



UNIVERSIDAD
NACIONAL
DE COLOMBIA

**Identificación y caracterización de una
proteína de unión al gen *icaA* y evaluación de
su potencial participación en la formación de
biofilm en *Staphylococcus aureus*.**

Javier Antonio Escobar Pérez

Universidad Nacional de Colombia
Facultad de Ciencias, Instituto de Biotecnología, Doctorado en Biotecnología
Bogotá D.C., Colombia

2018

Identificación y caracterización de una proteína de unión al gen *icaA* y evaluación de su potencial participación en la formación de biofilm en *Staphylococcus aureus*.

Javier Antonio Escobar Pérez

Tesis presentada como requisito parcial para optar al título de:

Doctor en Biotecnología

Director:

Ph.D. Jaime Castellanos

Codirectora:

Ph.D. Natasha Vanegas Gómez

Línea de Investigación:

Estudio de factores de virulencia y elementos genéticos móviles bacterianos

Grupo de Investigación:

Laboratorio de Genética Molecular Bacteriana, Universidad El Bosque

Universidad Nacional de Colombia

Facultad de Ciencias, Instituto de Biotecnología, Doctorado en Biotecnología

Bogotá D.C., Colombia

2018

*Para mis hijos Andrés, Gabby y Amelia, y mi
esposa Marley. El ATP de todos mis días.*

“There is nothing more frightful than ignorance in action.”

Johann Wolfgang Goethe

Agradecimientos

Quiero expresar mis más sinceros agradecimientos a todas las personas que de alguna manera contribuyeron a la realización de este estudio y que por olvido (no intencional) no son nombrados. Gracias infinitas:

- A Mi familia, por soportar estos años de postergación de innumerables actividades familiares debido a mi eterna frase “estoy realizando mi tesis”.
- A Jaime Castellanos y Natasha, por su invaluable apoyo y confianza, no solo durante mi tesis, sino desde que inicie mi carrera como investigador en la Universidad El Bosque.
- A todos los docentes del Laboratorio de Genética Molecular Bacteriana, compartimos más de ocho horas diarias; y especialmente a Zaida y Katterine, porque sin su desinteresada colaboración este proceso se hubiera demorado mucho más.
- Al Doctor Otero por creer en nosotros y apoyarnos siempre....muchas gracias.
- A la Universidad El Bosque por acogerme y permitirme realizarme como investigador, algo tan difícil de realizar en nuestro país.
- A la Universidad Nacional de Colombia (mi “*alma mater*”),
- Al Instituto de Biotecnología, a sus docentes por compartir sus experiencias y expandir mis fronteras del conocimiento.
- Al Doctor Iñigo Lasa de la Universidad de Pamplona en España, por su exagerada amabilidad y por enseñarnos a realizar los experimentos de mutagénesis en *Staphylococcus aureus*.

Resumen

El biofilm es definido como una comunidad microbiana altamente estructurada, compleja, inmersas dentro de una matriz de biopolímeros y adheridas a un sustrato o interfase inerte o viviente. Las bacterias dentro del biofilm están protegidas del ataque del sistema inmune del hospedero, de los antibióticos y otros agentes antimicrobianos. *Staphylococcus aureus* es una bacteria Gram positiva considerada como un patógeno oportunista, ya que hace parte del microbioma humano pero tiene la capacidad de causar infecciones. Esta bacteria genera un gran impacto en la salud pública mundial, ya que es el cuarto microorganismo recuperado de infecciones de pacientes hospitalizados. Adicionalmente, *S. aureus* tiene la capacidad de formar biofilm sobre diferentes superficies, incluyendo implantes médicos. El biofilm en *S. aureus* es producido por el operón *ica*, el cual está conformado por los genes *icaA*, *icaB*, *icaC* e *icaD*, que codifican las enzimas relacionadas con la producción del polisacárido de adhesión intercelular (PIA), componente principal de la matriz extracelular del biofilm en este microorganismo. En trabajos realizados previamente en el Laboratorio de Genética Molecular Bacteriana (LGMB) de la Universidad el Bosque se determinó la existencia de una región dentro del gen *icaA* altamente conservada en varias especies de *Staphylococcus* spp y que además contiene una secuencia palindrómica. Además existen indicios de proteínas que se unen de manera específica a esta secuencia palindrómica. El objetivo de este estudio fue identificar y caracterizar una proteína de unión a esta secuencia palindrómica del gen *icaA* y determinar su posible participación en la formación del biofilm en *S. aureus*. A partir de ensayos de retardamiento en gel y un proceso de purificación basado en cromatografía de proteínas de alta resolución se logró establecer la existencia de dos proteínas que reconocen esta secuencia palindrómica, SarX y una proteína hipotética, previamente no descrita. Ya se ha determinado que SarX tiene la habilidad para unirse al gen *icaA*. Sin embargo, la proteína hipotética correspondió a un ortólogo de la proteína YheA de *Bacillus subtilis* y además contiene un dominio denominado *com_ylbF*, el cual también está presente en las proteínas YlbF y YmcA de esta misma bacteria. Adicionalmente, se realizó un análisis estructural “*in silico*” de la proteína YheA, y de las

