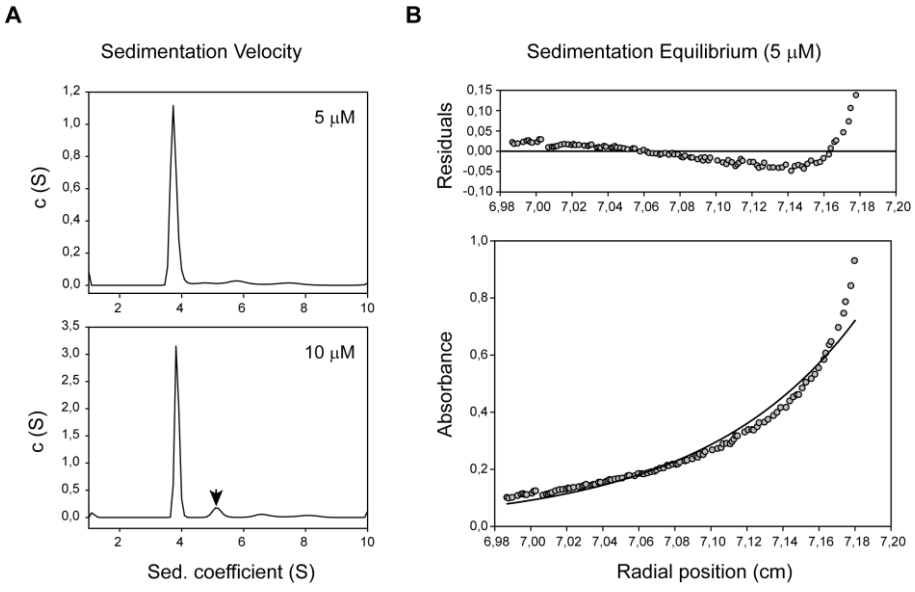


Figure 14. Analytical ultracentrifugation of Mga*Spn*. (**A**) Sedimentation velocity profiles of Mga*Spn* (5 µM and 10 µM) at 43,000 rpm and 12ºC. The lines represent the best fits to a model of a single sedimenting species (3.8 S; $S_{20,w}$ = 5.7 S). The arrow indicates a molecular species of higher sedimentation coefficient (5.1 S) that appears at 10 µM and represents a 9.3% of total. (**B**) Sedimentation equilibrium profile of Mga*Spn* (5 µM) at 9,500 rpm and 10ºC. Grey circles represents the experimental data and; the continuous line is the best-fit Mw (167,719 ± 894 Da).

2.4. Secondary structure content and thermal stability of Mga*Spn*

The secondary structure content of the Mga*Spn* regulator was analysed by computational methods and by far-ultraviolet (far-UV) circular dichroism (CD) spectroscopic analyses. According to the prediction programs used (Table 7), Mga*Spn* has a high content of α-helices (57.4-66.3%) and a low content of β-strands (8.1-9.7%). The distribution of secondary structure elements predicted by the SABLE program (Adamczak *et al*., 2005) is shown in Figure 15A.

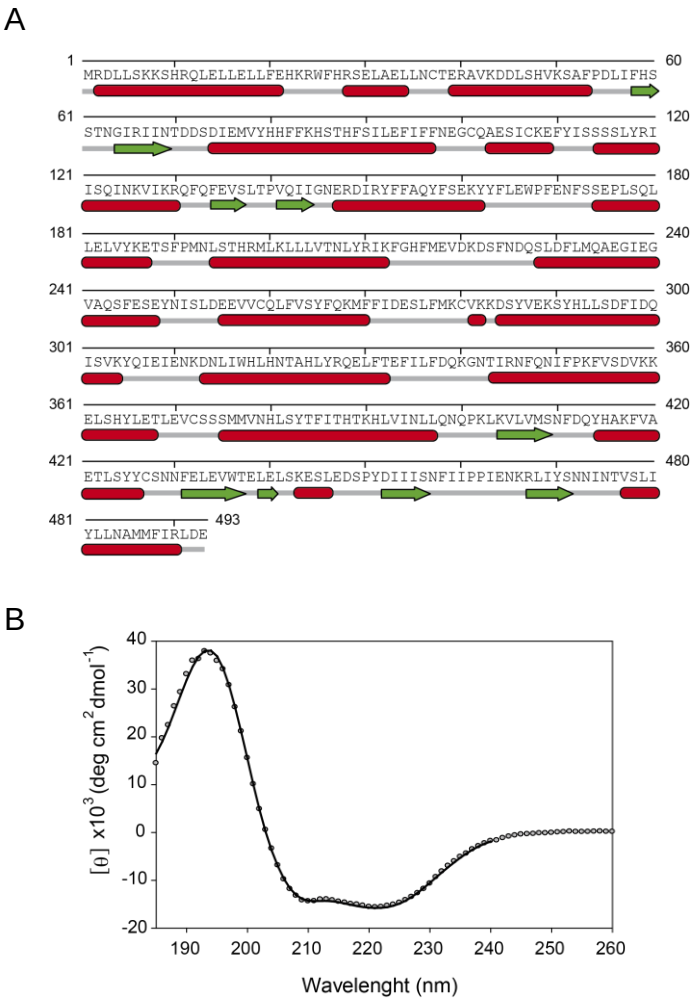


Figure 15. Secondary structure content of Mga*Spn*. (**A**) *In silico* prediction of the secondary structure content of Mga*Spn*. The amino-acid sequence of Mga*Spn* is shown as well as the distribution of secondary structure elements predicted by the SABLE program (Adamczak *et al*., 2005). α-helices are represented by red cylinders while β-sheets are represented by green arrows. (**B**) CD spectrum of Mga*Spn* in the far-UV region at 4ºC. Data were obtained using a 0.2-mm optical path-length quartz cell in CD buffer. Experimental data (grey circles) and the fitted data using the CONTINLL algorithm (continuous line) are represented. The spectra are expressed as mean residues ellipticities (y axis) at given wavelength (x axis).

These predictions were further confirmed by CD analyses (Figure 15B). The CD spectrum of native Mga*Spn* (12 µM) was measured in the far-UV region (185-260 nm). It is characterized by two minima at 208 and 222 nm, and a maximum at around 192 nm, indicative of the presence of α-helical structures. Deconvolution of the CD spectrum by different methods (Table 7) gave a consensus average of 55.3% α-helices, 10.5% β-strands and 37.8% random conformation and turns.

Table 7. Secondary structure content of Mga*Spn*

		α-helix	*β-strand*	*Turns*	Unordered
Deconvolution method	SELCON3	54.1	10.6	13.3	21.5
	CONTIN	53.6	1.4	13.1	22
	CDSSTR	54.5	12	14.7	18.7
	K2D	59	8	N.D.	34
Average		**55.3±2.5**	**10.5±1.8**	**13.7±0.9**	**24.1±6.8**
Prediction program	SABLE	60.7	8.5	N.D.	30.8
	PSIPred	66.3	8.1	N.D.	25.6
	JPred	57.4	9.1	N.D.	33.5
	NPS@	61.2	9.7	N.D.	29
	PredictProtein	60	8.5	N.D.	31.4
Average		**61.1±3.2**	**8.8±0.6**	**N.D.**	**30.1±2.9**

Data and standard deviations are expressed in percentage. N.D. non determined.

CD spectroscopic studies at different temperatures were done to analyse the thermal stability of Mga*Spn* (12 µM). Denaturation of the protein in response to temperature produces changes in the ellipticity values. These changes were monitored in the far-UV region (190-260 nm) while the temperature was increased from 4ºC to 90ºC at the rate of 50ºC/h or 20ºC/h with similar results. CD profiles were recorded from 190 to 260 nm at 4, 12, 25, 37, 50, 60, 70 and 90ºC (Figure 16). At 50ºC there was a progressive loss of ellipticity indicating that denaturation of Mga*Spn* starts around this temperature. Above 50ºC, the ellipticity strongly decreased as the temperature was increased. After the maximum temperature (90ºC) was reached, the sample was cooled to the initial temperature (4ºC) to study the reversibility of Mga*Spn* denaturation. The spectrum of Mga*Spn* was not recovered. Similar results were obtained when the maximum temperature was limited to 50ºC (not shown) indicating that, once protein denaturation starts, Mga*Spn* is not able to refold its secondary structure to its native form.

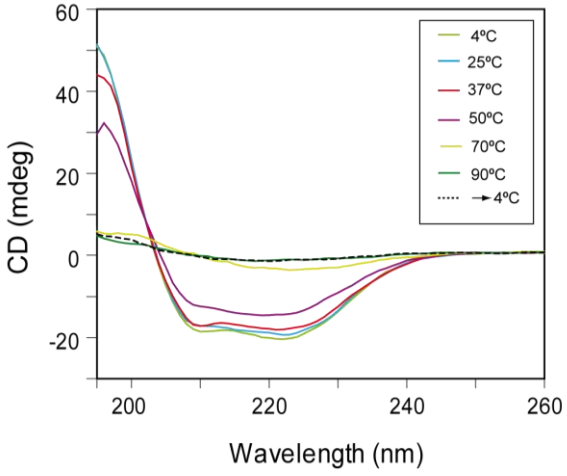


Figure 16. Temperature-associated changes in the secondary structure of Mga*Spn*. CD spectra measured during thermal denaturing (solid lines) and refolding (heated sample re-cooled to 4ºC; dotted line) processes. A perceptible loss of secondary structure was observed from 50ºC. Re-cooled samples did not recover the secondary structure once they were denatured, indicating an irreversible damage to the Mga*Spn* secondary structure.

Chapter 3

Activator role of the Mga*Spn* transcriptional regulator

In this chapter, we describe an in depth transcriptional analysis of the *spr1623-spr1626* operon and the ability of Mga*Spn* to activate its transcription. We published this study in 2012 (Solano-Collado *et al*., 2012; see Related publications) and it was the first report describing the activator role of the pneumococcal Mga-like regulator.

3.1. Construction of pneumococcal R6 mutant strains

To study the effect of Mga*Spn* on gene expression, we constructed two derivatives of the pneumococcal R6 strain: a deletion mutant strain (R6Δ*mga*) which does not synthesize Mga*Spn* and a strain designed to overproduce Mga*Spn* (R6/pDL*PsulA::mga*). The latter strain carries the *mgaSpn* gene cloned into a plasmid (Figure 17B). Specifically, we first constructed the *PsulA::mga* fusion gene, in which the *Pmga* promoter of the *mgaSpn* gene was replaced with the promoter of the pneumococcal *sulA* gene (*PsulA*) (Lacks *et al*., 1995). The activity of the *PsulA* promoter region had been tested previously using a promoter-probe vector (Ruiz-Cruz *et al*., 2010). The *PsulA::mga* fusion gene was then inserted into the pDL287 plasmid (LeBlanc *et al*., 1993) generating the pDL*PsulA*::*mga* recombinant plasmid. Pneumococcal R6 cells either containing or lacking the recombinant plasmid were grown under standard laboratory conditions, and the intracellular amount of Mga*Spn* was estimated by Western-blotting (Figure 17C). Compared to R6 cells without plasmid (Figure 17C, lane 1) or R6 cells carrying pDL287 (not shown), the amount of Mga*Spn* increased about 8-fold in cells harbouring the pDL*PsulA*::*mga* recombinant plasmid (Figure 17C, lane 3). The higher intracellular level of Mga*Spn* did not affect the bacterial growth rate (not shown).

For the construction of the R6Δ*mga* mutant strain (Figure 17A), the chromosomal region that includes the *mgaSpn* gene and its promoter (*Pmga*) (coordinates 1596826 to 1598431) was replaced with the *cat* gene of the pC194 plasmid, which confers chloramphenicol resistance (Horinouchi and Weisblum, 1982). To do this, we first constructed a cassette composed of the adjacent regions of *mgaSpn* flanking the *cat* gene (described in Methods, Section 3.1). Then, competent R6 cells were transformed with the *cat* cassette and by homologous recombination the *mgaSpn* gene was replaced with the *cat* gene. Cells lacking the *mgaSpn* gene were selected based on its resistance to chloramphenicol (1.5 µg/ml). As expected, the R6Δ*mga* strain did not synthesize Mga*Spn* (Figure 17C, lane 2).

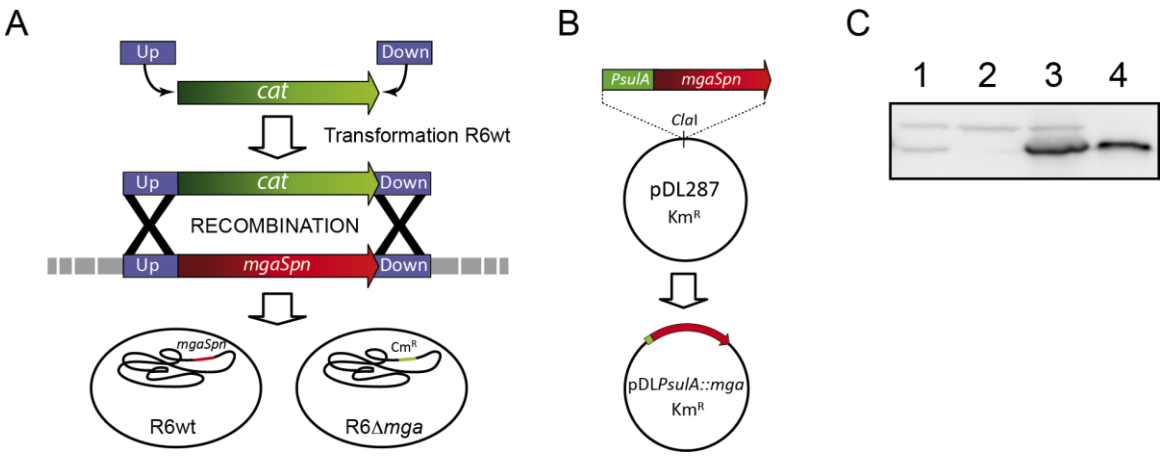


Figure 17. Construction of pneumococcal R6 mutant strains and detection of Mga*Spn* in whole-cell extracts by Western-blotting. (**A**) Construction of the R6Δ*mga* strain. The *mgaSpn* gene and its promoter were replaced with the *cat* gene (chloramphenicol resistance) of the pC194 plasmid. First, two DNA fragments that flank the *mgaSpn* gene (called Up and Down) were PCR-amplified and ligated to the *cat* fragment. The cassette generated was used to transform competent R6 cells. Finally, two homologous recombination events led to the replacement of the *mgaSpn* gene with the *cat* gene. Cells that lacked the *mgaSpn* gene were selected based on their resistance to chloramphenicol (1.5 µg/ml). (**B**) Construction of a mutant strain that overproduces Mga*Spn*. The *Pmga* promoter of the *mgaSpn* gene was replaced with the promoter of the *sulA* gene (Lacks *et al*., 1995). The fusion *PsulA::mga* gene was then inserted into the *Cla*I site of plasmid pDL287 (LeBlanc *et al*., 1993), generating the pDL*PsulA::mga* recombinant plasmid, which was further used to transform pneumococcal R6 cells. (**C**) Detection of Mga*Spn* in pneumococcal cell extracts by Western-blotting using polyclonal antibodies against the Mga*Spn*-His protein. Total proteins from R6 cells (lane 1), R6Δ*mga* (lane 2), and pDL*PsulA::mga*-harbouring cells (lane 3) were separated by SDS-PAGE (10% polyacrylamide). His-tagged Mga*Spn* protein (6 ng) (lane 4) and pre-stained proteins (Invitrogen) (not shown) were run in the same gel. Cross-reactive bands were used as loading controls.

3.2. Preliminary proteomic assays

To identify the target genes of the Mga*Spn* regulator, we analysed the effect of the absence of Mga*Spn* on the pattern of global gene expression comparing the R6 wild-type and R6Δ*mga* strains by proteomics. Pneumococcal cells were grown under standard laboratory conditions. These experiments were carried out in collaboration with Dr. J. A. López at the Proteomics Facility of the Centro Nacional de Investigaciones Cardiovasculares (CNIC, Madrid). The results obtained showed significant changes (increased or decreased levels in the absence of Mga*Spn*) in 10 candidates, including proteins involved in competence, in metabolism of amino acids, heat shock proteins and proteins of unknown function. Interestingly, one of the proteins whose levels decreased (2.3-fold) in the absence of Mga*Spn* was the putative general stress protein 24, encoded by the *spr1625* gene, indicating that Mga*Spn* might be acting, directly or indirectly, as a positive regulator of the *srp1625* gene. This fact, together with the location of *spr1625*

close to *mgaSpn* on the R6 genome (Figure 18), led us to work on the validation of this result. The transcriptional analysis of the *spr1625* gene, as well as its regulation by Mga*Spn* is described in the following sections.

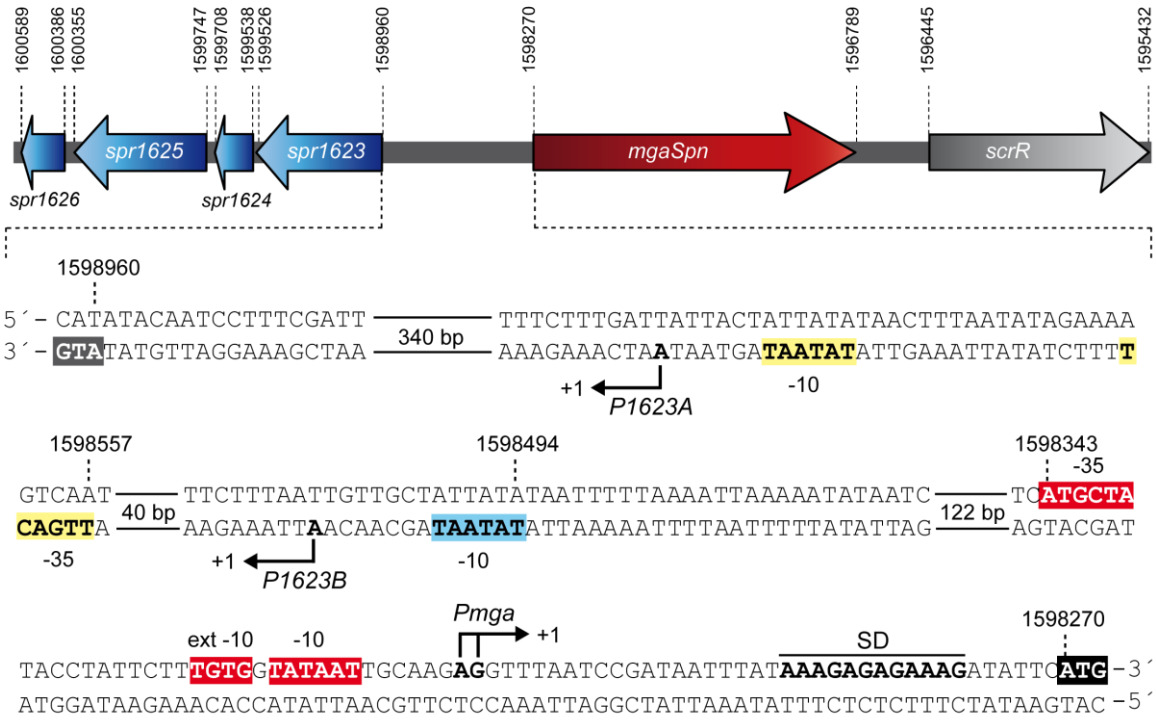


Figure 18. Genetic map of the region spanning the 1596789 and 1600589 coordinates of the pneumococcal R6 genome (Hoskins *et al.*, 2001). For each gene, the coordinates of the start and stop codons are indicated. The nucleotide sequence of the region spanning the start codon (ATG) of the *mgaSpn* gene (coordinate 1598270) and the start codon (ATG) of the *spr1623* gene (coordinate 1598960) is shown. The putative Shine-Dalgarno sequence (SD) of the *mgaSpn* gene is indicated (bold letters). The main sequence elements (-35 box, -10 box and extended –10 box) of the promoters identified in this work (*Pmga* in red, *P1623B* in blue and *P1623A* in yellow), as well as the transcription start sites (+1 position, arrows), are indicated.

3.3. Transcription of the *spr1623-spr1626* operon in pneumococcal R6 cells

The *spr1625* gene is located upstream of the *mgaSpn* gene and on the complementary strand. It is the third of four genes of unknown function that appear to be organized in an operon (*spr1623-spr1626*) (Figure 18). Due to this organization, the *mgaSpn* gene and the putative four-gene operon would be divergently transcribed. The ATG codon at coordinate 1598960 is probably the translation initiation codon of the *spr1623* gene, and it is located at 689-bp from the ATG initiation codon of the *mgaSpn* gene (coordinate 1598270). To investigate whether the putative operon was transcribed under our experimental conditions, we performed RT-PCR assays. Initially, the 1623B

primer, which anneals with the *spr1623* transcript, was used for extension on total RNA isolated from R6 cells (Figure 19). The cDNA extension product was then amplified by PCR using either the 1623B and 1623C or the 1623B and 1623A oligonucleotides. With the pair of primers 1623B and 1623C, a PCR product with the mobility expected for a 695-bp fragment was detected. This product was not visualized in the negative control (total RNA as template). When the cDNA product was amplified using the 1623B and 1623A primers, no PCR products were detected. However, these primers amplified an 892-bp region when chromosomal DNA was used as template (positive control) (Figure 19). These results indicate that the *spr1623* gene is transcribed under our experimental conditions. In addition, we performed RT-PCR experiments to elucidate whether the four genes were transcribed as a single polycistronic mRNA. Specifically, the 1626A primer was used for the cDNA synthesis (Figure 19). Amplification of the cDNA product with the 1626A and 1623C primers generated a product with the mobility expected for a 1,917-bp fragment. This product was not detected when total RNA was used as template for the PCR reactions (negative control) but it was synthesized in the positive control. Collectively, these results indicate that the *spr1623*, *spr1624*, *spr1625* and *spr1626* genes are transcribed into a polycistronic mRNA molecule from a site(s) located between the 1598433 and 1598630 coordinates.

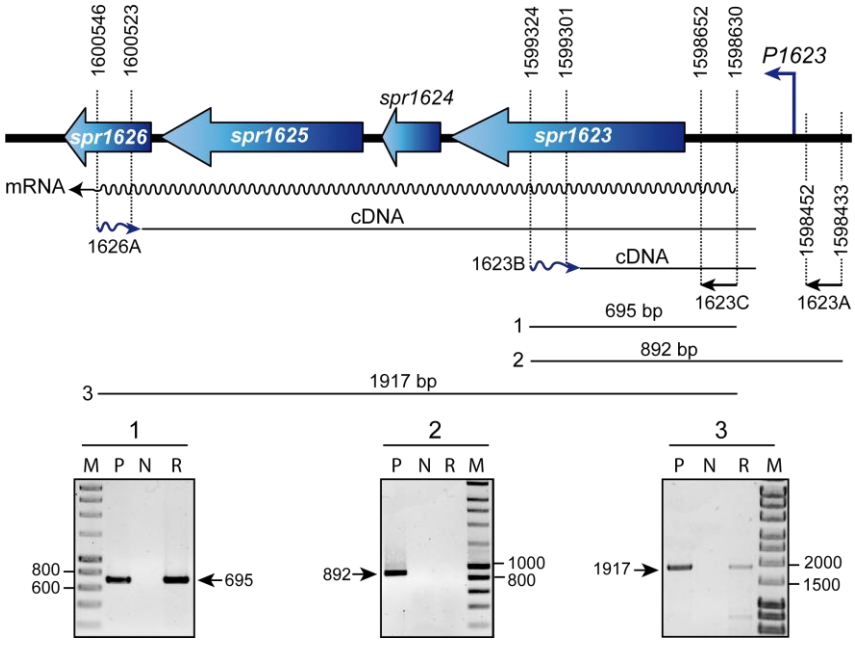


Figure 19. The *spr1623-spr1626* genes constitute an operon. RT-PCR experiments were performed using total RNA isolated from R6 cells. The positions of the oligonucleotides used (1623A, 1623B, 1623C and 1626A) are shown. RT-PCRs (lanes R) were subjected to agarose (0.8%) gel electrophoresis and visualized with ethidium bromide. As negative controls (lane N), RT-PCRs were carried out without adding the reverse transcriptase. The sizes of the PCR-amplified DNA fragments (1, 2, and 3) using genomic DNA as template (lanes P, positive control) are indicated. Lanes M, DNA fragments used as molecular weight markers (in bp) (HyperLadder I, Bioline).

Additionally, we carried out a sequence analysis of the region spanning coordinates 1598960 to 1598270 (intergenic region between *spr1623* and *mgaSpn*) with the BPROM program. This analysis predicted a promoter (hereon named *P1623A*; Figure 18) that exhibits a canonical -10 hexamer (5´-**TATAAT**-3´) and a near consensus -35 hexamer (5´-**TTGAC**t-3´). The distance between these two elements is 17 nt, which is the optimum spacer length in promoters recognised by an RNAP carrying a housekeeping σ factor similar to the *E. coli* σ^{70} (Haugen *et al*., 2008). To investigate whether the predicted promoter *P1623A* was functional *in vivo*, we performed primer extension experiments using total RNA from R6 cells and the PDA oligonucleotide, which anneals with a sequence located downstream of the *P1623A* promoter (Figure 20). We were able to detect two cDNA products of 106 and 191 nt, which may correspond to transcription initiation events at coordinates 1598592 (*P1623A* promoter) and 1598507, respectively. Therefore, the pneumococcal RNAP recognised not only the *P1623A* promoter but also a promoter sequence located upstream of the *P1623A* promoter, which was not identified by bioinformatics programs. Such a promoter sequence (termed *P1623B*) has a consensus -10 hexamer (5´-**TATAAT**-3´) but lacks a -35 element (Figure 18). To confirm the functionality of the *P1623B* promoter *in vivo*, primer extension experiments were performed using the PDB primer. A cDNA product of 60 nt was synthesized, indicating that the *P1623B* promoter was functional (Figure 20).

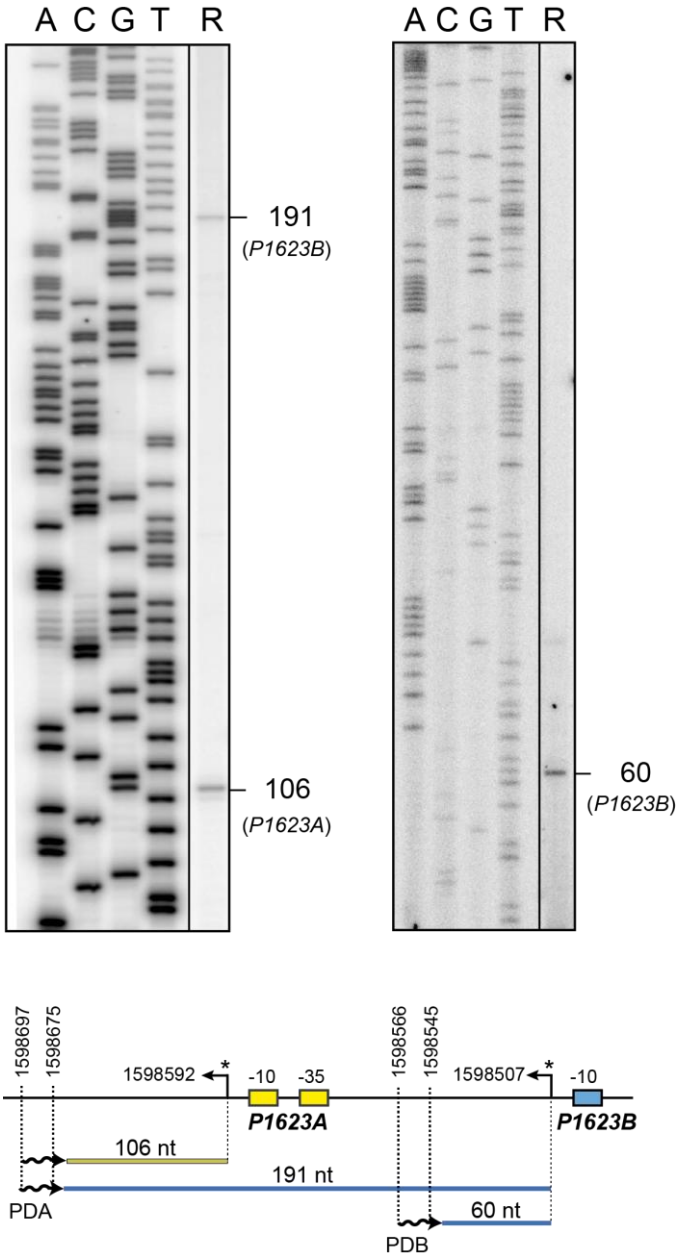


Figure 20. The *spr1623-spr1626* operon is transcribed from promoters *P1623A* and *P1623B*. Primer extension experiments were carried out using total RNA from R6 cells and either the PDA (left gel) or the PDB (right gel) primers. The sizes of the cDNA extension products (lanes R) are indicated in nucleotides on the right of the gels. A, C, G, and T sequence ladders were used as DNA size markers. They were prepared using M13mp18 DNA (Yanisch-Perron *et al.*, 1985) and the −40 M13 primer (5´-GTTTTCCCAGTCACGAC-3´) (left gel) or a PCR-amplified DNA fragment from the *E. faecalis* V583 genome and the Rev primer (5´-GATTTCTTCAATTTGTTCCATC-3´) (right gel). The asterisks in the scheme below the gels indicate the transcription start sites identified in this study.

3.4. Mga*Spn* activates the *P1623B* promoter *in vivo*

To investigate whether the Mga*Spn* protein influenced the activity of a particular promoter *in vivo*, we used the R6/pDL*PsulA::mga* strain that overproduce Mga*Spn* (see section 3.1 and Figure 17B). Specifically, we analysed the effect of the Mga*Spn* overproduction on the activity of the chromosomal *P1623A* and *P1623B* promoters by primer extension. For the cDNA synthesis, we used the 5´-labelled PDA oligonucleotide, which anneals to the *spr1623* transcript (see Figure 21). As a loading control, we

included in the reactions the 5´-labelled PErpoE primer, which anneals to the *rpoE* transcript. The *rpoE* gene (*spr0437* in the R6 genome) encodes the delta subunit of the RNAP (Figure 21). When total RNA from R6/pDL287 cells was used as template (Figure 21, lane 1; low levels of Mga*Spn*), three cDNA products of 106 nt, 191 nt and 231 nt were detected, that correspond with products synthesized from the *P1623A* promoter, the *P1623B* promoter and the *PrpoE* promoter, respectively. Unlike the 231 nt product, which served as a loading control, the amount of the 106 nt and 191 nt cDNAs increased 2.6-fold and 4.5-fold, respectively, when total RNA from R6/pDL*PsulA*::*mga* cells was used (Figure 21, lane 2; overproduction of Mga*Spn*). Therefore, the overproduction of Mga*Spn* led to the activation of both the *P1623A* and *P1623B* promoters, although the effect appeared to be greater on the activity of the *P1623B* promoter.

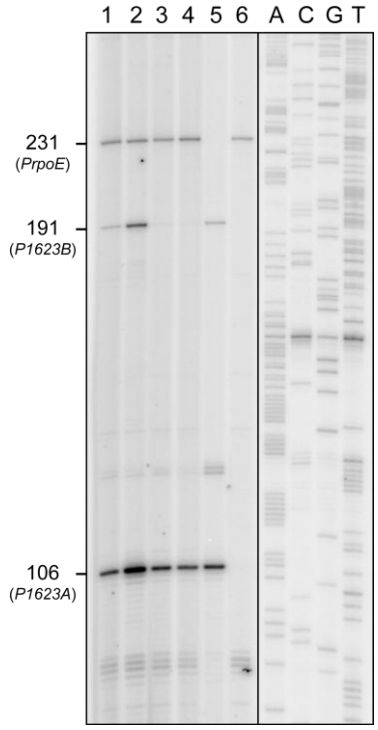


Figure 21. Mga*Spn* mediates activation of the *P1623B* promoter. Primer extension reactions were performed using total RNA from R6/pDL287 (lanes 1, 5, and 6), R6/pDL*PsulA::mga* (lane 2), R6Δ*mga*/pDL287 (lane 3), or R6Δ*mga*/pDL*PsulA::mga* (lane 4) cells. Used as primers were 5´-labelled oligonucleotides: a mix of the PDA and PErpoE primers (lanes 1, 2, 3, and 4), primer PDA (lane 5), or primer PErpoE (lane 6). The sizes of the cDNA extension products are indicated on the left of the gel in nucleotides: 106 nt for the *P1623A* promoter, 191 nt for the *P1623B* promoter, and 231 nt for the *PrpoE* promoter. A, C, G, and T sequence ladders were used as DNA size markers. They were prepared using a PCR-amplified DNA fragment from the *E. faecalis* V583 genome and the Fw primer (5´-CGTTTGAGCAATATAATCGTTTG-3´).

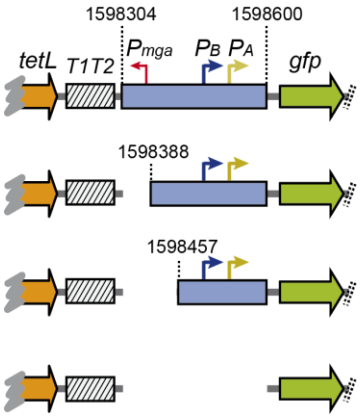
We next examined the activity of the chromosomal *P1623A* and *P1623B* promoters in R6Δ*mga* cells carrying either plasmid pDL287 (absence of Mga*Spn*) (Figure 21, lane 3) or plasmid pDL*PsulA*::*mga* (overproduction of Mga*Spn*) (Figure 21, lane 4). The R6Δ*mga* strain lacks the *mgaSpn* gene and its promoter (*Pmga*), but conserves the *P1623A* and *P1623B* promoter sequences (see Figure 18). As before, a mix of the 5´-labelled PDA and PErpoE primers was used in primer extension assays. Compared to R6/pDL287 cells (lane 1; low levels of Mga*Spn*), the 191 nt product was not detected in

R6Δ*mga*/pDL287 cells (lane 3; absence of Mga*Spn*), although no changes were found in the amount of the 106 nt and 231 nt cDNAs. Thus, low intracellular levels of Mga*Spn* resulted in the activation of the *P1623B* promoter without affecting the activity of the *P1623A* promoter. However, unexpectedly, the activity of the *P1623B* promoter on the R6Δ*mga* genome did not change in the presence of the pDL*PsulA*::*mga* plasmid (lane 4; overproduction of Mga*Spn*). These results suggest that the genome of the R6Δ*mga* strain lacks not only the *mgaSpn* gene (including the *Pmga* promoter) but also a site required for Mga*Spn*-mediated activation of the *P1623B* promoter (see below).

3.5. Mapping the site required for Mga*Spn*-mediated activation of the *P1623B* promoter

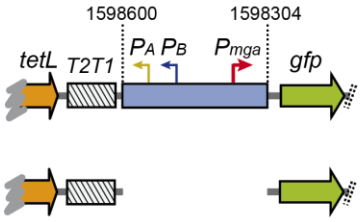
In order to delimit the region necessary for the activation of the *P1623B* promoter by Mga*Spn*, we carried out a deletion analysis using the promoter-probe vector pAST (Figure 22A) (Ruiz-Cruz *et al*., 2010). This vector carries a MCS between the *T1-T2* tandem terminators of the *E. coli rrnB* rRNA operon and a promoter-less *gfp* gene. The presence of the *T1-T2* terminators results in an efficient transcriptional termination of the *tetL* gene (tetracycline resistance, Lacks *et al*., 1986). Plasmid pAST constitutes a useful tool for measuring the promoter activity of regions inserted into the MCS. For the deletion analyses, three R6 chromosomal regions were inserted independently into the *Sac*I site of pAST (Figure 22A): *(i)* the *PAB* region (coordinates 1598304-1598600); *(ii)* the *PAB*Δ84 region (coordinates 1598388-1598600) and *(iii)* the *PAB*Δ153 region (coordinates 1598457-1598600), generating the recombinant plasmids pAST-*PAB,* pAST-*PAB*Δ84 and pAST-*PAB*Δ153, respectively. In all these constructions, expression of the *gfp* gene is under control of both, the *P1623A* and *P1623B* promoters, as shown in Figure 22A. Subsequently, each recombinant plasmid was introduced into R6 and R6Δ*mga* cells, and the promoter activity of each chromosomal region was evaluated by measuring *gfp* expression (fluorescence assays) (Figure 22A). Compared to R6Δ*mga* cells (absence of Mga*Spn*), the promoter activity of the *PAB* and *PAB*Δ84 regions was twice that observed in R6 cells (low levels of Mga*Spn*). However, the promoter activity of the *PAB*Δ153 region was similar in both genetic backgrounds. These results suggest that the chromosomal region spanning coordinates 1598388 and 1598457 contains sequences that are needed for Mga*Spn*-mediated activation of the *P1623A* and/or *P1623B* promoters.

A



	Intensity of fluorescence	
	R6	R6Δ*mga*
pAST-*PAB*	10.75 ± 0.75	5.64 ± 0.38
pAST-*PAB*Δ84	13.03 ± 1.37	6.36 ± 0.24
pAST-*PAB*Δ153	7.26 ± 0.86	6.74 ± 0.08
pAST	0.23 ± 0.13	0.25 ± 0.04

B



	R6	R6Δ*mga*
pAST2-*Pmga*	2.17 ± 0.07	2.21 ± 0.29
pAST2	1.09 ± 0.07	0.89 ± 0.06

Figure 22. Fluorescence assays. (**A**) Activity of the *P1623A* (*PA*) and *P1623B* (*PB*) promoters. The promoter-probe vector pAST was described previously (Ruiz-Cruz *et al.*, 2010). The positions of the *tetL* (tetracycline resistance) and *gfp* (green fluorescent protein) genes are shown. The *T1T2* box represents the tandem terminators *T1* and *T2* of the *E. coli rrnB* rRNA operon. Blue boxes represent DNA fragments of the R6 genome used for the deletion analysis. (**B**) Activity of the *Pmga* promoter. Plasmid pAST2 was described (Ruiz-Cruz *et al.*, 2010). Compared to pAST, it carries the *T2T1rrnB* region inserted in the opposite orientation (box *T2T1*). The position of promoter *Pmga* is indicated. The intensity of fluorescence (arbitrary units) corresponds to 0.8 ml of culture (OD_{650} = 0.3). In each case, three independent cultures were analysed.

We further evaluated the promoter activity of each chromosomal region by primer extension (Figure 23). For these assays, we used a mix of the 5´-labelled INTgfp and ASTtetL primers, which anneal to the *gfp* and *tetL* transcripts, respectively. The *tetL* gene of pAST plasmid was used as internal control. When total RNA from R6 cells (low levels of Mga*Spn*) harbouring plasmid pAST-*PAB*Δ84 (Figure 23, lane 2) was used, we detected three cDNA extension products of 102 nt, 111 nt and 196 nt, which correspond with transcription from *PtetL*, *P1623A* and *P1623B* promoters, respectively. Unlike the 102 nt and 111 nt cDNAs, the amount of the 196 nt cDNA decreased 5-fold when total RNA from R6 cells harbouring plasmid pAST-*PAB*Δ153 was used (Figure 23, lane 3). Thus, deletion of the region spanning coordinates 1598388 and 1598457 reduced the activity of promoter *P1623B* but not the activity of promoter *P1623A*. Such a specific decrease in the activity of *P1623B* promoter was also observed in R6Δ*mga* cells

(absence of Mga*Spn*) carrying either *PAB*Δ84 (Figure 23, lane 4) or *PAB*Δ153 (Figure 23, lane 5).

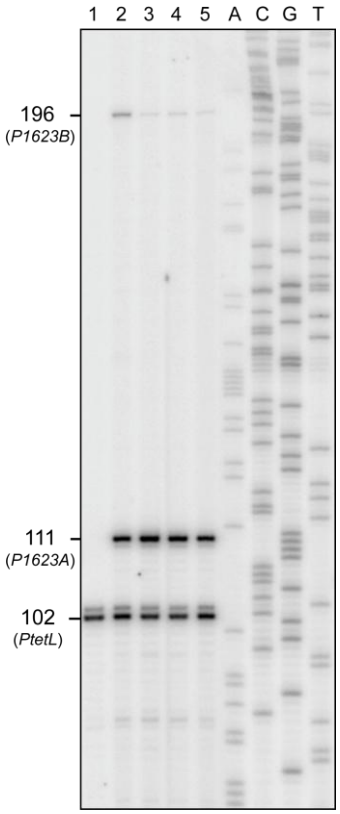


Figure 23. Genomic region needed for Mga*Spn*-mediated activation of the *P1623B* promoter. Primer extension reactions were carried out using total RNA isolated from R6/pAST-*PAB*Δ84 (lanes 1 and 2), R6/pAST-*PAB*Δ153 (lane 3), R6Δ*mga*/pAST-*PAB*Δ84 (lane 4), and R6Δ*mga*/pAST-*PAB*Δ153 (lane 5) cells. Used as primers were 5´-labelled oligonucleotides: a mix of the INTgfp and ASTtetL primer (lanes 2, 3, 4, and 5) or the ASTtetL primer (lane 1). The sizes of the cDNA extension products are indicated on the left of the gel in nucleotides: 102 nt for the *PtetL* promoter, 111 nt for the *P1623A* promoter, and 196 nt for the *P1623B* promoter. Dideoxy-mediated chain-termination sequencing reactions using pAST DNA and the INTgfp primer were run in the same gel as DNA size markers (lanes A, C, G, and T).

These results therefore demonstrate that Mga*Spn* is able to act, directly or indirectly, as a positive regulator of the *P1623B* promoter *in vivo*. This activation requires a 70-bp region (termed henceforth as the *PB* activation region; coordinates 1598388 to 1598457), which maps between the *P1623B* and *Pmga* divergent promoters and is 50-bp upstream of the *P1623B* transcription start site (coordinate 1598507) (Figure 18 and 22A).

3.6. Mga*Spn* does not influence the activity of the *Pmga* promoter *in vivo*

The Mga protein of *S. pyogenes* regulates positively the expression of its own gene during the exponential phase of growth in order to amplify the Mga regulon (McIver *et al*., 1999). Therefore, we investigated whether Mga*Spn* was able to affect the activity of its own promoter *in vivo*. Moreover, our data showed that the site required for Mga*Spn*-mediated activation of the *P1623B* promoter is located upstream of the *Pmga* promoter (Figure 18 and 22A). This fact suggested that Mga*Spn* might also influence the activity of its own promoter. To test this hypothesis we constructed a transcriptional fusion between *Pmga* and the *gfp* gene of plasmid pAST2 (Figure 22B). Compared to the promoter-probe vector pAST, plasmid pAST2 (referred to as pAS-*T2T1 rrnB* in (Ruiz-Cruz *et al*., 2010) carries the *T1T2rrnB* transcriptional terminator region inserted in the opposite orientation (*T2T1rrnB* region). We inserted the *PAB* chromosomal region (coordinates 1598304-1598600) into the *Sac*I site of the pAST2 plasmid, generating the pAST2-*Pmga* plasmid (Figure 22B). Such a region carries the *P1623A*, *P1623B* and *Pmga* promoters. In pAST2-*Pmga*, expression of the *gfp* gene is under control of the *Pmga* promoter. The *T2T1rrnB* region functions as a transcriptional terminator signal of the *tetL* gene, although it is less efficient than the *T1T2rrnB* region (Ruiz-Cruz *et al*., 2010). In addition, the *T2T1rrnB* region ensures an efficient termination of the transcription initiated at both divergent promoters, *P1623A* and *P1623B*. Plasmids pAST2 and pAST2-*Pmga* were then introduced into R6 (low levels of Mga*Spn*) and R6Δ*mga* (absence of Mga*Spn*) strains, and *gfp* expression was monitored by fluorescence assays. The intensity of fluorescence in R6 cells carrying the pAST2-*Pmga* plasmid was about 2-fold higher than in R6 cells harbouring pAST2, indicating that the *Pmga* promoter was functional. However, the activity of the *Pmga* promoter was similar in both, R6 and R6Δ*mga* cells. Thus, Mga*Spn* did not influence the activity of the *Pmga* promoter placed on the plasmid.