- X Identificación y caracterización de una proteína de unión al gen *icaA* y evaluación de su potencial participación en la formación de biofilm en *Staphylococcus aureus*.
-

proteínas YlbF y YmcA de *S. aureus* y *B. subtilis*. Modelos tridimensionales (3D) de las tres proteínas fueron construidos usando los programas I-TASSER y Quark. A partir de este análisis dos nuevas características de YheA, YlbF y YmcA fueron encontradas: una estructura 3D altamente conservada, a pesar de sus bajos porcentajes de identidad <16% y la presencia de un motivo putativo específico de cada proteína localizado dentro del dominio *com_ylbF*. Esta proteína YheA tiene un alto contenido de glutamina (13,2%), a nivel general y en su motivo putativo (QQKQMQ), lo cual sugiere que este aminoácido podría estar participando en la función de la proteína. Basados en estos hallazgos, se sugiere renombrar YheA como proteína rica en glutamina o Qrp (por su nombre en inglés Glutamine-rich protein) en *S. aureus*. Usando ensayos de mutagénesis dirigida, la delección del gen *yheA/qrp*, que codifica para la proteína Qrp/YheA, produce una disminución en la formación del biofilm y un aumento en la hemólisis de eritrocitos por *S. aureus*. Estos resultados señalan que *S. aureus* posee una nueva familia de proteínas que contienen el dominio *com_ylbF*, de las cuales una de ellas, Qrp/YheA, está participando en la formación de biofilm a través de una interacción, directa o indirecta, con una secuencia palindrómica presente dentro del gen *icaA*.

Palabras clave: *Staphylococcus aureus*, biofilm, operón *ica*, Qrp/YheA, dominio *com_ylbF*.