With this experimental approach we cannot rule out the possibility that the results obtained were because of a single chromosomal copy of *mgaSpn* might not be sufficient for regulation of the *Pmga* promoter placed on a multi-copy plasmid. To test this hypothesis, we introduced the pAST2-*Pmga* recombinant plasmid into cells that overproduce Mga*Spn* (pDL*PsulA::mga*-harbouring R6Δ*mga* cells). It was also introduced into cells that do not synthesize Mga*Spn* (R6Δ*mga* carrying the pDL287 plasmid) as control. Again, expression of the *gfp* gene was monitored by fluorescence assays. No changes in the activity of the *Pmga* promoter were observed (not shown). This conclusion was further confirmed by primer extension experiments using total RNA

isolated from R6Δ*mga*/pDL287 and R6Δ*mga*/pDL*PsulA::mga* cells harbouring the pAST2-*Pmga* plasmid. As primers, a mix of the 5´-labelled INTgfp and ASTtetL oligonucleotides, which anneal to the *gfp* and *tetL* transcripts, respectively, were used (Figure 24). The *tetL* gene of pAST2 plasmid was used as internal control. As shown in Figure 24, we detected two cDNA extension products of 102 nt and 110 nt, which correspond with transcription from *PtetL* and *Pmga*, respectively. No differences were observed in the amount of cDNAs independently of the presence of high levels of intracellular Mga*Spn*. These results confirmed our previous findings and supported that Mga*Spn* did not influence the activity of the *Pmga* promoter under the conditions tested.

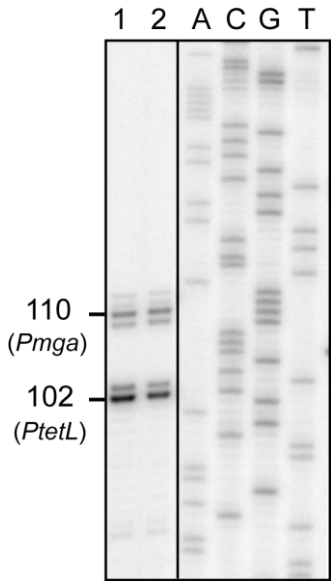


Figure 24. High levels of intracellular Mga*Spn* do not influence the activity of *Pmga*. Primer extension reactions were carried out using total RNA isolated from R6Δ*mga*/pDL287 (lane 1) and R6Δ*mga*/pDL*Psul::mga* cells (lane 2). A mix of the INTgfp and ASTtetL primers (lanes 1 and 2) was used for the extension reactions. The sizes of the cDNA extension products are indicated on the left of the gel in nucleotides: 102 nt for the *PtetL* promoter and 110 for the *Pmga* promoter. Dideoxy-mediated chain-termination sequencing reactions using pAST DNA and the INTgfp primer were run in the same gel as DNA size markers (lanes A, C, G, and T).

Chapter 4

DNA binding properties of the Mga*Spn* regulator

In this chapter, we have analysed the DNA binding properties of the Mga*Spn* regulatory protein by gel retardation, footprinting and electron microscopy techniques. Electron microscopy experiments were carried out at the Max-Planck Institute für molekulare Genetik (Berlín) under the supervision of Dr. Rudi Lurz. We published this study in 2013 (Solano-Collado *et al*., 2013; see Related publications) and, for the first time, we showed that a member of the Mga/AtxA family of transcriptional regulators is able to generate multimeric complexes on linear double-stranded DNAs.

4.1. Defining the optimal DNA-binding conditions of Mga*Spn*

To study the binding of Mga*Spn* to linear double-stranded DNAs, we used firstly a 222-bp DNA fragment (coordinates 1598298 to 1598519), which contains the *P1623B* and *Pmga* divergent promoters, as well as the *PB* activation region (coordinates 1598388 to 1598457) (Figure 25). This region is required for Mga*Spn*-mediated activation of the *P1623B* promoter *in vivo* (Results, Chapter 3) (Solano-Collado *et al*., 2012). Preliminary EMSA experiments were performed using as binding conditions those described in Methods (Section 11.1.1). First of all, we determined the time required for the binding reaction to reach equilibrium. The ^{32}P-labelled 222-bp DNA (1 nM) was mixed with Mga*Spn* (25 or 50 nM) in the absence of competitor DNA, and then, after different incubation times (1 to 9 min), reaction mixtures were loaded onto native polyacrylamide (5%) gels. Since binding and running buffers were different, and to rule out a possible re-equilibration of the sample, reaction mixtures were loaded onto the gel while the current was turned on (200 V). Under these conditions, the loading process itself acts to quench the association reaction, since the free DNA begins to be separated from the complexes as soon as the sample is loaded onto the running gel (Carey, 1988). When the last sample entered the gel, the voltage was reduced to 100 V. As shown in Figure 26, at 25 nM of Mga*Spn*, free DNA and four protein-DNA complexes were already detected at 1 min. The number of complexes observed did not vary at longer incubation times, indicating that reaction mixtures already had reached equilibrium at 1 min. When the protein concentration used was 50 nM, the time required for equilibrium to be reached was the same, although up to seven protein-DNA complexes were detected (not shown).

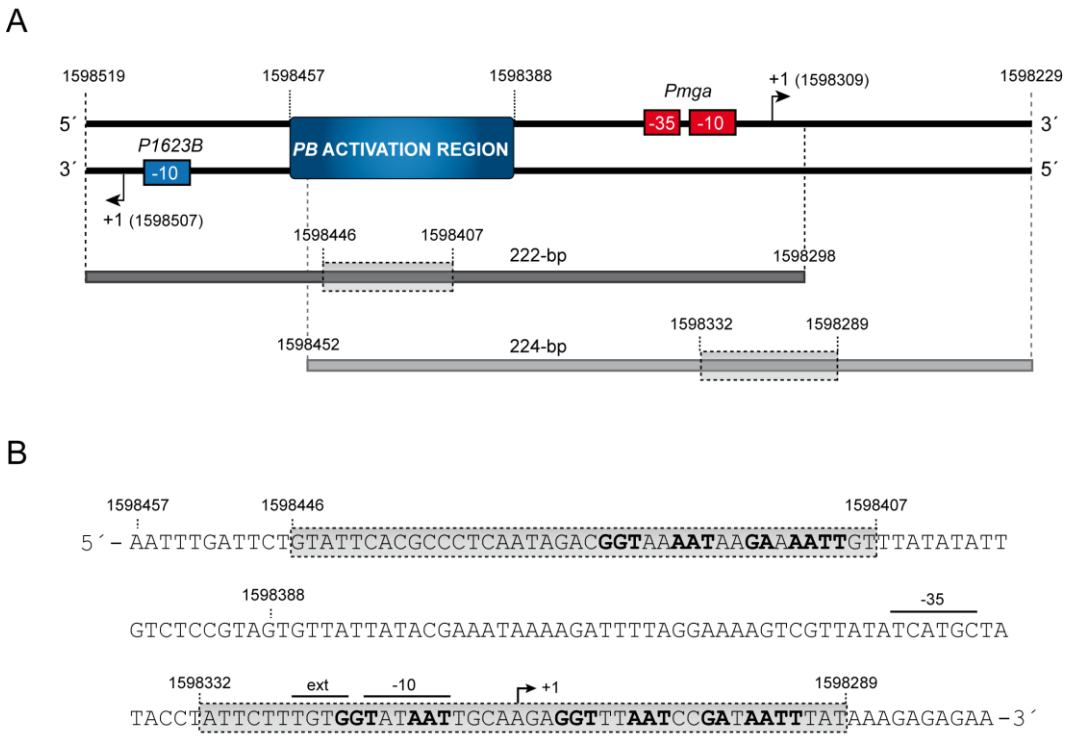


Figure 25. Scheme showing the relevant features of the 222-bp and 224-bp DNA fragments. (**A**) Region spanning coordinates 1598229 and 1598519 of the pneumococcal R6 genome. This region includes the *P1623B* and *Pmga* divergent promoters, as well as the *PB* activation region (Solano-Collado *et al.*, 2012). The transcription initiation site (+1) for each promoter is indicated with an arrow. Location of the 222-bp and 224-bp DNA fragments is indicated. The shadowed box on each DNA fragment indicates the primary Mga*Spn* binding site defined by hydroxyl radical footprinting assays in this work. Coordinates of the main elements are indicated. (**B**) Nucleotide sequence of the region spanning coordinates 1598279 and 1598457. It includes the two primary binding sites of Mga*Spn* (shadowed boxes). The two sequence elements (**GGT**(A/T)(A/T)**AAT** and **GA**(A/T)**AATT**) shared by both sites, are indicated. The main elements of the *Pmga* promoter are marked.

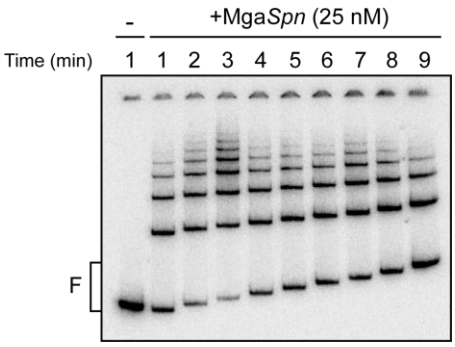


Figure 26. Time-course formation of Mga*Spn*-DNA complexes. The ^{32}P-labelled 222-bp DNA (1 nM) was mixed with Mga*Spn* (25 nM) in the absence of competitor DNA. At different time intervals (1 to 9 min), samples were loaded onto the running gel. Free (F) and bound DNA bands were visualized using a Fujifilm Image Analyzer (FLA-3000).

Since the interaction between a protein and its target DNA can be affected by salt concentration, we also analysed by EMSA the effect of NaCl concentration on the binding reaction. In this experiment, Mga*Spn* (50 nM) was incubated with the non-

labelled 222-bp DNA (10 nM) in the presence of different NaCl concentrations (from 20 to 300 mM). Reaction mixtures were incubated for 20 min. The results obtained are shown in Figure 27. Independently of the NaCl concentration, similar amounts of the C1, C2 and C3 complexes were formed. Thus, binding of Mga*Spn* to DNA is not affected by salt concentration in the range tested.

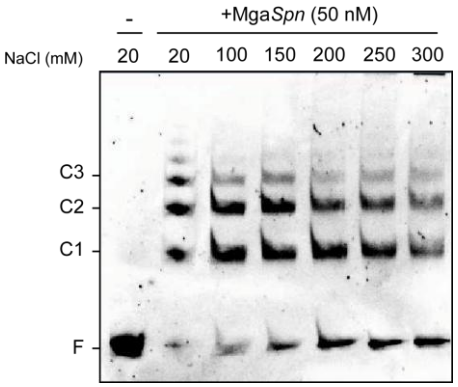


Figure 27. Effect of NaCl concentration on the formation of Mga*Spn*-DNA complexes. Non-labelled 222-bp DNA (10 nM) was incubated with Mga*Spn* (50 nM). Each binding reaction contained a particular concentration of NaCl (20 to 300 nM). Samples were subjected to electrophoresis on native polyacrylamide (5%) gels. Bands were stained with ethidium bromide (1 μg/ml) and visualized using a Gel-doc system (Bio-Rad). Free DNA (F) and Mga*Spn*-DNA complexes (C1, C2 and C3) are indicated.

4.2. Mga*Spn* binds to DNA forming multimeric complexes

To further analyse the interaction of Mga*Spn* with linear double-stranded DNA, the 222-bp DNA fragment was ^{32}P-labelled at the 5´-end of the coding strand, and was incubated with increasing concentrations of Mga*Spn* (20 to 480 nM) in the presence of non-labelled competitor calf thymus DNA (2 μg/ml) (Figure 28A). Free and bound DNAs were separated by electrophoresis on native polyacrylamide (5%) gels, and visualized by autoradiography. As it is shown in Figure 28A, at 20 nM of Mga*Spn*, we were able to visualize two protein-DNA complexes (C1 and C2) as well as free DNA. Interestingly, as the protein concentration was increased, higher-order complexes appeared sequentially whereas complexes with greater mobility were observed increasingly less. Moreover, protein-DNA complexes moving slower than complex C1 were detected before the total disappearance of unbound DNA, suggesting that the first site of interaction was not saturated prior to formation of higher-order complexes. This pattern of complexes is compatible with the formation of regular multimeric complexes, in which multiple protein units bind orderly on the same DNA molecule. To analyse whether cooperativity exists in the formation of such complexes, bands of free and bound DNAs were quantified with the Quantity One software. The percentage of each protein-DNA complex was calculated and plotted against Mga*Spn* concentration (Figure 28B). The curves obtained suggested

that several Mga*Spn* units interact with the DNA molecule in an apparently non-cooperative manner, at least in the formation of the first four complexes.

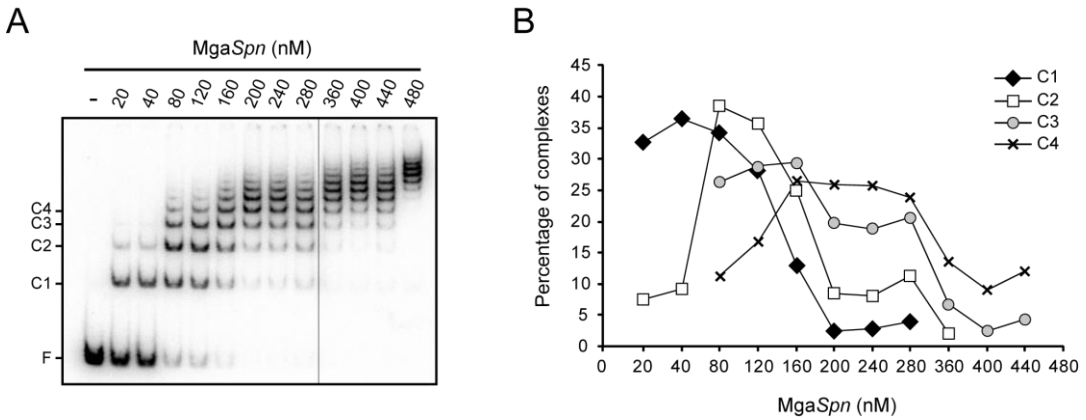


Figure 28. Formation of multimeric Mga*Spn*-DNA complexes. (**A**) EMSA analysis of the Mga*Spn*-DNA complexes. The ^{32}P-labelled 222-bp DNA fragment (2 nM) was incubated with increasing concentrations of Mga*Spn* in the presence of non-labelled competitor calf thymus DNA (2 μg/ml). Reactions were loaded onto a native gel (5% polyacrylamide). All the lanes displayed came from the same gel. Bands corresponding to free DNA (F) and to several Mga*Spn*-DNA complexes (C1, C2, C3 and C4) are indicated. (**B**) The autoradiograph shown in A was scanned and the percentage of the indicated complexes was plotted against the concentration of Mga*Spn*.

Additionally, we performed dissociation experiments (Figure 29). Essentially, Mga*Spn* (160 nM) was incubated with 2 nM of the ^{32}P-labelled 222-bp DNA for 20 min. Once the complexes were formed, different amounts of non-labelled competitor calf thymus DNA (0.5 to 10 μg/ml) were added to the binding reactions, which were incubated for a further 5 min under the same conditions. Finally, reaction mixtures were loaded onto native polyacrylamide (5%) gels. The results obtained showed that, as the amount of competitor DNA is increased, higher-order complexes gradually disappear, while moving faster complexes appear. Therefore, the protein units dissociate sequentially from the higher-order complexes.

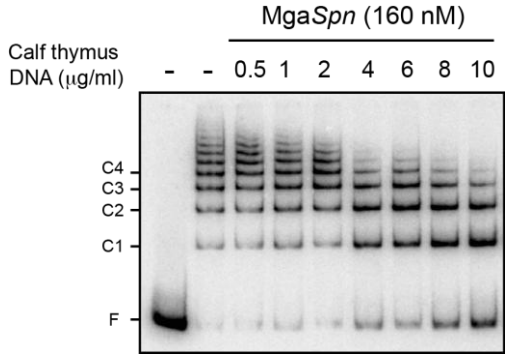


Figure 29. Dissociation of Mga*Spn*-DNA high-order complexes. EMSA was carried out incubating Mga*Spn* (160 nM) and 2 nM of ^{32}P-labelled 222-bp DNA. The indicated amount of non-labelled calf thymus DNA was added to pre-formed Mga*Spn*-DNA complexes. Radioactive bands were visualized using a Fujifilm Image Analyzer (FLA-3000).

By EMSA, we also estimated the affinity of Mga*Spn* for the 222-bp DNA. In this case, the concentration of the ^{32}P-labelled DNA was reduced to 0.1 nM and the Mga*Spn* concentration varied from 1 to 130 nM. Binding reactions were incubated for 20 min. Free and bound DNA were visualized using a Phosphorimager system, and the intensity of the bands was quantified using the Quantity One software. The protein concentration required to bind half the DNA was determined by measuring the decrease in free DNA rather than the increase in complexes (Figure 30), giving an indication of the approximate magnitude of the dissociation constant, K_d (Carey, 1988). With this approach, a K_d value of 50 nM was calculated. However, this value underestimates the affinity of Mga*Spn* for its first binding site on the 222-bp DNA, since, upon binding to this site, additional protein units bind sequentially to the same DNA molecule.

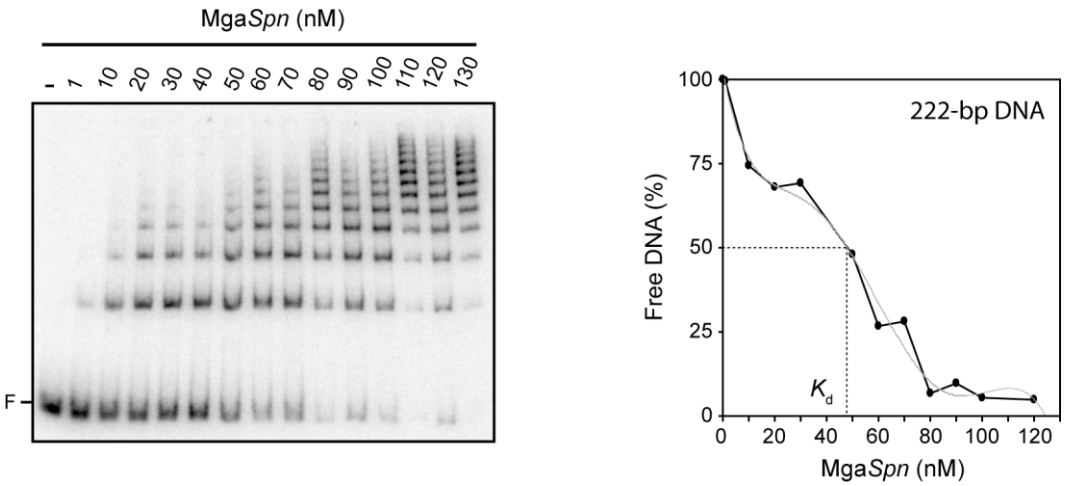


Figure 30. Affinity of Mga*Spn* for the 222-bp DNA. EMSA of the ^{32}P-labelled 222-bp DNA (0.1 nM) incubated with increasing concentrations of Mga*Spn* (1 to 130 nM) in the absence of competitor DNA (left panel). Binding mixtures were separated by native gel electrophoresis. The percentage of free DNA was calculated and plotted against Mga*Spn* concentration (right panel). From the amounts of unbound DNA the dissociation constant (K_d) was calculated (Carey, 1988).

4.3. Binding of Mga*Spn*-His and Mga*Spn* to the *PB* activation region

By *in vivo* experiments (Results, Chapter 3), we demonstrated that the *PB* activation region (Figure 17) is required for Mga*Spn*-mediated activation of the *P1623B* promoter. To study whether Mga*Spn* was able to interact with the *PB* activation region, we performed DNase I footprinting experiments. In a first approach, we used a His-tagged Mga*Spn* (Mga*Spn*-His) protein. This variant of Mga*Spn* carries six His residues at the C-terminal end and two additional amino acids (Leu and Glu) between the His-tag and the last amino acid of Mga*Spn*. Previously we had found by EMSA that Mga*Spn*-His

generated a pattern of protein-DNA complexes similar to that shown in Figure 28A. For the DNase I footprinting experiments, we used the 222-bp DNA fragment, which contains the *PB* activation region at internal position (at a distance of 62-bp and 90-bp from each DNA end, respectively) (Figure 25). The DNA fragment was radioactively labelled at the 5´-end of either the coding (Figure 31A) or the noncoding (Figure 31B) strand with respect to the *P1623B* promoter. Considering the ability of Mga*Spn*-His to form multimeric complexes, the binding reactions were performed in the presence of heparin to favour the formation of the complex C1 (Figure 32). As shown in Figure 31A, on the coding strand and in the presence of Mga*Spn*-His, a region from the position -52 to -90 relative to the transcription start site of the *P1623B* promoter was protected against DNase I digestion. In the case of the noncoding strand (Figure 31B), protections against DNase I digestion were observed from -57 to -79 and from -83 to -102. In addition, the -82 and the -104 positions were slightly more sensitive to DNase I cleavage. These results demonstrate that Mga*Spn*-His interacts with sequences located between the positions -52 and -102 with respect to the *P1623B* transcription start site (Figure 31C). Similar results were obtained using the untagged Mga*Spn* protein (not shown), indicating that the His-tag does not affect the DNA-binding activity of Mga*Spn*. Moreover, the site (coordinates 1598405 to 1598455) protected by both proteins (Mga*Spn*-His and Mga*Spn*) is included within the *PB* activation region (coordinates 1598388 to 1598457). Therefore, the Mga*Spn* regulator activates directly the *P1623B* promoter.

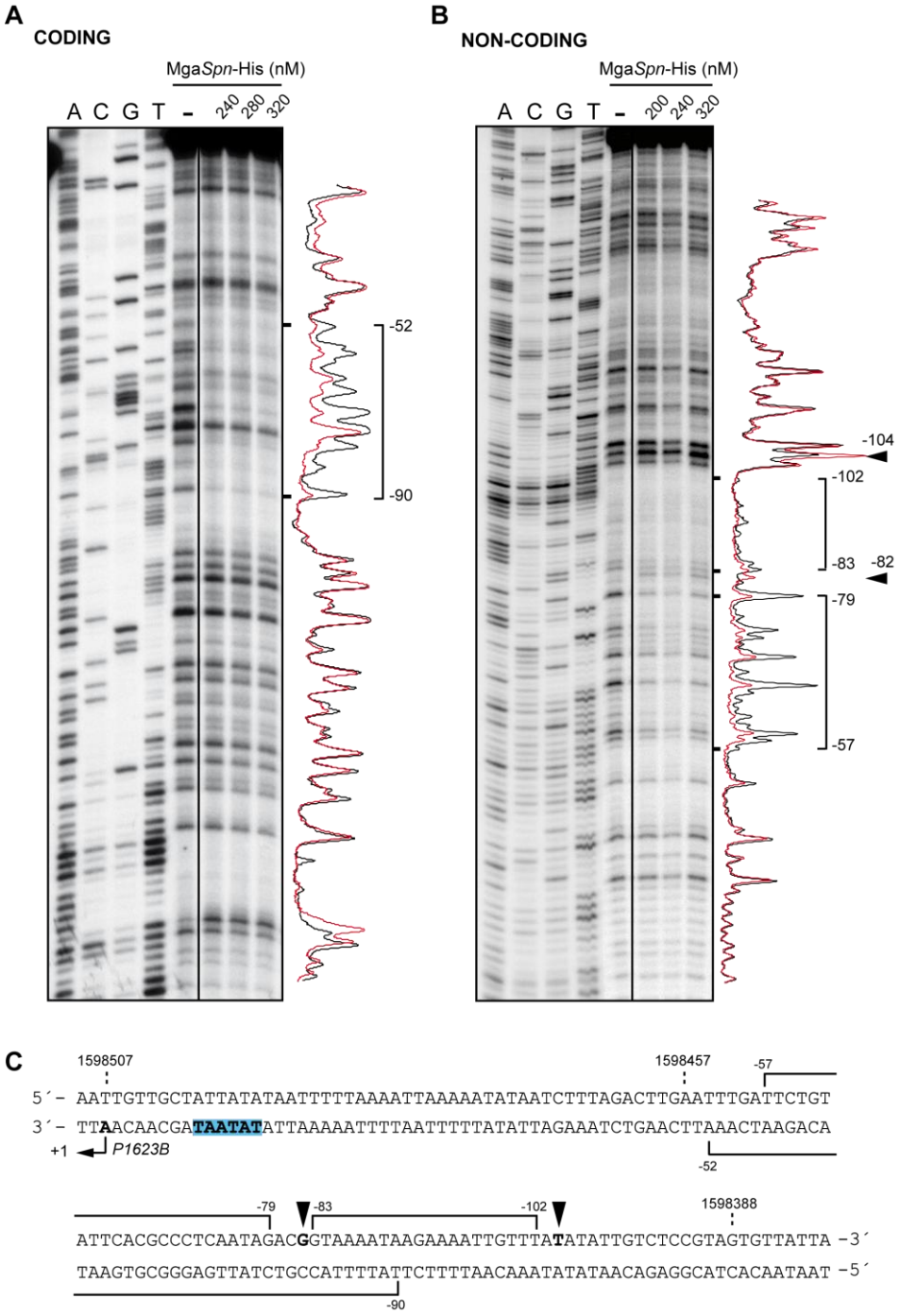


Figure 31. Analysis of the Mga*Spn*-His-DNA complexes formed in the presence of heparin by DNase I footprinting. The 222-bp DNA fragment (coordinates 1598298 to 1598519) was labelled at the 5´-end of either the coding (**A**) or the non-coding (**B**) strand. The labelled DNA (4 nM) was incubated with the indicated concentrations of Mga*Spn*-His in the presence of heparin (2 μg/ml) and the complexes formed were digested with DNase I. Dideoxy-mediated chain-termination sequencing reactions were run in the same gel (lanes A, C, G, and T). Densitometer scans corresponding to DNA without protein (black line) and DNA with protein (red line; 240 nM in panel A and 200 nM in panel B) are shown. The Mga*Spn*-His-protected regions are indicated with brackets. Arrowheads indicate positions that are slightly more sensitive to DNase I cleavage. Positions are relative to the transcription initiation site of the *P1623B* promoter. All the lanes displayed came from the same gel. (**C**) Nucleotide sequence of the region that spans coordinates 1598509 to 1598380 of the R6 genome. It includes the transcription start site of the *P1623B* promoter (coordinate 1598507), the region required for Mga*Spn*-mediated activation of the *P1623B* promoter (1598457 to 1598388), and the site recognized by Mga*Spn*-His (brackets).

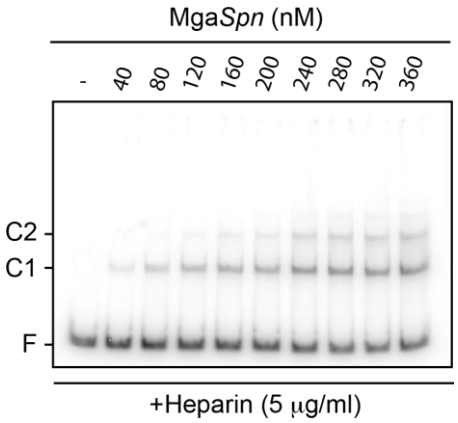


Figure 32. Mga*Spn*-DNA complexes formed in the presence of heparin. The 222-bp DNA (2 nM) was radioactively labelled and incubated with increasing concentrations of Mga*Spn* (40 to 360 nM) in the presence of heparin (5 μg/ml). Bands corresponding to free DNA (F) and to complexes C1 and C2 are indicated on the left of the gel.

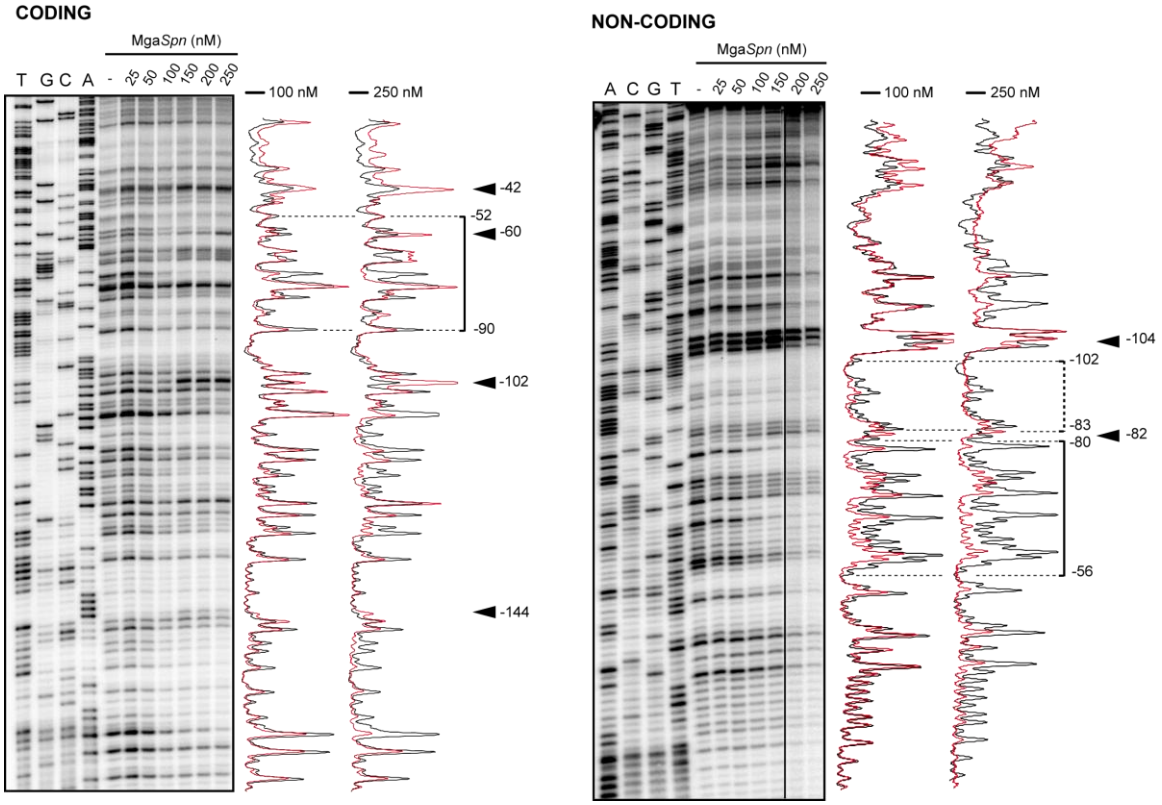


Figure 33. DNase I footprints of complexes formed by Mga*Spn* on the 222-bp DNA fragment. Coding and noncoding strands relative to the *P1623B* promoter were ^{32}P-labelled at the 5′-end. In this assay, the concentration of DNA was 4 nM and no competitor was used. All the lanes displayed came from the same gel. Densitometer scans corresponding to DNA without protein (black line) and DNA with the indicated concentration of protein (red line) are shown. The regions protected at 100 nM of Mga*Spn* are indicated with brackets. Arrowheads indicate positions that are slightly more sensitive to DNase I cleavage. The indicated positions are relative to the transcription start site of the *P1623B* promoter. Dideoxy-mediated chain termination sequencing reactions were run in the same gel (lanes A, C, G, T).

Binding of Mga*Spn* to the 222-bp DNA fragment was also analysed by DNase I footprinting assays in the absence of heparin (Figure 33). On the non-coding strand and at 100 nM of protein, Mga*Spn* recognised preferentially a site located between the positions -56 and -102 relative to the *P1623B* transcription start site. Such a primary binding site (coordinates 1598405 to 1598451) is located within the *PB* activation region. However, at higher protein concentrations (200 to 250 nM), regions flanking the primary binding site were also protected against DNase I digestion. On the coding strand and at 250 nM of Mga*Spn*, protected regions and hypersensitive sites (-144, -102, -60 and -42) were observed along the DNA fragment (Figure 33A). These results indicate that, Mga*Spn* does not bind randomly to the 222-bp DNA. It is able to bind preferentially to the *PB* activation region (primary binding site) and subsequently to spread along the adjacent DNA regions. This result is consistent with the pattern of protein-DNA complexes observed by EMSA (Figure 28A).

The specific contacts established by Mga*Spn* with its primary binding site on the 222-bp DNA were further analysed by hydroxyl radical footprinting assays (see Methods, Section 10.2). Once more, the 222-bp DNA was radioactively labelled at either the coding or the noncoding strand relative to the *P1623B* promoter. The binding reactions were performed in the presence of calf thymus DNA to favour the formation of the complex C1 (see Figures 28A and 29). To increase the footprint signal, after treatment with the hydroxyl radical cleavage reagent, the Mga*Spn*-DNA complex C1 was separated from the unbound DNA by electrophoresis on a native polyacrylamide gel (see Figure 28A). DNA from complex C1 and unbound DNA were excised from the gel, eluted from the acrylamide and loaded onto a sequencing gel (Figure 34A). On the coding strand, regions of decreased cutting were observed between the -60 and -96 positions relative to the transcription start site of the *P1623B* promoter. In the case of the noncoding strand, the decrease in hydroxyl radical cleavage occurred at regions located between positions -65 and -99. Specifically, Mga*Spn* protected four regions on each DNA strand. Each protected region covered three to five nucleotides, and the individual protected regions were separated by six to eight nucleotides. Protected regions were represented on a B-DNA (10.5-bp/helix turn) double helix model. As shown in Figure 34B, Mga*Spn* interacted with one face of the DNA double helix (coordinates 1598407 to 159846). Moreover, these results confirm that Mga*Spn* recognises preferentially the *PB* activation region on the 222-bp DNA.

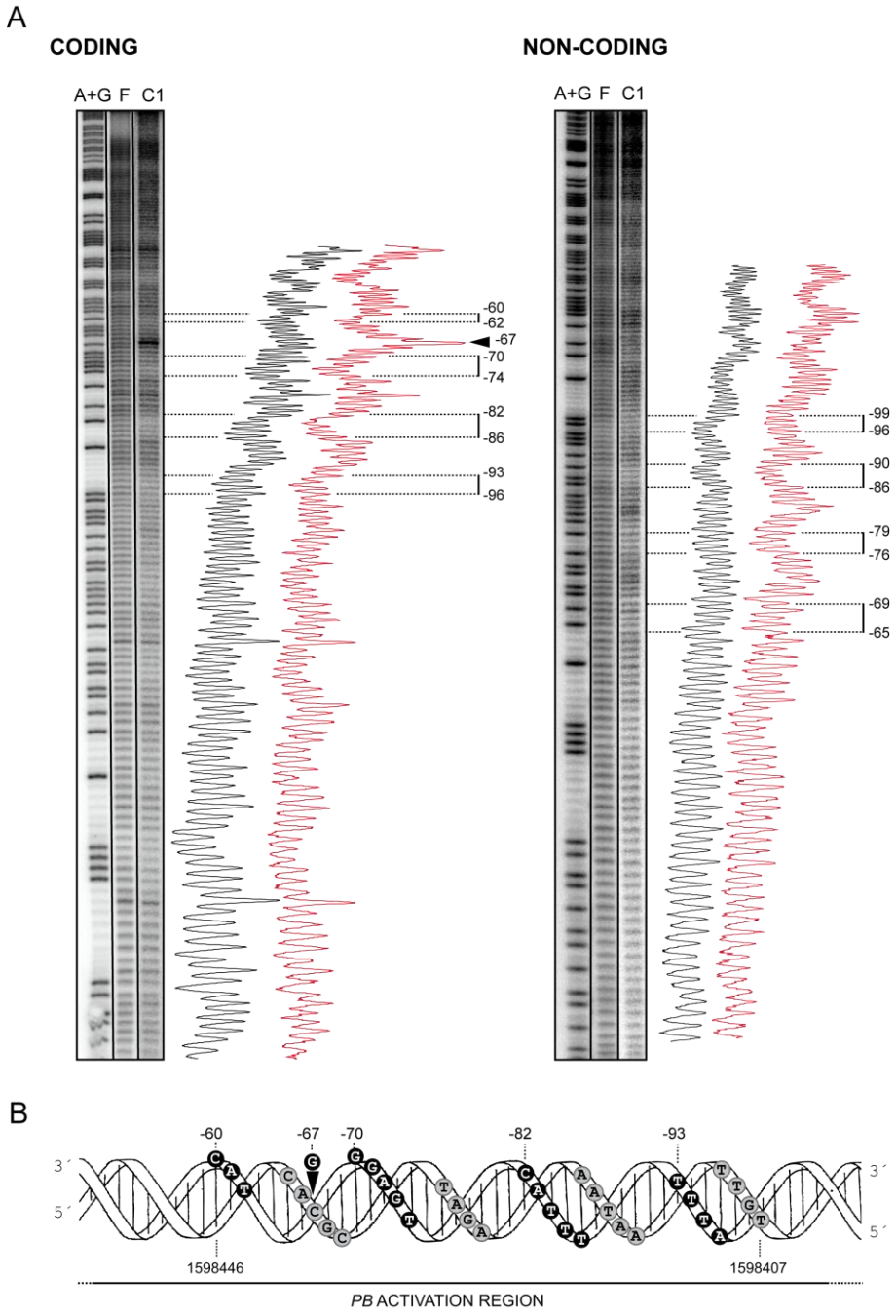


Figure 34. Mga*Spn* binds preferentially to the *PB* activation region on the 222-bp DNA. (**A**) Hydroxyl radical cleavage pattern of the 222-bp DNA without protein (F) and with Mga*Spn* bound to its primary site (complex C1). Coding and noncoding strands relative to the *P1623B* promoter were ^{32}P-labelled at the 5´-end and incubated with Mga*Spn* (DNA:protein molar ratio was 1:80). Densitometer scans from lanes F (black line) and C1 (red line) are shown. Numbers indicate positions relative to the transcription start site of the *P1623B* promoter. Regions protected by Mga*Spn* are indicated with brackets. The arrowhead indicates a position (-67) more sensitive to hydroxyl radical cleavage. All the lanes displayed came from the same gel. Lane C1 corresponds to a longer exposure time. A+G are products from Maxam-Gilbert adenine- and guanine-specific sequencing reactions performed on the respective labelled strands. (**B**) Schematic representation of the *PB* activation region on a B-DNA double helix. Mga*Spn* contacts, deduced from hydroxyl radical, are indicated as black (coding strand) and grey (noncoding strand) circles.

4.4. Mga*Spn* binds to double-stranded DNA with low sequence specificity

We decided to study whether Mga*Spn* was able to interact with DNAs either lacking the *PB* activation region or with it being positioned at one end of the DNA fragment. For this study, we used various linear DNA fragments, such as a 253-bp DNA from plasmid pUC19 (Yanisch-Perron *et al*., 1985), a 282-bp DNA from the coding region of the *mgaSpn* gene (1597232 to 1597513 of R6), and a 224-bp DNA (coordinates 1598229 to 1598452 of R6) which contains the *PB* activation region (positioned at one DNA end) and the *Pmga* promoter (Figure 25). In all cases (Figure 35), Mga*Spn* generated a pattern of complexes similar to that shown in Figure 28A (222-bp DNA fragment), which is compatible with the idea that multiple protein units bind to the same DNA molecule in an ordered fashion. Therefore, Mga*Spn* appears to bind to linear double-stranded DNAs with high affinity, but with low sequence specificity. An in-depth analysis of the interaction of Mga*Spn* with the 224-bp DNA fragment (Figure 25) supports this conclusion (see following sections).

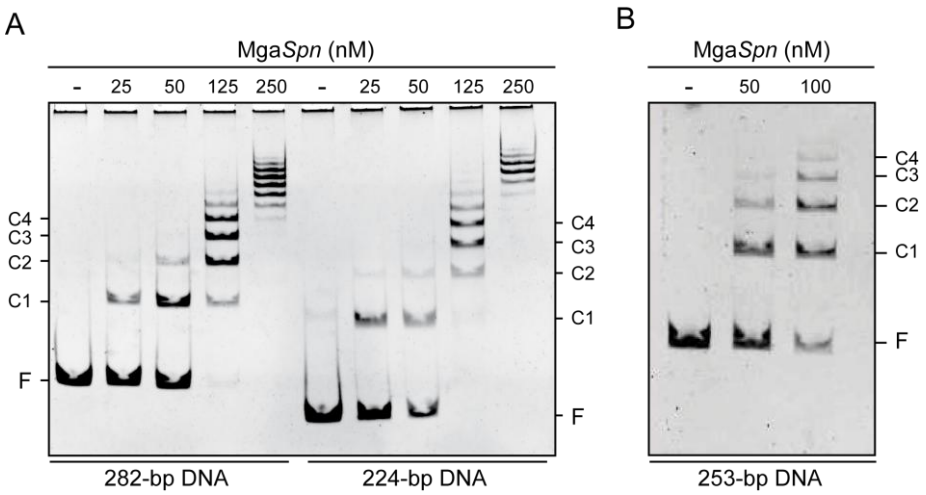


Figure 35. Mga*Spn* binds to linear dsDNAs with low sequence specificity. (**A**) EMSA analysis of the complexes formed by Mga*Spn* and two DNA fragments from the R6 chromosome. The 282-bp DNA came from an internal region of the *mgaSpn* gene (coordinates 1597232 to 1597513). The 224-bp DNA contains the *Pmga* promoter and the *PB* activation region positioned at one end of the molecule (coordinates 1598229 to 1598452) (Figure 25). The non-labelled DNAs (10 nM) were incubated with the indicated concentrations of Mga*Spn*. Bands corresponding to free (F) and several Mga*Spn*-DNA complexes (C1, C2, C3 and C4) are indicated. (**B**) Binding of Mga*Spn* to a dsDNA fragment (10 nM) from pUC19 plasmid. Free DNA (F) and Mga*Spn*-DNA complexes (C1, C2, C3 and C4) are indicated.

4.5. Binding of Mga*Spn* to the *Pmga* promoter region

Considering that Mga*Spn* binds to the 224-bp DNA (Figures 25 and 35), which includes the *Pmga* promoter but lacks the *P1623B* promoter, we also studied in detail the characteristics of this interaction. In the 224-bp DNA, the *PB* activation region is positioned at one DNA end (Figure 25). We first performed DNase I footprinting experiments labelling the 224-bp DNA at the 5´-end of either the coding or the noncoding strand relative to the *Pmga* promoter. Binding reactions contained 2 nM of ^{32}P-labelled DNA and increasing concentrations of Mga*Spn* (Figure 36). On the coding strand and at 40 nM of Mga*Spn*, sites more sensitive to DNase I digestion were observed along the DNA fragment (positions -68, -57, -30, -16, +13, +26 and +30 relative to the transcription start site of the *Pmga* promoter). These sites were spaced with a regular frequency of alternatively 10-13 and 26-27 nt. On the noncoding strand and at 40 nM of Mga*Spn*, the hypersensitive sites (positions +26 -15, -52 and -94) were spaced 36-41 nt. On both strands, the hypersensitive sites were flanking regions protected against DNase I digestion; a few unprotected sites were also observed. The pattern of protections and hypersensitive sites rules out that Mga*Spn* bind randomly to the 224-bp DNA and suggests a regular positioning of the protein along the DNA molecule. This is consistent with the idea of binding to a preferential site and then spreading (ordered positioning) along the adjacent DNA regions.

To identify the primary site recognised by Mga*Spn* on the 224-bp DNA, protein-DNA complexes were formed under conditions that favoured the formation of complex C1. These protein-DNA complexes were then subjected to hydroxyl radical treatment. To increase the footprint signal of complex C1, DNA complexed to Mga*Spn* was separated from free DNA before being loaded on a sequencing gel. The hydroxyl radical footprint pattern of Mga*Spn* bound to its primary site is shown in Figure 37A. On the coding strand, Mga*Spn* protected four regions within the sequence spanning the -21 and +21 positions relative to the *Pmga* transcription start site. Each protected region covered six to seven nt, and they were separated by five to six nt. On the noncoding strand, Mga*Spn* protected four regions of five nt within the sequence spanning the +18 and -23 positions. The protected regions were separated by six to eight nt. Sites more sensitive to hydroxyl radical cleavage were localized between the -25 and -34 positions. This fact indicates that the binding of Mga*Spn* to its primary site might induce a conformational change within the -35 element. Therefore, the primary site recognised by Mga*Spn* on the 224-bp DNA is the *Pmga* promoter region (Figure 37B). Specifically, it interacts with sequences that map between the -23 and +21 positions (coordinates 1598332 and 1598289). These results demonstrate that Mga*Spn* does not recognise the *PB* activation region as the

primary binding site when it is positioned at one end of the DNA molecule, but rather it shifts to the *Pmga* promoter.

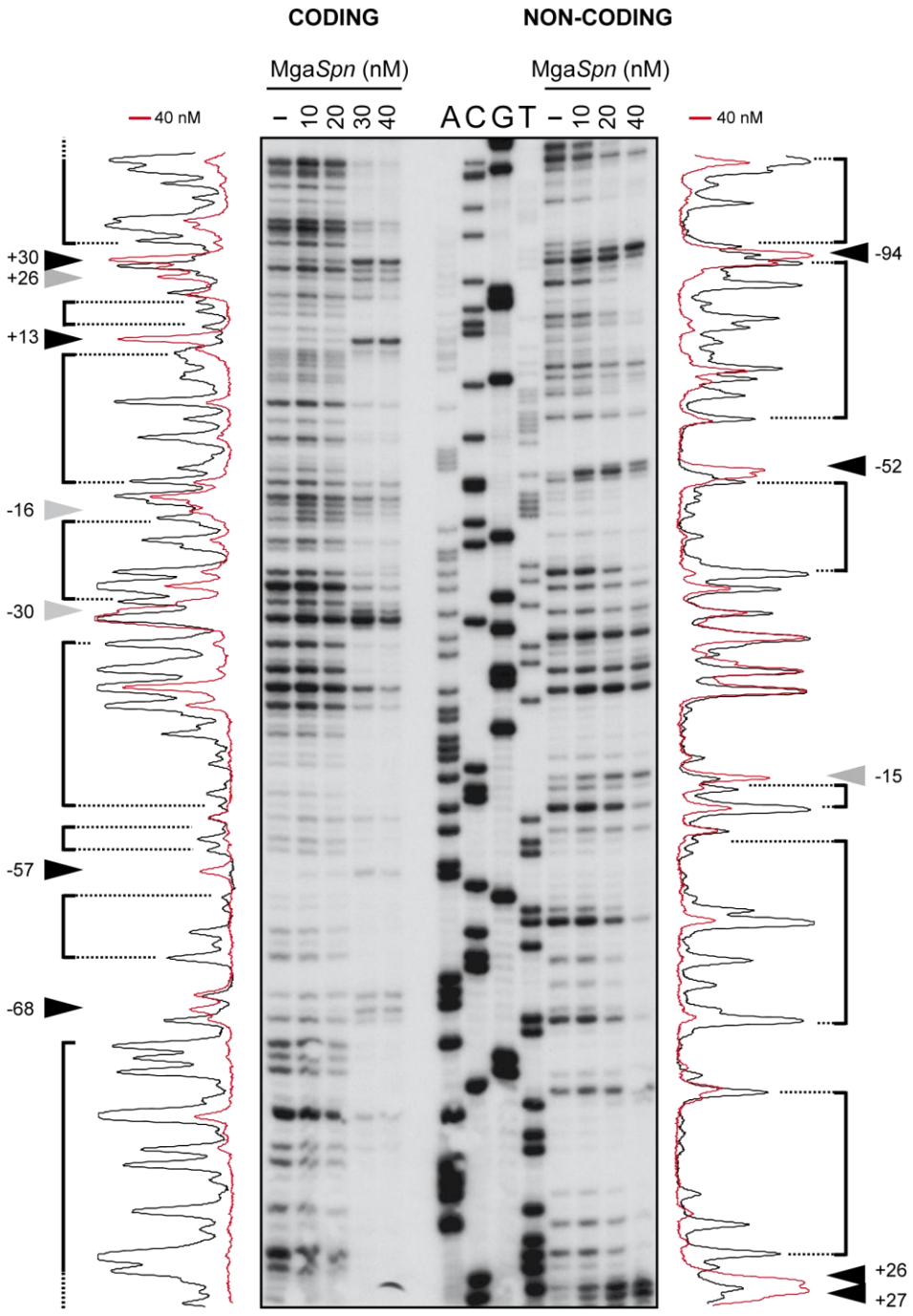


Figure 36. DNase I footprints of complexes formed by Mga*Spn* on the 224-bp DNA. Coding and noncoding strands relative to the *Pmga* promoter were ^{32}P-labelled at the 5´-end. In this assay, the concentration of DNA was 2 nM. Densitometer scans corresponding to free DNA (black line) and bound DNA (40 nM of Mga*Spn*; red line) are shown. Numbers indicate positions relative to the transcription start site of the *Pmga* promoter. Regions protected against DNase I digestion (brackets) and hypersensitive sites (arrowheads) are indicated. Dideoxy-mediated chain termination sequencing reactions on the non-coding strand were run in the same gel (A, C, G, T).

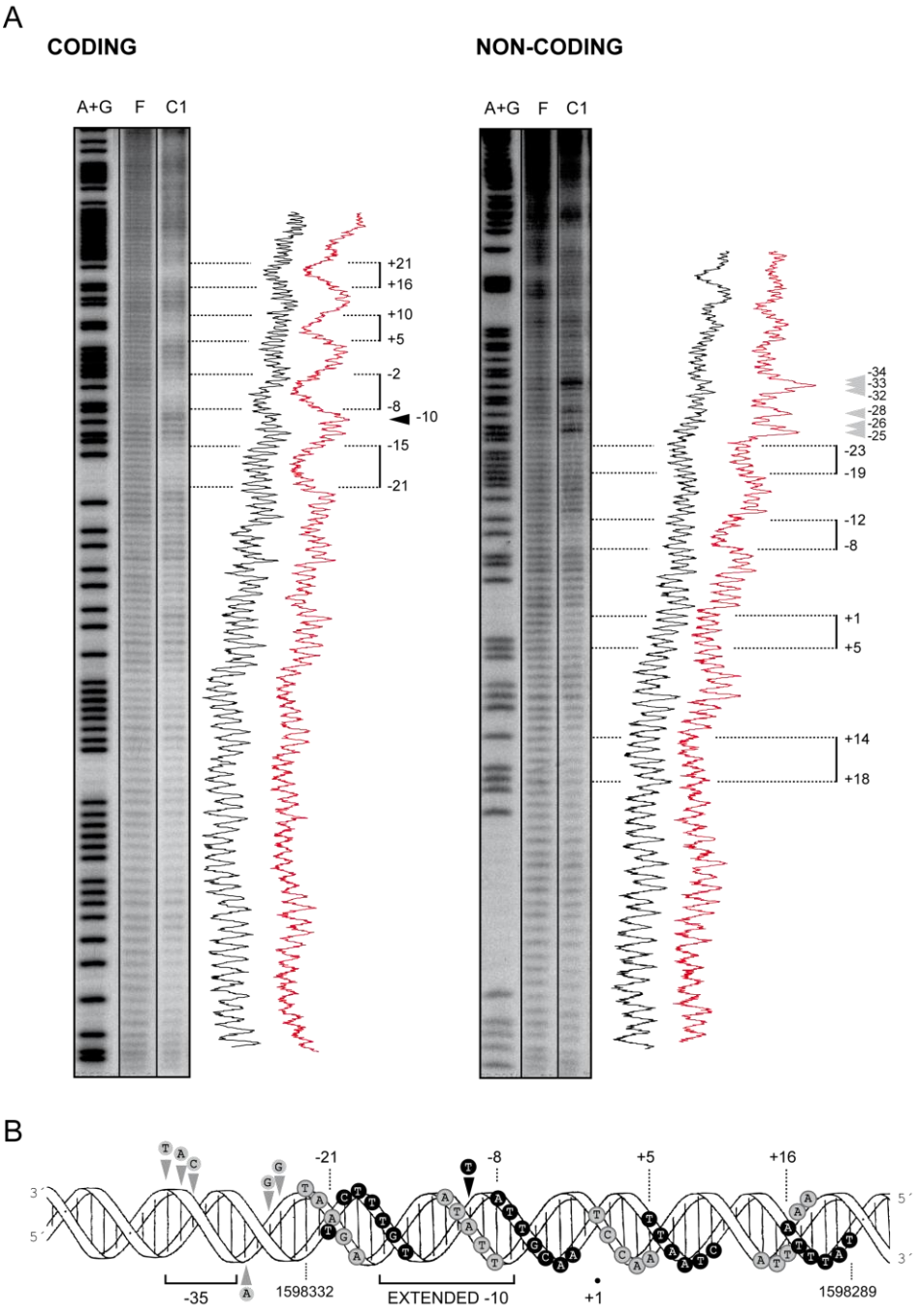


Figure 37. Mga*Spn* binds preferentially to the *Pmga* promoter region on the 224-bp DNA. (**A**) Hydroxyl radical digestion pattern of the 224-bp DNA without (F) and with Mga*Spn* bound to its primary site (complex C1). Coding and noncoding strands relative to the *Pmga* promoter were ^{32}P-labelled at the 5´-end and incubated with Mga*Spn* (DNA:protein molar ratio was 1:80). Densitometer scans from lanes F (black line) and C1 (red line) are shown. Numbers indicate positions relative to the transcription start site of the *Pmga* promoter. Regions protected by Mga*Spn* are indicated with brackets. Arrowheads indicate positions more sensitive to hydroxyl radical cleavage. All the lanes displayed came from the same gel. Lane C1 corresponds to a longer exposure time. A+G are products from Maxam-Gilbert sequencing reactions performed on the respective labelled strands. (**B**) B- form DNA of the *Pmga* promoter region showing Mga*Spn* contacts as deduced from hydroxyl radical. Black and grey circles indicate protein contacts on the coding and noncoding strand, respectively.

4.6. Analysis of the minimum DNA size required for Mga*Spn* binding

Hydroxyl radical experiments showed that, in the formation of complex C1, Mga*Spn* interacted with a region of 40-bp (primary binding site) on the 222-bp DNA (Figure 34B). Similar results were obtained with the 224-bp DNA (44-bp; Figure 37B). However, EMSA with both fragments showed the formation of up to 10 complexes. Therefore, to define the minimum DNA length necessary for Mga*Spn* binding, we performed EMSA experiments using small DNA fragments (from 20-bp to 40-bp) (Figure 38). The 40-bp fragment contained the -35 and -10 hexamers of the *Pmga* promoter, whereas the 26-bp and 32-bp DNAs contained only the -10 hexamer. The DNA fragments were incubated with increasing concentrations of Mga*Spn*. As shown in Figure 38, two Mga*Spn*-DNA complexes were detected with DNA fragments of 40-bp and 32-bp, whereas only one complex was observed with the 26-bp DNA. Similar results were obtained with DNA fragments that had the same size but different sequence (not shown). Moreover, no complexes were formed with the 20-bp DNA. Therefore, the minimum DNA size required for Mga*Spn* binding is between 20-bp and 26-bp.

In conclusion, since binding of Mga*Spn* to a 40-bp DNA fragment generates two complexes, and the minimum DNA size required for Mga*Spn* binding is between 20-bp and 26-bp, we propose that two Mga*Spn* units bound to the primary site of the 222-bp could constitute the C1 complex (40-bp, primary site). Sequential binding of additional Mga*Spn* units to complex C1 would then explain the formation of the additional complexes (C2 to C10) observed by EMSA (Figure 29).

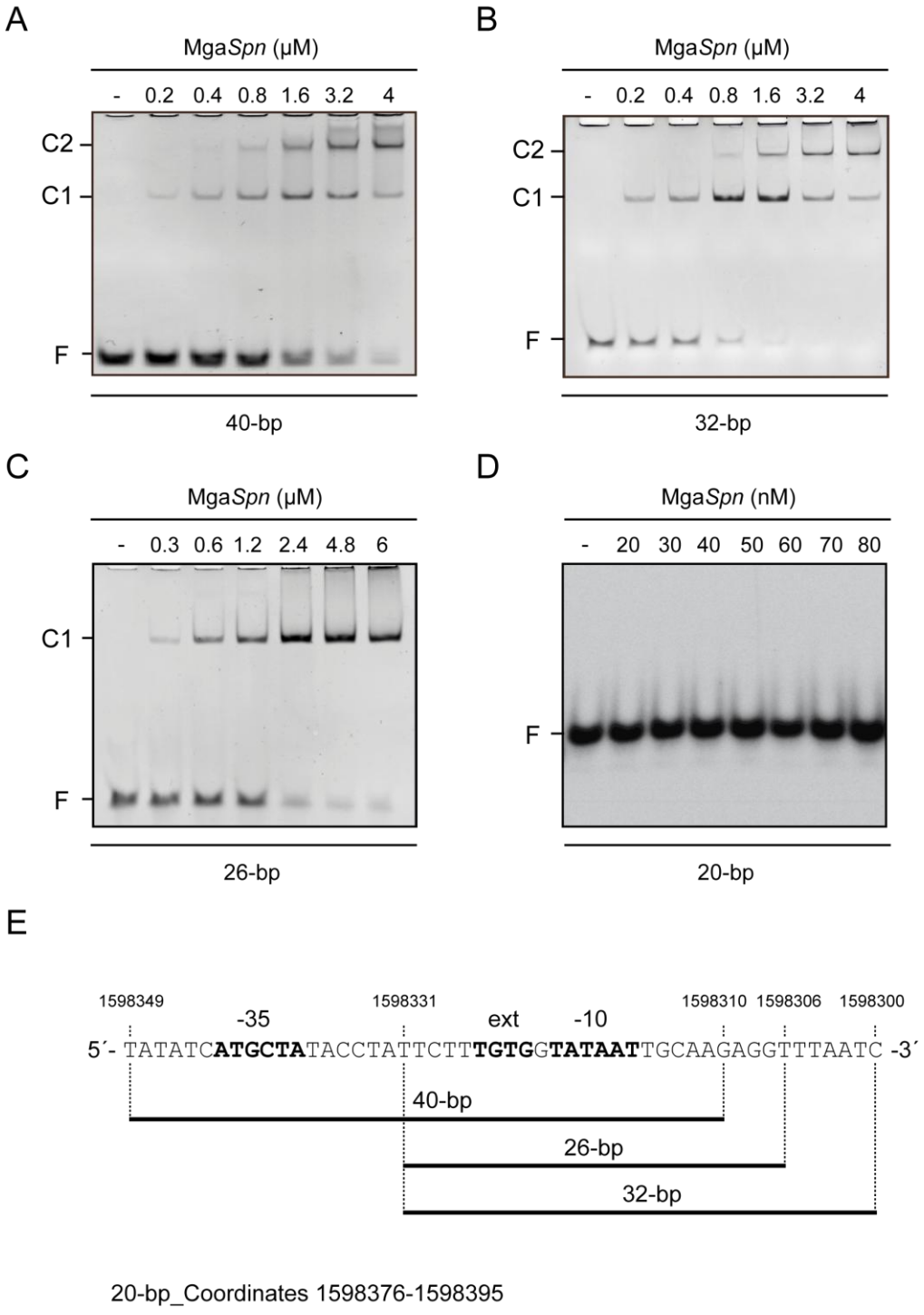


Figure 38. Binding of Mga*Spn* to small DNA fragments. EMSA analysis of the interaction of Mga*Spn* with DNA fragments of 40-bp (**A**), 32-bp (**B**), 26-bp (**C**) and 20-bp (**D**). Only the 20-bp DNA was radioactively labelled. The indicated concentration of Mga*Spn* was mixed with 200 nM of the 40-bp DNA, with 200 nM of the 32-bp DNA, with 300 nM of the 26-bp DNA or with 2 nM of the 20-bp DNA. Binding reactions were analysed by native gel electrophoresis. Bands corresponding to free DNA (F) and to Mga*Spn*-DNA complexes (C1 and C2) are indicated. (E) Nucleotide sequence of the oligonucleotides used to generate the 40-bp, 32-bp and 26-bp dsDNAs. The 20-bp DNA was from a region (coordinates 1598376-1598395) located upstream of the *Pmga* promoter. The main elements of the *Pmga* promoter are shown in bold.

4.7. Mga*Spn* binds to the *PB* activation region rather than to the *Pmga* promoter on long linear DNAs

The hydroxyl radical footprinting experiments performed with the 222-bp and 224-bp DNAs (Figures 34 and 37) revealed that the requirement for the recognition of the *PB* activation region as the primary binding site is its location internally within the DNA molecule. When it was positioned at one DNA end, the primary binding site recognised by Mga*Spn* shifted to the *Pmga* promoter region (see Figure 25). Since footprinting experiments are limited by the length of the DNA molecule, electron microscopy was used to further analyse the binding of Mga*Spn* to longer DNA molecules: 640-bp, 1418-bp and 1458-bp. The common feature of all these DNAs is the internal location of both *PB* activation region and the *Pmga* promoter. First, we performed the binding reactions at low protein/DNA ratios to favour the formation of complexes formed by one unit of Mga*Spn* bound to the DNA molecule. Electron micrographs of the complexes are shown in Figure 39A. We determined the Mga*Spn* binding position by measuring the DNA length in the protein-DNA complexes as well as the distance from the end of the DNA fragment to the protein-binding site. The distribution of the Mga*Spn* binding positions on each DNA fragment is represented in Figure 39A. On the 640-bp DNA (coordinates 1598010 to 1598649), of 171 complexes examined, the majority (80%) showed an Mga*Spn* unit bound to sequences located at a distance of 213-bp from one end of the DNA. Thus, Mga*Spn* bound preferentially either around coordinate 1598223 (Mga*Spn* coding region) or around 1598436 (*PB* activation region). On the 1418-bp DNA (coordinates 1597232 to 1598649), 62% of 156 complexes examined showed Mga*Spn* binding at a peak around 214-bp from the nearest DNA end. This result positioned Mga*Spn* around coordinate 1597446 (Mga*Spn* coding region) or 1598435 (*PB* activation region). On the 1458-bp DNA (coordinates 1598188 to 1599645), 63% of 182 complexes showed Mga*Spn* binding at a peak around 245-bp from one DNA end. Therefore, Mga*Spn* bound either around coordinate 1598433 (*PB* activation region) or 1599400 (downstream of the *P1623B* promoter). The same experiments were carried out using other DNA fragments, giving results (not shown) consistent with those already described. Taken collectively, the above results demonstrate that Mga*Spn* binds to the *PB* activation region rather than to the *Pmga* promoter when both sites are located at internal positions on the same DNA.

We next performed experiments at high protein/DNA ratios. Under these conditions, we detected DNA molecules totally or partially covered by Mga*Spn* (Figure 39B). The contour length of DNA molecules, both naked and covered with protein was measured to

find out whether the DNA was compacted. The results obtained show that their lengths do not vary as a result of protein binding, indicating that the DNA is not wrap around a core of protein molecules, and therefore compaction of the DNA does not occur. Overall, the electron microscopy results support that Mga*Spn* is able to spread along the DNA upon preferential binding to a particular site.

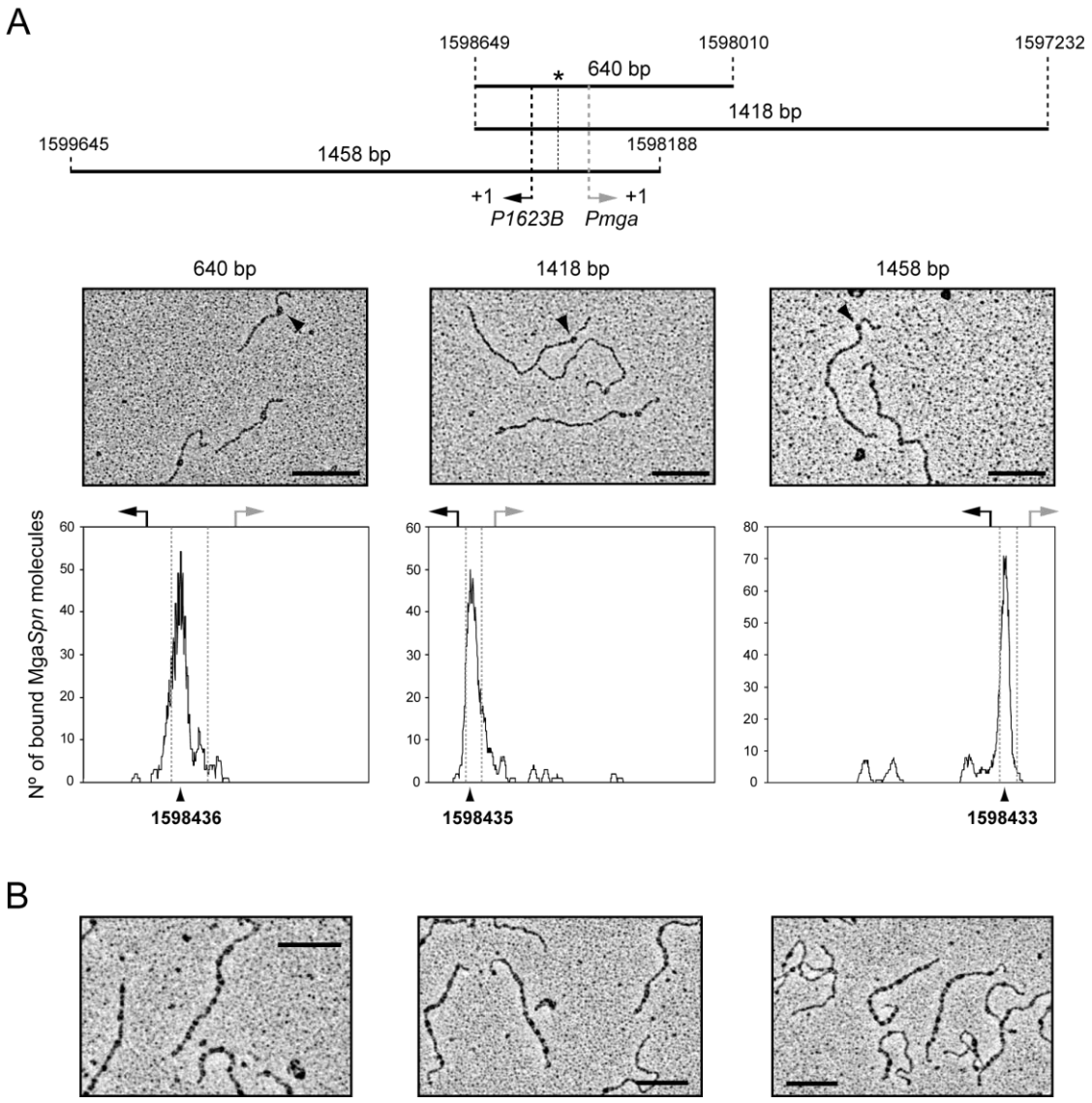


Figure 39. Electron microscopy analysis of Mga*Spn*-DNA complexes. (**A**) Mga*Spn* was incubated with different DNA fragments (640-bp, 1418-bp and 1458-bp) at low protein/DNA ratios to favour the formation of C1 complexes (Mga*Spn* bound to its primary site). The position of the *PB* activation region is indicated with an asterisk. Electron micrographs of the protein-DNA complexes are shown, and representative Mga*Spn*-DNA complexes are indicated with a black arrow. The distribution of the Mga*Spn* positions on the three DNA fragments is shown. Position of the transcription start site of the *P1623B* (black arrow) and *Pmga* (grey arrow) promoters is indicated. (**B**) Mga*Spn* was incubated with the 1418-bp DNA fragment at high protein/DNA ratios. Electron micrographs of DNA molecules covered by Mga*Spn* are shown. Scale bar, 500 bp.

4.8. Local DNA conformations might contribute to the DNA-binding specificity of Mga*Spn*

We have observed that the two sites recognised preferentially by Mga*Spn* (*PB* activation region and *Pmga* promoter) share a low sequence identity: the **GGT**(A/T)(A/T)**AATT** and **GA**(A/T)**AATT** sequence elements (Figure 25B). Nevertheless, despite this identity, Mga*Spn* interacted preferentially with the *PB* activation region rather than with the *Pmga* promoter when both elements were located at internal positions on long linear DNAs (see above, Section 4.7). Moreover, when the *PB* activation region was placed at one DNA end, Mga*Spn* recognised preferentially the *Pmga* promoter (Figure 25A). In order to investigate whether both primary sites shared common structural features, we calculated the curvature/propensity plots of both the 222-bp and the 224-bp DNA fragments with the bend.it server (Vlahovicek *et al*., 2003). In the case of the 222-bp DNA (Figure 40), the profile showed a peak of potential curvature at position 102 (included in the *PB* activation region; positions 73-112). The magnitude of the curvature propensity (9.5) is within the range calculated for experimentally tested curved motifs (Gabrielian *et al*., 1997). Flanking the peak of potential curvature, we found two regions of bendability (positions 74-88 and 122-134). In the case of the 224-bp DNA (Figure 40), the *Pmga* promoter region (positions 120-163) contains one peak of potential curvature (position 153; magnitude 9.1), which is also flanked by regions of bendability (positions 105-114 and 174-180). Therefore, the two sites recognised by Mga*Spn* contain a potential intrinsic curvature, which is surrounded by regions with the capacity of being easily bent.

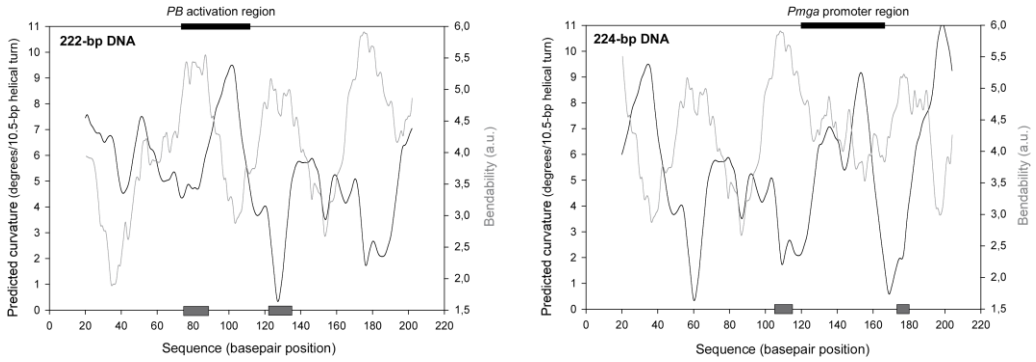


Figure 40. Bendability/curvature propensity plots of the 222-bp and 224-bp DNA fragments according to the bend.it program (Vlahovicek *et al*., 2003). The primary binding sites of Mga*Spn* are indicated (black boxes). Grey boxes indicate regions of bendability.

Subsequently, to analyse in greater detail whether Mga*Spn* interacted preferentially with intrinsically curved DNAs, we performed EMSA experiments using the curved (C) DNA and the non-curved (NC) DNA. Both DNAs had a similar A+T content. The C DNA fragment (321-bp; 72.3% A+T content) came from the *E. faecalis* V583 genome, while the NC DNA fragment (322-bp; 71.1% A+T content) was from the *S. pneumoniae* R6 genome. The latter fragment carried the sequence spanning the -22 to +299 positions relative to the *Pmga* transcription start site. According to *in silico* predictions of intrinsic DNA curvature using the bend.it server (Vlahovicek *et al*., 2003), the C DNA fragment showed a potential intrinsic curvature with a magnitude greater than 12 within the region spanning the positions 150 to 200. This magnitude of curvature is higher than the magnitude of the curvature predicted for the NC DNA fragment (Figure 41B). In addition, compared to the NC DNA fragment, the C DNA fragment showed an anomalously slow electrophoretic mobility on native polyacrylamide gels (Figure 41B), which is a characteristic of curved DNA fragments (Diekmann, 1987). This property is due to the higher friction of more curved molecules in the narrow pores of the polyacrylamide gel that slows down the rate of migration. We carried out competitive EMSA experiments using the radioactively labelled 222-bp DNA fragment. It was incubated with increasing concentrations of Mga*Spn* in the presence of non-labelled competitor DNAs, either C DNA or NC DNA (Figure 41A). At all protein concentrations, the amount of labelled DNA that remained unbound was higher when C DNA was used as a competitor (Figure 41C). These results gave us a first indication that Mga*Spn* was preferentially interacting with the intrinsically curved DNA.

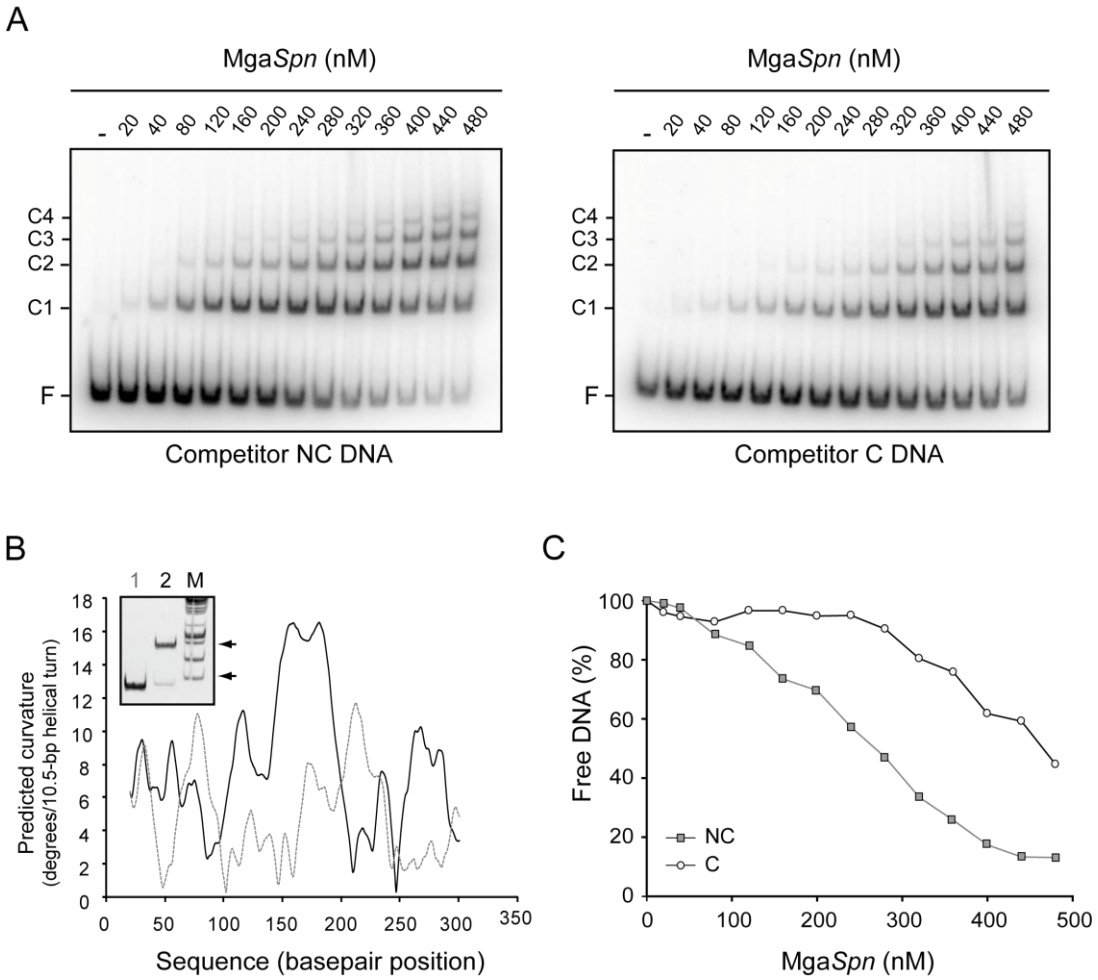


Figure 41. Binding of Mga*Spn* to a naturally occurring curved DNA. (**A**) Competitive EMSA. The ^{32}P-labelled 222-bp DNA fragment (2 nM) was incubated with the indicated concentration of Mga*Spn* in the presence of non-labelled competitor DNA (30 nM), either NC DNA (left) or C DNA (right). Bands corresponding to free DNA (F) and to several Mga*Spn*-DNA complexes (C1, C2, C3 and C4) are indicated. (**B**) Curvature-propensity plot of the NC DNA (grey line) and C DNA (black line). Inset shows the electrophoretic mobility of the NC DNA (lane 1) and C DNA (lane 2) on a native polyacrylamide (5%) gel. Lane M, DNA fragments used as molecular weight markers (HyperLadder I, Bioline). Arrows indicate the position of the 400-bp and 800-bp fragments. (**C**) The autoradiographs shown in A were scanned and the percentage of free DNA was plotted against the concentration of Mga*Spn*. NC DNA (grey squares) and C DNA (white circles) were used as competitors.

Moreover, by EMSA, we estimated the affinity of Mga*Spn* for both C DNA and NC DNA. Specifically, ^{32}P-labelled DNA (0.1 nM) was incubated with increasing concentrations of Mga*Spn* (1 to 130 nM). On both DNAs, Mga*Spn* generated multiple protein-DNA complexes (Figure 42). The DNA that remained unbound was quantified with the Quantity One software and plotted against Mga*Spn* concentration. The half-maximal binding point was determined by measuring the decrease in free DNA instead of the increase of the various complexes. For the C DNA (Figure 42A), the apparent K_d was around 12 nM, significantly lower than that obtained for the NC DNA (Figure 42B), which was around 95 nM. These results indicate that Mga*Spn* has higher affinity for the

naturally occurring curved DNA, and are in agreement with the results obtained from the competitive gel retardation assay (Figure 41).

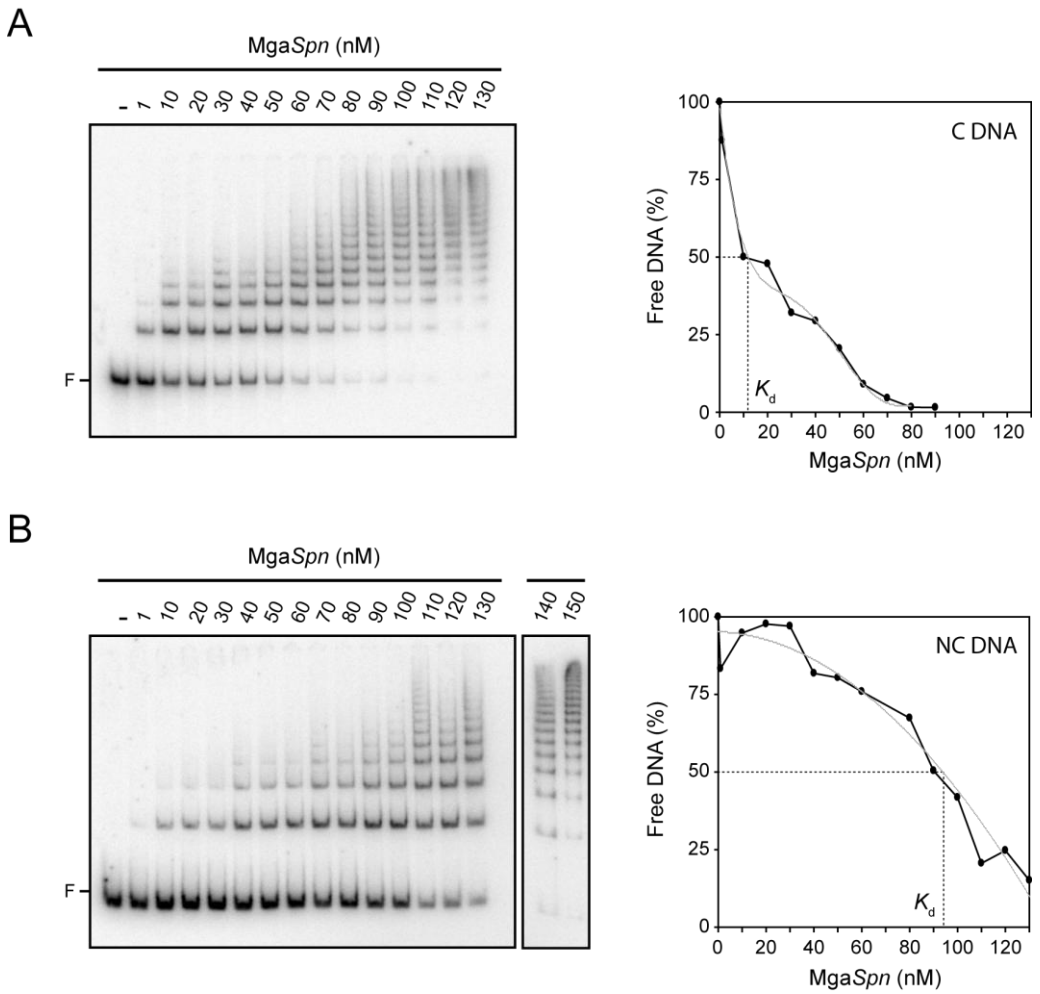


Figure 42. Affinity of Mga*Spn* for the C and NC DNAs. EMSA of either ^{32}P-labelled C DNA (0.1 nM) (A) or ^{32}P-labelled NC DNA (0.1 nM) (B) incubated with increasing concentrations of Mga*Spn* in the absence of competitor DNA (left panels). Bands corresponding to free (F) DNA are indicated. The percentage of free DNA was calculated and plotted against Mga*Spn* concentration (right panels). The apparent dissociation constant (K_d) was calculated by measuring the decrease in free DNA rather than the increase in complexes (Carey, 1988).

Discussion

Activator role of the Mga*Spn* virulence transcriptional regulator

Pathogenic bacteria must be able to survive to different conditions found during the diverse stages of infection. This survival depends on the ability to sense and respond to external stimuli via global transcriptional regulators whose activity and/or intracellular concentration change in response to environmental variations. In combination with TCSs, various *stand-alone* response regulators have been implicated in the regulation of virulence gene expression, as it is the case of the Mga global regulator of GAS, whose signal transduction components have yet to be fully defined (McIver, 2009). In 2002, signature-tagged mutagenesis in *S. pneumoniae* TIGR4 evidenced the existence of a putative virulence transcriptional regulator, which was an Mga orthologue encoded by the chromosomal *sp1800 (mgrA)* gene (Hava and Camilli, 2002; Hemsley *et al*., 2003). In GAS, the *mga* regulatory gene is located upstream of some of the genes it regulates. Nevertheless, in the pneumococcal TIGR4 strain, Hemsley *et al*. (2003) reported that the *mgrA* gene did not affect transcription of its neighbouring genes. These authors found that MgrA was able to repress the expression of the *rlrA* pathogenicity islet, a cluster comprising seven genes some of which had been previously characterized as virulence factors (Hava *et al*., 2003).

In *S. pneumoniae* R6, whose genomic sequence was published in 2001 (Hoskins *et al*., 2001), the *mgaSpn* gene is equivalent to the *mgrA* gene of the TIGR4 strain. In fact, this gene has been found in all the pneumococcal strains whose genomes have been totally or partially sequenced. In this work we have analysed 26 pneumococcal strains whose genomes have been totally sequenced: R6, D39, TIGR4, Hungary 19A-6, 670-6B, 70585, A026, AP200, ATCC 700669, CGSP14, G54, INV104, INV200, JJA, OXC141, P1031, PCS8235, SPN034156, SPN034183, SPN994038, SPN994039, SPNA45, ST556, TCH8431/19A, Taiwan19F-14 and gamPNI0373. According to protein sequence database similarity searches, Mga*Spn* (MgrA in TIGR4) is highly conserved among the above-mentioned pneumococcal strains. Compared to R6 and D39, Mga*Spn* has two amino acid changes (I309M and V358I) in the TIGR4, G54, JJA, ATCC 700669, P1031, TCH8431/19A, AP200, INV200, gamPNI0373, SPN034156, A026 and CGSP14 strains, three amino acid changes (C280Y, I309M, and V358I) in the Taiwan19F-14 and ST556 strains, three amino acid changes (I309M, V358I, and L362I) in the 670-6B and Hungary 19A-6, three amino acid changes (I309M, V358I, and P450S) in the 70585, PCS8235, SPN034183, OXC141, INV104, SPN994038 and SPN994039 strains and four amino acid changes (S86L, I309M, V358I, and P450S) in the SPNA45 strain. Unlike the *mgaSpn* gene, the *rlrA* pathogenicity islet has been found only in a small group of

pneumococcal strains (e. g. in TIGR4 but not in R6 and D39) (Paterson and Mitchell, 2006), suggesting that it may has been acquired by horizontal gene transfer, as it is flanked by two IS*1167* insertion sequences. This fact indicated that the *rlrA* islet might not be the main target of the pneumococcal Mga-like regulator.

In this work, we have performed an *in vivo* transcriptional analysis of the region spanning coordinates 1596789 to 1600589 of the pneumococcal R6 genome. This region includes the *mgaSpn* regulatory gene and four genes (*spr1623*, *spr1624*, *spr1625* and *spr1626*) of unknown function that are highly conserved in the TIGR4 strain (see below). We have identified the promoter of the *mgaSpn* gene (named *Pmga*) and its transcription initiation site (coordinate 1598308). In addition, we have shown that the ATG codon at coordinate 1598270 is likely the translation start site of the *mgaSpn* gene, as it is preceded by a consensus Shine-Dalgarno sequence. Translation from this ATG codon, which differs from the one annotated by the NCBI (Results, Chapter 1), generates a product of 493 amino acids.

Regarding the *spr1623*, *spr1624*, *spr1625* and *spr1626* genes, we have shown that they constitute an operon, which is transcribed from two promoters (*P1623A* and *P1623B*). Furthermore, we have demonstrated that, unlike the MgrA regulator of TIGR4 (Hemsley *et al*., 2003), Mga*Spn* is able to act, directly, as a positive regulator of the *spr1623-spr1626* operon. Mga*Spn* activates the *P1623B* promoter *in vivo*, and therefore the expression of the *spr1623-spr1626* operon, through direct interaction with sequences located upstream of the *P1623B* promoter (*PB* activation region). This discrepancy between the results presented here and the results previously obtained by Hemsley *et al.* (2003) might be due to the use of different pneumococcal strains and/or the use of different bacterial growth conditions. Actually, even though transcriptional regulators seem to be conserved among the majority of the *S. pneumoniae* strains, some of them modulate the transcriptional profile differently depending on the bacterial strain and/or serotype as well as the environmental conditions (Blue and Mitchell, 2003, Hendriksen *et al*., 2009), (Hendriksen *et al*., 2007).

According to the Protein Cluster database (Klimke *et al*., 2009), the *spr1623*, *spr1624*, *spr1625* and *spr1626* products are identical to those encoded by the 26 pneumococcal genomes mentioned above. However, except in R6 and D39, the operon in the other strains contains an additional ORF (*sp1801* in TIGR4). It is located upstream of the equivalent *spr1623* gene and it might encode a product of 54 residues (hypothetical protein, putative transglycosylase associated protein). Its absence in both R6 and D39 is due to the deletion of one nucleotide between coordinates 1598751 and

1598752, which results in a truncated ORF that might encode a product of 20 amino acids, which is not contemplated by the NCBI. The function of the genes that constitute the *spr1623-spr1626* operon is unknown. They encode products of 188 (*spr1623*; hypothetical protein), 56 (*spr1624*; putative lipoprotein), 202 (*spr1625*; putative general stress protein 24), and 67 (*spr1626*; hypothetical protein) residues. However, several observations suggest that the *spr1623-spr1626* operon might play a role in virulence. On the one hand, Hemsley *et al*. (2003) reported the characterization of two transposon insertion mutants in TIGR4. One of them (STM206) carries a transposon inserted ~300 bp upstream of the transcription start site of the *sp1800* gene (*mgaSpn* in R6). The authors demonstrated that this mutant strain was attenuated for both nasopharyngeal carriage and lung infection and it was much more affected in virulence than a mutant strain (AC1272) that carries a transposon inserted within the coding sequence of the *sp1800* gene. In this work we have shown that the distance between the translation start codon of the *mgaSpn* gene (coordinate 1598270) and the transcription initiation sites of the *spr1623-spr1626* operon is 323 (from the *P1623A* promoter; coordinate 1598592) and 238 nucleotides (from the *P1623B* promoter; coordinate 1598507), respectively. Hence, the transposon in strain STM206 could be affecting the expression of the *sp1801-sp1805* operon (*spr1623-spr1626* operon in R6). If this were the case, the attenuation phenotype of the STM206 mutant strain (Hemsley *et al*., 2003) would indicate an important role of the operon in pneumococcal virulence. In addition, the *spr1625* product (202 amino acids) has homology (69% similarity) to the product of the *E. faecalis gls24* gene (*EF0080* in strain V583; 186 amino acids). In *E. faecalis*, the *gls24* gene was shown to be a general stress-inducible gene involved in bile-salts resistance (Giard *et al*., 2000) and its importance in enterococcal virulence using a mouse peritonitis model was already reported (Teng *et al*., 2005). Moreover, in GAS, the levels of Gls24 were enhanced when cultures were grown in hyaluronic acid-enriched medium (which simulates an infection situation) (Zhang *et al*., 2007) and during murine soft tissue infection (Graham *et al*., 2006). Additionally, it has been recently reported a direct effect of Gls24 in GAS survival in whole-blood and in neutrophil resistance (Tsatsaronis *et al*., 2013). Further work has to be done to confirm the role of the *spr1623-spr1626* operon in pneumococcal virulence.

Mga*Spn* is a member of the Mga/AtxA family of global regulators

The pneumococcal Mga*Spn* regulator shares sequence homology with the Mga global regulator of GAS (42.6% of similarity and 21.4% of identity) and with the AtxA global regulator of *B. anthracis* (39.9% of similarity and 20.7% of identity). Although no structural data are available for Mga or AtxA, both regulators were shown to have a similar organization of known or predicted functional domains (Hondorp *et al*., 2013). According to several programs for sequence-based prediction of secondary structures (Results, Chapter 2), the Mga*Spn* regulatory protein has a high content of α-helices (57.4-66.3%). We confirmed this prediction by CD spectroscopic analyses (average of 55.3%). Moreover, using the Pfam (protein domain prediction) and Phyre2 (protein structure prediction) programs (Punta *et al*., 2012; Kelley and Sternberg, 2009), we have found that the organization of functional domains in Mga*Spn* is similar to that found in Mga and AtxA.

The N-terminal region of Mga*Spn* contains two putative HTH DNA binding motifs, the so-called HTH-Mga (residues 6-65) and Mga (residues 71-158). HTH motifs are also present within the N-terminal region of the Mga and AtxA regulators. At present, studies on the interaction of AtxA with DNA are not available. However, in the Mga regulator, the HTH-3 and HTH-4 motifs were shown to be involved in the interaction with DNA (McIver and Myles, 2002). Moreover, the small CMD-1 motif was shown to be required for Mga-dependent activation of some promoters (Vahling and McIver, 2006). Sequence alignment of the first 150 amino acids of the Mga and Mga*Spn* regulators reveals the existence of several conserved residues within the Mga*Spn* region that spans the positions 74 and 130 (Figure 43). Thus, the most conserved region includes the HTH-4 motif of Mga. These findings indicate that the N-terminal region of Mga*Spn* may be implicated in the recognition and binding to DNA.

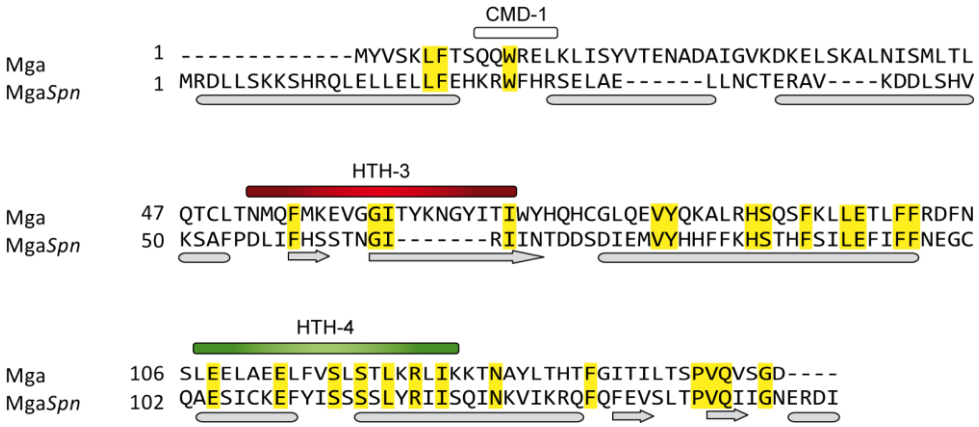


Figure 43. Sequence alignment of the N-terminal region of Mga and Mga*Spn* using the Clustal Omega program. In Mga, the three motifs (CMD-1, HTH-3 and HTH-4) involved in DNA binding are indicated (McIver and Myles, 2002; Vahling and McIver, 2006). Identical residues are shown in yellow. For Mga*Spn*, predicted secondary structure elements (SABLE program) are indicated (cilinders: α-helices; arrows: β-sheets).

Like Mga and AtxA, the central region of Mga*Spn* contains two putative PRDs, which are structural domains found in bacterial transcriptional regulators (antiterminators and activators) involved in the expression of catabolic operons. Such regulators are usually phosphorylated by components of the PTS on conserved histidine residues located within the PRDs. This phosphorylation is known to modulate the activity of the regulator (reviewed by Joyet *et al*., 2013). Amino acid sequence alignment of AtxA and other PRD-containing proteins led to the identification of two histidine residues (H199 in PRD1 and H379 in PRD2) as potential phosphorylation sites (Tsvetanova *et al*., 2007). By site-directed mutagenesis, histidine residues were mutated to alanine (mimicking an unphosphorylated state) or to aspartate (mimicking a phosphorylated state). These experiments indicated that the activity of AtxA is influenced by phosphorylation/dephosphorylation events within the PRDs. Indeed, phosphorylation of the conserved histidine (H199) in PRD1 increases AtxA activity, whereas phosphorylation of the conserved histidine (H379) in PRD2 decreases AtxA function (Tsvetanova *et al*., 2007). The Mga global regulator also contains two PRDs located at the central region. Hondorp *et al*. (2013) showed that phosphorylation of H204 and H270 (PRD1) inactivates Mga, whereas phosphorylation of H324 (PRD2) increases its function. Using the Clustal Omega program (Sievers *et al*., 2011), we have done an amino acid sequence alignment of the PRDs of Mga*Spn* with the PRDs of some transcriptional regulators (Figure 44). Such a multiple alignment includes MtlR (*Geobacillus stearothermophilus*) and LicR (*B. subtilis*), which are related to carbon utilization, and the AtxA and Mga virulence regulators (Figure 44). According to it,

several histidine residues of Mga*Spn* (H197 in PRD1 and H320 and H388 in PRD2) are potential sites to be phosphorylated by components of the PTS, suggesting that PTS-mediated phosphorylation of Mga*Spn* might modulate its activity in response to carbohydrate availability.

Figure 44. Sequence alignment of some PRDs using the Clustal Omega program. The PRDs of MtlR (*G. stearothermophilus*), LicR (*B. subtilis*), AtxA (*B. anthracis*), Mga (*S. pyogenes*) and Mga*Spn* are shown. Conserved histidine (yellow), arginine (blue) and aspartate (orange) amino acids are pointed out. The residues that link the PRD1 and PRD2 domains are indicated in grey. Residues that have been shown to increase or decrease the activity of the regulator upon phosphorylation are indicated in green or red, respectively.

In the Mga and AtxA regulators, a motif with structural homology to a region of the EIIB component used by the PTS (EIIB-like motif) has been found at their C-terminal region. A truncated Mga protein that lacks the last 139 amino acids (including the EIIB-like motif) was shown to be able to bind DNA with the same affinity as the wild-type protein. However it caused a drastic decrease in the transcript levels of two target genes, *arp* (M-like protein) and *sof* (fibronecting-binding protein). Moreover, co-immunoprecipitation, gel filtration chromatography and sedimentation equilibrium experiments showed that the truncated Mga protein was defective in oligomerization.

Based on those results, Hondorp *et al*. (2012) established a correlation between Mga-mediated transcriptional activation *in vivo* and the ability of Mga to form oligomers *in vitro*. In the case of AtxA, Hammerstrom *et al*. (2011) reported that AtxA exists, *in vivo*, in a homo-oligomeric state. Moreover, they found a correlation between AtxA dimerization and AtxA activity when bacteria were grown in specific conditions. They also showed that AtxA variants lacking the last 91 amino acids (includes the EIIB-like motif) are defective in the establishment of AtxA-AtxA interactions. Hence, in Mga and AtxA, the EIIB-like motif might play a role in protein-protein interactions. In addition, phosphorylation within the PRD1 domain of Mga appears to affects its oligomerization and its ability to regulate gene expression *in vivo* (Hondorp *et al*., 2013). In this work, we have demonstrated that, under our experimental conditions (100 mM and 250 mM NaCl), the untagged Mga*Spn* protein behaves as a dimer in solution. Moreover, it is able to form higher-order molecular species in a concentration-dependent manner. Interestingly, the Phyre2 protein fold recognition server predicts a putative EIIB-like motif at the C-terminal region of Mga*Spn*. Further work is required to elucidate whether the EIIB-like motif is involved in protein-protein interactions.

Collectively, the prediction of functional domains supports that the pneumococcal Mga*Spn* protein is a member of the Mga/Atxa family of global regulators. The predicted domain organization indicates that *(i)* Mga*Spn* binds DNA presumably through the N-terminal region, *(ii)* its activity might be modulated via phosphorylation of conserved histidine residues found in the putative PRDs, and *(iii)* the C-terminal region might be involved in its ability to dimerize and/or oligomerize.

Mga controls the expression of about 10% of the GAS genome (Ribardo and McIver, 2006) and AtxA influences the expression of more than a hundred genes located on the chromosome and the virulence plasmids of *B. anthracis* (Fouet, 2010). To date, the pneumococcal Mga*Spn* regulator is known to repress, directly or indirectly, the expression of several genes located within the *rlrA* pathogenicity islet (Hemsley *et al*., 2003) and to activate directly the expression of a four-gene operon (*spr1623-spr1626*) of unknown function (this work; Solano-Collado *et al*., 2012). Furthermore, we have identified a few putative Mga*Spn* target genes (metabolism of amino acids, competence and heat shock response) in a first proteomic analysis in which bacteria (producing or not Mga*Spn*) were grown under standard laboratory conditions (sucrose as carbon source). The low expression level of *mgaSpn* found in such conditions could explain the small number of putative Mga*Spn* target genes identified. In addition, under the conditions used, the Mga*Spn* regulator might not be fully active. Therefore, it would be

interesting to identify the specific external stimuli that influence the intracellular concentration and/or activity of the Mga*Spn* regulator. This knowledge is essential for the design of further proteomic and transcriptomic analyses and, consequently, for the definition of the Mga*Spn* regulon.

Mga*Spn* likely recognizes particular DNA conformations to achieve DNA-binding specificity

The recognition of specific DNA sequences and/or particular DNA structures by transcription factors is an essential step in the control of gene expression. In general, protein-DNA interactions can be divided in two main categories: those when the protein recognizes the unique chemical signatures of the DNA bases (base readout) and those when the protein recognizes a sequence-dependent DNA shape (shape readout). Nevertheless, it has been argued that any one DNA-binding protein is likely to use a combination of readout mechanisms to achieve DNA-binding specificity (Rohs *et al*., 2010). Regarding the Mga global transcriptional regulator of GAS, several binding sites have been identified. Although a consensus Mga-binding site was initially proposed (McIver *et al*., 1995), new sequence alignments of all established Mga-binding regions revealed a very low sequence identity (13.4%) (Hause and McIver, 2012). EMSA and DNase I footprinting experiments showed that Mga interacts with large non-palindromic DNA regions (45 to 59-bp). Unlike other many regulatory proteins, the regions found recognized by Mga did not present discernible symmetry (Hause and McIver, 2012). To determine the core nucleotides involved in functional Mga-DNA interactions, Hause and McIver (2012) carried out a mutational analysis in some target promoters and established that Mga binds to DNA in a promoter-specific manner. In the case of the AtxA regulator, two DNA-binding motifs have been predicted but no DNA-binding studies have been reported. Although sequence similarities in the promoter regions of AtxA-regulated genes are not apparent, a detailed analysis of the anthrax toxin genes showed a high proportion of A+T in the promoter regions, suggesting that they might exhibit intrinsic curvature. This DNA feature was further confirmed by *in silico* and *in vitro* studies (gel-retardation and circular permutation assays), supporting that AtxA may recognize a structural feature of the DNA rather than a specific sequence (Hadjifrangiskou and Koehler, 2008).

Our study (Results, Chapter 4) has shown that Mga*Spn* is able to interact with linear dsDNAs of different sources (e. g. the pneumococcal genome, pUC19 plasmid and the enterococcal genome). By gel-retardation assays, Mga*Spn* was shown to bind to any

tested DNA, generating a similar pattern of complexes; thus, it binds DNA with high affinity but with low sequence specificity. However, when the DNAs used carried the *PB* activation region at different positions (222-bp and 224-bp DNAs), DNase I and hydroxyl radical footprinting assays demonstrated that Mga*Spn* did not bind randomly at all. When the *PB* activation region was located at internal position on the DNA (222-bp DNA), Mga*Spn* recognized such a region as primary binding site (40-bp; positions -60 to -99 of the *P1623B* promoter). However, when the *PB* activation region was placed at one DNA end (224-bp DNA), Mga*Spn* interacted preferentially with the *Pmga* promoter region (44-bp; positions -23 to +21), which is located at internal position on this DNA fragment. We have found that both primary Mga*Spn*-binding sites, the *PB* activation region and the *Pmga* promoter region, share a common sequence element: **GGT**(A/T)(A/T)**AAT** and **GA**(A/T)**AATT** (see Chapter 4; Figure 25). Nevertheless, despite this identity, when both sites were positioned at internal position on the same DNA molecule (640 to 1458-bp), electron microscopy experiments showed that the *PB* activation region was the preferred target of Mga*Spn*. This observation suggested that Mga*Spn* might recognize a particular DNA conformation rather than a particular DNA sequence to establish specific DNA interactions.

We have analysed the two sites recognized preferentially by Mga*Spn* with the bend.it server (Vlahovicek *et al*., 2003). This analysis revealed that both sites contain a potential intrinsic curvature flanked by regions of bendability. Moreover, we have demonstrated that Mga*Spn* has a high affinity for a naturally occurring curved DNA. Our study suggests that local DNA conformations might contribute to the DNA-binding specificity of Mga*Spn.* Therefore, a general feature of the Mga/AtxA family of regulators might be their ability to bind DNA with little or no sequence specificity, as it is the case of many bacterial nucleoid-associated proteins that are able to influence transcription in either a positive or negative manner (Browning *et al*., 2010; Dillon and Dorman, 2010). We suggest that a preference for particular DNA structures might contribute to the capacity of the Mga/AtxA family of regulators to control the expression of a wide range of genes.

Formation of multimeric Mga*Spn*-DNA complexes

During this Thesis we have studied in detail the interaction of Mga*Spn* with linear double-stranded DNAs. An interesting finding during the progress of this work has been the ability of the untagged Mga*Spn* protein to generate multimeric protein-DNA complexes. Gel retardation assays with different DNAs showed that, before disappearance of the free DNA, additional protein units bound sequentially to the Mga*Spn*-DNA complex C1 generating higher-order complexes. The analysis of the interaction of Mga*Spn* with the 222-bp and 224-bp DNAs by DNase I and hydroxyl radical footprinting experiments demonstrated that Mga*Spn* interacted with a particular site (*PB* activation region or *Pmga* promoter) on both fragments. After this first event, Mga*Spn* was able to spread along the adjacent DNA regions and, therefore, to generate multimeric protein-DNA complexes. Considering that Mga*Spn* occupies a region of ~40-bp on both primary sites and the minimum DNA region necessary for Mga*Spn* binding has been found to be between 20 and 26-bp, we propose that two Mga*Spn* units might be bound in complex C1.

At present, formation of multimeric protein-DNA complexes has not been shown for other regulators of the Mga/AtxA family, and until 2011 very little was known about the biochemical properties of the members of this family. In such a year, Hammerstrom *et al*. (2011) found that the AtxA virulence regulator forms dimers, tetramers and higher molecular species under certain conditions (bacterial cultures grown in elevated CO_2/bicarbonate). In 2012, Hondorp *et al*. (2012) reported a correlation between the capacity of a His-tagged Mga protein to oligomerize in solution (without DNA) and its ability to activate transcription. Also, DNA binding was found to be necessary but insufficient for fully transcriptional activation. Concerning the Mga*Spn* protein, under our experimental conditions, it forms, in the main, dimers in solution, but it is able to form higher-order molecular species during sedimentation equilibrium experiments carried out at higher protein concentrations. Moreover, Mga*Spn* is able to generate multimeric protein-DNA complexes in the presence of linear double-stranded DNAs. We propose that the spreading phenomenon described in this work might be a useful mechanism used by Mga*Spn* to regulate promoters located at a certain distance from the initial protein-binding site.

There are other examples of regulators with ability to oligomerize, such as several bacterial nucleoid-associated proteins. These proteins, which act as both architectural proteins and global transcriptional regulators, exist primarily as dimers, but can multimerize into higher-order oligomers. It is the case, for instance, of the H-NS protein

from enteric bacteria. Depending on different parameters such as protein concentration, temperature and monovalent cations, different association states of the protein are found. H-NS forms dimers at low protein concentrations but can multimerize into higher-order oligomers as its concentration increases (Smyth *et al*., 2000). H-NS has been shown to regulate gene expression through a variety of mechanisms (Fang and Rimsky, 2008), and H-NS oligomerization is essential for its function. As pointed out by Dillon and Dorman (2010), one feature of the nucleoid-associated proteins that makes them excellent regulators of gene expression at the global level is their promiscuity in their interactions with DNA. Initially, H-NS was shown to bind preferentially to DNA containing curved regions (Yanisch-Perron *et al*., 1985). Later on, Lang *et al*. (2007) identified high-affinity DNA-binding sites for H-NS in AT-rich regions of the chromosomal DNA. They proposed that H-NS binds initially to high-affinity sites (nucleation sites) and then spreads along the AT-rich DNA regions. In this model, H-NS spreading from specific sites would enable the silencing of extensive regions of the bacterial chromosome. Recently, Shin *et al.* (2012) investigated the H-NS-mediated repression of the *LEE5* promoter. Their results supported a new mechanism by which DNA-bound proteins communicate with each other. Basically, H-NS binds to a cluster of A tracks located upstream of the *LEE5* promoter (position -114) and then spreads, presumably through oligomerization, to a site at the promoter where H-NS makes specific contacts with the RNA polymerase. Thus, in this model, H-NS spreading on DNA would facilitate encounters between distant regulatory elements.

It has been proposed that sequence-dependent DNA structures may be critical components in the assembly of higher-order protein-DNA complexes (Rohs *et al*., 2010; Serrano *et al*., 1993). Thus, we analysed the nucleotide sequence of the regions (stretches of 50-bp) flanking the primary binding sites of Mga*Spn*. Compared to the global A+T content (60.3%) of the pneumococcal R6 genome, such adjacent regions display a high A+T content (74-88%), which might facilitate Mga*Spn* spreading due to the potential bendability of the DNA in these regions. Hydroxyl radical footprinting assays with the 224-bp DNA suggested that binding of Mga*Spn* to its primary site (*Pmga* promoter) induces a conformational change within the -35 region. Then, Mga*Spn*-dependent changes in the conformation of the DNA may promote sequential binding of additional protein units. Therefore, a combination of local variations in DNA structure might be directing the specificity of Mga*Spn* for a particular DNA region.

Role of Mga*Spn* in self-regulation

It is known that the Mga global transcriptional regulator of GAS is able to activate the expression of its own gene in order to amplify the Mga regulon during the first stages of infection (McIver *et al*., 1999). In this work, we have identified the promoter of the pneumococcal *mgaSpn* gene. By RT-PCR, transcriptional fusions and primer extension assays we have shown that the *mgaSpn* regulatory gene is transcribed *in vivo* from the *Pmga* promoter. In addition, by *in vitro* transcription experiments, we have demonstrated that this promoter is recognized by the pneumococcal housekeeping σ^{43} factor. The *PB* activation region (70-bp), identified in this work between the *P1623B* and *Pmga* divergent promoters, is located 80 nucleotides upstream of the *Pmga* transcription initiation site. Since Mga*Spn* interacts with such a region to activate the *P1623B* promoter, we initially proposed that binding of Mga*Spn* to the *PB* activation region might also influence the activity of its own promoter *in vivo*. However, we found that it was not the case, at least using a *Pmga-gfp* transcriptional fusion in which the *Pmga* promoter extended to the +6 position. Clearly, the activity of the *Pmga* promoter did not change in response to the intracellular level of Mga*Spn*. Subsequent to this study, we found that, in addition to the *PB* activation region, Mga*Spn* is able to recognize the *Pmga* promoter region as primary binding site. Specifically, hydroxyl radical footprinting experiments demonstrated that Mga*Spn* binds to sequences located between the -23 and +21 positions. Further work is needed to determine whether binding of Mga*Spn* to such a site influences the expression of its own gene *in vivo*. Interestingly, binding of H-NS to two sites located at a distance has been reported for many of its target genes, and in some cases one of the binding sites is located downstream of the promoter (Fang and Rimsky, 2008; Cendra Mdel *et al*., 2013).

Possible mechanism(s) of Mga*Spn*-mediated transcriptional regulation

In prokaryotes, the strength of a particular promoter can be modulated by transcription factors (activator or repressor proteins) that bind to specific DNA sites. Different modes of activation and repression by this transcription factors have been described (for a review see Browning and Busby, 2004; Minchin and Busby, 2009; van Hijum *et al*., 2009). However, the transcriptional regulation mechanism(s) mediated by the Mga/AtxA family of global regulators remains as a key question in the field. Regarding the Mga regulator of GAS, several binding sites have been identified using either an MBP-tagged or a His-tagged Mga protein (gel retardation and DNase I footprinting assays) (McIver *et al*., 1995; McIver *et al*., 1999). Based on the number of

binding sites and their position with respect to the start of transcription, three categories of Mga-regulated promoters have been proposed: category A promoters (the majority of the Mga-regulated promoters) contain a single proximal binding site of 45-bp centred at the -54 position relative to the start of transcription; category B promoters (*PsclA* and *Psof/sfbX*) have only a distal Mga-binding site of 45-bp (centred at the -168 position from the start of transcription), and category C promoters (*Pmga*) have two 59-bp binding sites (Almengor and McIver, 2004; Hondorp and McIver, 2007; Hause and McIver, 2012). Although no experimental data are available, the overlapping between the Mga-binding site and the -35 element in category A promoters suggests that Mga might interact with the α-subunit of the RNAP, whereas Mga binding to a distal site in category B promoters suggests that the interaction with the RNAP might require DNA bending (Almengor *et al*., 2006). Concerning the AtxA regulator, AtxA-DNA interactions have not been reported yet, even though it is known that AtxA influences the activity of many promoters *in vivo* and has two putative HTH DNA-binding motifs within the N-terminal region.

The work presented in Chapter 3 has demonstrated that Mga*Spn* acts as a positive regulator of the *spr1623-spr1626* operon (Solano-Collado *et al*., 2012). It activates directly the *P1623B* promoter through interaction with the *PB* activation region. Such an interaction has been extensively studied by EMSA and footprinting techniques (Chapter 4) (Solano-Collado *et al*., 2013) showing that Mga*Spn* binds to a region located between positions -60 and -99 of the *P1623B* promoter. This location could allow its interaction with the α subunit of the RNAP. If this were the case, such interaction would recruit the RNAP to the promoter, stabilize the initiation complex and activate transcription, as it has been described for other transcriptional activators (van Hijum *et al*., 2009). Alternatively, Mga*Spn* might activate transcription initiation at the *P1623B* promoter by inducing a conformational change in the promoter region. Both mechanisms have been reported for the nucleoid-associated protein Fis, which binds at hundreds of DNA targets in *E. coli* and affects gene expression (positively or negatively) by at least six different mechanisms (reviewed by Browning *et al*., 2010). Furthermore, the ability of Mga*Spn* to spread on the DNA generating multimeric complexes suggests that Mga*Spn* spreading from specific sites might also play an important role in the regulation of particular promoters, as it is the case of the H-NS global modulator (Shin *et al*., 2012).

Conclusions

This study has contributed to the molecular knowledge of a novel pneumococcal virulence transcriptional regulator, the Mga*Spn* protein, and has led to the following conclusions:

1. The *mgaSpn* regulatory gene is transcribed *in vivo* from the *Pmga* promoter, which is recognized *in vitro* by the housekeeping σ^{43} factor of *S. pneumoniae*. The transcription start site of the *mgaSpn* gene is located 39 nucleotides upstream of the translation start codon.

2. The *spr1623-spr1626* operon, which is adjacent to the *mgaSpn* gene, is transcribed *in vivo* from two promoters, *P1623A* and *P1623B*. These promoters are divergent from the *Pmga* promoter.

3. Mga*Spn* acts directly as a positive transcriptional regulator. It activates the *P1623B* promoter *in vivo* and, therefore, the expression of the *spr1623-spr1626* operon. This activation requires binding of Mga*Spn* to a site located upstream of the *P1623B* promoter (positions -60 to -99).

4. Mga*Spn* is highly conserved in the pneumococcal strains whose genomes have been totally or partially sequenced. It is predicted to have the same organization of functional domains as the Mga and AtxA global response regulators.

5. Mga*Spn* has a high content of α-helices (around 55%). In solution, and under the conditions tested, Mga*Spn* forms predominantly dimers; however, it tends to form higher-order molecular species as a function of protein concentration.

6. Mga*Spn* binds to linear double-stranded DNAs with high affinity, but with low sequence specificity.

7. Local DNA conformations (e. g. intrinsic curvature) might contribute to the DNA-binding specificity of Mga*Spn*.

8. Upon binding to its primary site, Mga*Spn* is able to spread (presumably through oligomerization) along the adjacent DNA regions generating multimeric protein-DNA complexes.

References

Adamczak, R., Porollo, A., and Meller, J. (2005). Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* **59**, 467-475.

Aiyar, S.E., Gourse, R.L., and Ross, W. (1998). Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit. *Proc Natl Acad Sci USA* **95**, 14652-14657.

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

Almengor, A.C., Walters, M.S., and McIver, K.S. (2006). Mga is sufficient to activate transcription in vitro of *sof-sfbX* and other Mga-regulated virulence genes in the group A *Streptococcus*. *J Bacteriol* **188**, 2038-2047.

Andrade, M.A., Chacon, P., Merelo, J.J., and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* **6**, 383-390.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernández, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **40,** W597-603.

Barrett, J.F., and Hoch, J.A. (1998). Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob Agents Chemother* **42**, 1529-1536.

Beier, D., and Gross, R. (2006). Regulation of bacterial virulence by two-component systems. *Current Opinion in Microbiology Cell Regulation* **9**, 143-152.

Bender, M.H., and Yother, J. (2001). CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae*. *J Biol Chem* **276**, 47966-47974.

Bentley, S.D., Aanensen, D.M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Kaltoft, M.S., Barrell, B., Reeves, P.R., Parkhill, J., and Spratt, B.G. (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* **2**, e31.

Bethe, G., Nau, R., Wellmer, A., Hakenbeck, R., Reinert, R.R., Heinz, H.P., and Zysk, G. (2001). The cell wall-associated serine protease PrtA: a highly conserved virulence factor of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **205**, 99-104.

Blue, C.E., and Mitchell, T.J. (2003). Contribution of a response regulator to the virulence of *Streptococcus pneumoniae* is strain dependent. *Infect Immun* **71**, 4405-4413.

Bogaert, D., De Groot, R., and Hermans, P.W. (2004). *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**, 144-154.

Bongiorni, C., Fukushima, T., Wilson, A.C., Chiang, C., Mansilla, M.C., Hoch, J.A., and Perego, M. (2008). Dual promoters control expression of the *Bacillus anthracis* virulence factor AtxA. *J Bacteriol* **190**, 6483-6492.

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. *Mol Microbiol* **40**, 572-585.

Brown, J.S., Gilliland, S.M., Ruiz-Albert, J., and Holden, D.W. (2002). Characterization of *pit*, a *Streptococcus pneumoniae* iron uptake ABC transporter. *Infect Immun* **70**, 4389-4398.

Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**, 57-65.

Browning, D.F., Grainger, D.C., and Busby, S.J. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* **13**, 773-780.

Brückner, R., Nuhn, M., Reichmann, P., Weber, B., and Hakenbeck, R. (2004). Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. *Int J Med Microbiol* **294**, 157-168.

Burdett, V. (1980). Identification of tetracycline-resistant R-plasmids in *Streptococcus agalactiae* (group B). *Antimicrob Agents Chemother* **18**, 753-760.

Carey, J. (1988). Gel retardation at low pH resolves *trp* repressor-DNA complexes for quantitative study. *Proc Natl Acad Sci USA* **85**, 975-979.

Carey, J. (1991). Gel retardation. *Methods Enzymol* **208**, 103-117.

Cendra Mdel, M., Juarez, A., Madrid, C., and Torrents, E. (2013). H-NS is a novel transcriptional modulator of the ribonucleotide reductase genes in *Escherichia coli*. *J Bacteriol* **195**, 4255-4263.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**, W197-201.

Cole, J.L. (2004). Analysis of heterogeneous interactions. *Methods Enzymol* **384**, 212-232.

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem Sci* **25**, 147-150.

Compton, L.A., and Johnson, W.C., Jr. (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* **155**, 155-167.

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38.

Cunningham, M.W. (2000). Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **13**, 470-511.

Dai, Z., and Koehler, T.M. (1997). Regulation of anthrax toxin activator gene (*atxA*) expression in *Bacillus anthracis*: temperature, not CO2/bicarbonate, affects AtxA synthesis. *Infect Immun* **65**, 2576-2582.

deHaseth, P.L., Zupancic, M.L., and Record, M.T., Jr. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol* **180**, 3019-3025.

Deutscher, J., Francke, C., and Postma, P.W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939-1031.

Diekmann, S. (1987). Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. *Nucleic Acids Res* **15**, 247-265.

Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* **8**, 185-195.

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

Espinosa, M., López, P., Pérez-Ureña, M.T., and Lacks, S.A. (1982). Interspecific plasmid transfer between *Streptococcus pneumoniae* and *Bacillus subtilis*. *Mol Gen Genet* **188**, 195-201.

Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W., Niu, W., Ebright, R.H., and Gourse, R.L. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* **13**, 2134-2147.

Fang, F.C., and Rimsky, S. (2008). New insights into transcriptional regulation by H-NS. *Curr Opin Microbiol* **11**, 113-120.

Fouet, A. (2010). AtxA, a *Bacillus anthracis* global virulence regulator. *Res Microbiol,* **161**, 735-742.

Fredrick, K., Caramori, T., Chen, Y.F., Galizzi, A., and Helmann, J.D. (1995). Promoter architecture in the flagellar regulon of *Bacillus subtilis*: high-level expression of flagellin by the sigma D RNA polymerase requires an upstream promoter element. *Proc Natl Acad Sci USA* **92**, 2582-2586.

Gabrielian, A., Vlahovicek, K., and Pongor, S. (1997). Distribution of sequence-dependent curvature in genomic DNA sequences. *FEBS Lett* **406**, 69-74.

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

Giard, J.C., Rince, A., Capiaux, H., Auffray, Y., and Hartke, A. (2000). Inactivation of the stress- and starvation-inducible *gls24* operon has a pleiotrophic effect on cell morphology, stress sensitivity, and gene expression in *Enterococcus faecalis*. *J Bacteriol* **182**, 4512-4520.

Graham, M.R., Virtaneva, K., Porcella, S.F., Gardner, D.J., Long, R.D., Welty, D.M., Barry, W.T., Johnson, C.A., Parkins, L.D., Wright, F.A., and Musser, J.M. (2006). Analysis of the transcriptome of group A *Streptococcus* in mouse soft tissue infection. *Am J Pathol* **169**, 927-942.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology* **57**, 441-466.

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

Hajaj, B., Yesilkaya, H., Benisty, R., David, M., Andrew, P.W., and Porat, N. (2012). Thiol peroxidase is an important component of *Streptococcus pneumoniae* in oxygenated environments. *Infect Immun* **80**, 4333-4343.

Hall-Stoodley, L., Nistico, L., Sambanthamoorthy, K., Dice, B., Nguyen, D., Mershon, W.J., Johnson, C., Hu, F.Z., Stoodley, P., Ehrlich, G.D., and Post, J.C. (2008). Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in *Streptococcus pneumoniae* clinical isolates. *BMC Microbiol* **8**, 173.

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. *Mol Microbiol* **82**, 634-647.

Hanahan, D. (1983). Studies in transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-580.

Haugen, S.P., Ross, W., and Gourse, R.L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* **6**, 507-519.

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. *J Bacteriol* **194**, 4904-4919.

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

Hava, D.L., Hemsley, C.J., and Camilli, A. (2003). Transcriptional Regulation in the *Streptococcus pneumoniae rlrA* Pathogenicity Islet by RlrA. *J Bacteriol* **185**, 413-421.

Havarstein, L.S., Martin, B., Johnsborg, O., Granadel, C., and Claverys, J.P. (2006). New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Mol Microbiol* **59**, 1297-1307.

Heim, R., Cubitt, A.B., Tsien, R.Y. (1995). Improved green fluorescence. *Nature* **373**, 663-664.

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*. *J Bacteriol* **185**, 6640-6647.

Hendriksen, W.T., Bootsma, H.J., van Diepen, A., Estevão, S., Kuipers, O.P., de Groot, R., and Hermans, P.W.M. (2009). Strain-specific impact of PsaR of *Streptococcus pneumoniae* on global gene expression and virulence. *Microbiology* **155**, 1569-1579.

Hendriksen, W.T., Silva, N., Bootsma, H.J., Blue, C.E., Paterson, G.K., Kerr, A.R., de Jong, A., Kuipers, O.P., Hermans, P.W., and Mitchell, T.J. (2007). Regulation of gene expression in *Streptococcus pneumoniae* by response regulator 09 is strain dependent. *J Bacteriol* **189**, 1382-1389.

Hirst, R.A., Kadioglu, A., O'Callaghan, C., and Andrew, P.W. (2004). The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin Exp Immunol* **138**, 195-201.

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*. *Mol Microbiol* **88**, 1176-1193.

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. *Mol Microbiol* **83**, 953-967.

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

Horinouchi, S., and Weisblum, B. (1982). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol* **150**, 815-825.

Hoskins, J., Alborn, W.E., Jr., 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. Jr, Skatrud, P.L., Glass, J.I. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**, 5709-5717.

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. *Infect Immun* **78**, 704-715.

Hynes, W.L., and Walton, S.L. (2000). Hyaluronidases of Gram-positive bacteria. *FEMS Microbiol Lett* **183**, 201-207.

Johnston, J.W., Myers, L.E., Ochs, M.M., Benjamin, W.H., Jr., Briles, D.E., and Hollingshead, S.K. (2004). Lipoprotein PsaA in virulence of *Streptococcus pneumoniae*: surface accessibility and role in protection from superoxide. *Infect Immun* **72**, 5858-5867.

Jomaa, M., Yuste, J., Paton, J.C., Jones, C., Dougan, G., and Brown, J.S. (2005). Antibodies to the iron uptake ABC transporter lipoproteins PiaA and PiuA promote opsonophagocytosis of *Streptococcus pneumoniae*. *Infect Immun* **73**, 6852-6859.

Joyet, P., Bouraoui, H., Ake, F.M., Derkaoui, M., Zebre, A.C., Cao, T.N., Ventroux, M., Nessler, S., Noirot-Gros, M.F., Deutscher, J., and Milohanic, E. (2013). Transcription regulators controlled by interaction with enzyme IIB components of the phosphoenolpyruvate:sugar phosphotransferase system. *Biochim Biophys Acta* **1834**, 1415-1424.

Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**, 288-301.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371.

Kharat, A.S., and Tomasz, A. (2003). Inactivation of the *srtA* gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells *in vitro. Infect Immun* **71**, 2758-2765.

Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciufo, S., Fedorov, B., Kiryutin, B., O'Neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I., and Tatusova, T. (2009). The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res* **37**, D216-223.

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

Lacks, S., and Greenberg, B. (1977). Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J Mol Biol* **114**, 153-168.

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

Lacks, S.A., Greenberg, B., and Lopez, P. (1995). A cluster of four genes encoding enzymes for five steps in the folate biosynthetic pathway of *Streptococcus pneumoniae*. *J Bacteriol* **177**, 66-74.

Lacks, S.A., Lopez, P., Greenberg, B., and Espinosa, M. (1986). Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. *Journal of Molecular Biology* **192**, 753-765.

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. *J Bacteriol* **189**, 38-51.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Analytical ultracentrifugation in biochemistry and polymer sciences. Harding SE, Rowe A, Horton JC, editors Cambridge: Royal Society of Chemistry, pp. 90-125.

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

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press New York*.

Maxam, A.M., and Gilbert, W. (1980). Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol* **65**, 499-560.

McAllister, L.J., Tseng, H.J., Ogunniyi, A.D., Jennings, M.P., McEwan, A.G., and Paton, J.C. (2004). Molecular analysis of the *psa* permease complex of *Streptococcus pneumoniae*. *Mol Microbiol* **53**, 889-901.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405.

McIver, K., Heath, A., Green, B., and Scott, J. (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.

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

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**, 1591-1601.

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. *J Bacteriol* **181**, 5373-5383.

Miller, W.G., and Lindow, S.E. (1997). An improved GFP cloning cassette designed for prokaryotic transcriptional fusions. *Gene* **191**, 149-153.

Minchin, S.D., and Busby, S.J. (2009). Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* **47**, 6-12.

Mitchell, A.M., and Mitchell, T.J. (2010). *Streptococcus pneumoniae*: virulence factors and variation. *Clin Microbiol Infect* **16**, 411-418.

Mitchell, J.E., Zheng, D., Busby, S.J.W., and Minchin, S.D. (2003). Identification and analysis of 'extended –10' promoters in *Escherichia coli*. *Nucleic Acids Res* **31**, 4689-4695.

Morona, J.K., Miller, D.C., Morona, R., and Paton, J.C. (2004). The effect that mutations in the conserved capsular polysaccharide biosynthesis genes *cpsA*, *cpsB*, and *cpsD* have on virulence of *Streptococcus pneumoniae*. *J Infect Dis* **189**, 1905-1913.

Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., and Weiser, J.N. (2007). Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infect Immun* **75**, 83-90.

Osipiuk, J., Wu, R., Jedrzejczak, R., Moy, S., and Joachimiak, A. (2011). Putative Mga family transcriptional regulator from *Enterococcus faecalis*.

Paterson, G.K., Blue, C.E., and Mitchell, T.J. (2006). Role of two-component systems in the virulence of *Streptococcus pneumoniae*. *J Med Microbiol* **55**, 355-363.

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

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 Microbiol Lett* **285**, 170-176.

Paton, J.C., Lock, R.A., and Hansman, D.J. (1983). Effect of immunization with pneumolysin on survival time of mice challenged with *Streptococcus pneumoniae*. *Infect Immun* **40**, 548-552.

Perry, J., Koteva, K., and Wright, G. (2011). Receptor domains of two-component signal transduction systems. *Mol Biosyst* **7**, 1388-1398.

Pessen, H., and Kumosinski, T.F. (1985). Measurements of protein hydration by various techniques. *Methods Enzymol* **117**, 219-255.

Provencher, S.W., and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., and Finn, R.D. (2012). The Pfam protein families database. *Nucleic Acids Res* **40**, D290-301.

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**, 491-508.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* **79**, 233-269.

Ross, W., Aiyar, S.E., Salomon, J., and Gourse, R.L. (1998). *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J Bacteriol* **180**, 5375-5383.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**, 1407-1413.

Ross, W., and Gourse, R.L. (2009). Analysis of RNA polymerase-promoter complex formation. *Methods Related to Bacterial Transcriptional Control* **47**, 13-24.

Rost, B., Yachdav, G., and Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res* **32**, W321-326.

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.

Sabelnikov, A.G., Greenberg, B., and Lacks, S.A. (1995). An extended -10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. *Journal of Molecular Biology* **250**, 144-155.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* **412**, 754-771.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual. *Cold Spring Harbor Laboratory Press New York*.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463-5467.

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**, 368-379.

Schuck, P., and Rossmanith, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54**, 328-341.

Serrano, M., Salas, M., and Hermoso, J.M. (1993). Multimeric complexes formed by DNA-binding proteins of low sequence specificity. *Trends Biochem Sci* **18**, 202-206.

Shin, M., Lagda, A.C., Lee, J.W., Bhat, A., Rhee, J.H., Kim, J.S., Takeyasu, K., and Choy, H.E. (2012). Gene silencing by H-NS from distal DNA site. *Mol Microbiol* **86**, 707-719.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Smyth, C.P., Lundback, T., Renzoni, D., Siligardi, G., Beavil, R., Layton, M., Sidebotham, J.M., Hinton, J.C., Driscoll, P.C., Higgins, C.F., *et al.* (2000). Oligomerization of the chromatin-structuring protein H-NS. *Mol Microbiol* **36**, 962-972.

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

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.

Spiess, E., and Lurz, R. (1988). Electron microscopic analysis of nucleic acids and nucleic acid-protein complexes. *Methods Microbiol* **20**, 293-323.

Sreerama, N., and Woody, R.W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* **209**, 32-44.

Stassi, D.L., Lopez, P., Espinosa, M., and Lacks, S.A. (1981). Cloning of chromosomal genes in *Streptococcus pneumoniae. Proc Natl Acad Sci USA* **78**, 7028-7032.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**, 183-215.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130.

Teng, F., Nannini, E.C., and Murray, B.E. (2005). Importance of *gls24* in virulence and stress response of *Enterococcus faecalis* and use of the Gls24 protein as a possible immunotherapy target. *J Infect Dis* **191**, 472-480.

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**, 498-506.

Tong, H.H., Blue, L.E., James, M.A., and DeMaria, T.F. (2000). Evaluation of the virulence of a *Streptococcus pneumoniae* neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect Immun* **68**, 921-924.

Tsatsaronis, J.A., Hollands, A., Cole, J.N., Maamary, P.G., Gillen, C.M., Zakour, N.L., Kotb, M., Nizet, V., Beatson, S.A., Walker, M.J., and Sanderson-Smith, M.L. (2013). Streptococcal collagen-like protein A and general stress protein 24 are immunomodulating virulence factors of group A *Streptococcus*. *FASEB J* **27**, 2633-2643.

Tseng, H.J., McEwan, A.G., Paton, J.C., and Jennings, M.P. (2002). Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress. *Infect Immun* **70**, 1635-1639.

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**, 644-655.

Tullius, T.D., and Dombroski, B.A. (1986). Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc Natl Acad Sci USA* **83**, 5469-5473.

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

van der Poll, T., and Opal, S.M. (2009). Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* **374**, 1543-1556.

van Hijum, S.A., Medema, M.H., and Kuipers, O.P. (2009). Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev* **73**, 481-509, Table of Contents.

van Holde, K.E. (1985). *Physical Biochemistry*. 2nd edn. Prentice Hall: Englewoods Cliffs.

Vlahovicek, K., Kajan, L., and Pongor, S. (2003). DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res* **31**, 3686-3687.

Voskuil, M.I., and Chambliss, G.H. (1998). The -16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. *Nucleic Acids Res* **26**, 3584-3590.

Weiss-Salz, I., and Yagupsky, P. (2010). Asymptomatic Carriage of Respiratory Pathogens: "The Wolf shall Dwell with the Lamb…and a Little Child shall Lead them". *The Open Infectious Diseases Journal* **4**, 11-15.

Wooldridge, K.G., and Williams, P.H. (1993). Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiol Rev* **12**, 325-348.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yother, J. (2011). Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annu Rev Microbiol* **65**, 563-581.

Zhang, M., McDonald, F.M., Sturrock, S.S., Charnock, S.J., Humphery-Smith, I., and Black, G.W. (2007). Group A *Streptococcus* cell-associated pathogenic proteins as revealed by growth in hyaluronic acid-enriched media. *Proteomics* **7**, 1379-1390.