Abstract

The biofilm is a complex and high structured microbial community, which is immersed within a matrix composed of biopolymers with ability to adhere to inert or living substrates. The bacteria within the biofilm are protected the host defenses, antibiotics and other antimicrobial compounds. *Staphylococcus aureus* is a Gram positive bacterium, an opportunistic pathogen that generates a high clinic impact in the public health due to that it is the fourth bacterium recovered from infections in hospitalized patients. In addition, *S. aureus* has the ability to form biofilm on different substrates, including medical implants. The biofilm in *S. aureus* is produced by *ica* operon, which is composed for the *icaA*, *icaB*, *icaC* and *icaD* genes, encoding the enzymes responsible for the Polysaccharide Intercellular Adhesin (PIA) synthesis, principal biopolymer of the biofilm. In previous studies performed in the Laboratorio de Genetica Molecular Bacteriana of the Universidad El Bosque, a high conserved region was identified within *icaA* gene of the several Staphylococcal species, and interestingly, this region contained a palindromic sequence. There are experimental evidences that suggest the existence of proteins that recognize specifically this palindromic sequence. For these reasons, the objective of this study was to identify and to characterize one protein that recognizes this palindromic sequence and to establish whether it participates in the biofilm formation in *S. aureus*. Using Electrophoretic Mobility Shift Assay (EMSA) and FPLC-based protein purification, we achieve to identify two proteins that recognize this palindromic sequence, SarX and a hypothetical protein, which has not been previously described. It is known that SarX protein has the ability for binding to *icaA* gene. However, the hypothetical protein was an ortholog of the YheA protein of *Bacillus subtilis* and it contains the *com_ylbF* domain, which has also been identified in the proteins YlbF and YmcA in this same bacterium. Then, we performed an “*in silico*” structural analysis of the YheA protein, as well as the YlbF and YmcA. Three-dimensional (3D) models of these three proteins, in both *S. aureus* and *B. subtilis*, were built using the I-TASSER and Quark programs. From these analysis two new and important characteristics for YheA, YlbF and YmcA were found: a highly conserved 3D structure (despite their low amino acid identity, <16%) and the presence of a putative conserved motif located in the central region of the domain, which could be involved in its function. The YheA protein has a high content of glutamine (13.2%), even within its putative motif (QQKQMQ), which suggests that this amino acid could to be participating in the protein function. Based on these findings, we suggest

renaming YheA as glutamine-rich protein (Qrp) in *S. aureus*. Interestingly, the *yheA/qrp* gene deletion in the *S. aureus* genome produces a decrease in the biofilm formation but an increase in the red blood cells hemolysis. Our results indicate that *S. aureus* possesses a new protein family that contain the *com_ylbF* domain, and that least one of these, Qrp/YheA, is participating in the biofilm formation through either direct or indirect interaction with a palindromic sequence located within the *icaA* gene.

Keywords: *Staphylococcus aureus*, biofilm, *ica* operon, Qrp/YheA, *com:ylbF* domain.

Contenido

	Pág.
Resumen	IX
Lista de figuras.....	XV
Lista de tablas	XXI
Introducción	1
1. Capítulo 1.....	5
1.1 Generalidades de <i>Staphylococcus aureus</i>	5
1.2 Formación, estructura y adherencia del Biofilm o biopelícula en Gram positivos.....	6
1.2.1 Biofilm en <i>Staphylococcus aureus</i>	8
1.2.2 Dispersión del biofilm	14
1.2.3 Secuencias palindrómicas y regulación génica	15
2. Capítulo 2. Identificación de proteínas de unión al gen <i>icaA</i> en <i>Staphylococcus aureus</i> 17	17
2.1 Materiales y métodos.....	17
2.1.1 Cepas bacterianas.	17
2.1.2 Obtención de bacterias en crecimiento planctónico y formando biofilm.17	17
2.1.3 Obtención del extracto de proteínas totales de <i>S. aureus</i>	18
2.1.4 Detección y seguimiento de la proteína de unión al gen <i>icaA</i> por retardamiento en gel.	18
2.1.5 Identificación de las proteínas de unión al gen <i>icaA</i>	19
2.2 Resultados.....	20
2.2.1 Diseño de la sonda de DNA.	20
2.2.2 Detección y purificación parcial de las proteínas de unión al gen <i>icaA</i> por cromatografía líquida rápida de proteínas (FPLC).	21
2.2.3 Identificación de las proteínas por espectrometría de masas MALDI- TOF	28
3. Capítulo 3. Características moleculares y estructurales de las proteínas SarX y YheA. 33	33
3.1 Materiales y métodos.....	33
3.1.1 Predicción de las características fisicoquímicas de las proteínas.	33
3.1.2 Alineamientos pareados y múltiples	33
3.1.3 Modelamiento de la estructura tridimensional (3D).....	35
3.2 Resultados.....	35
3.2.1 Análisis de la proteína SarX	35

3.2.2	Análisis de la proteína YheA (proteína hipotética P1/P9)	38
4.	Capítulo 4. Evaluación de la participación del gen <i>qrp</i> en la formación del biofilm y hemólisis en <i>Staphylococcus aureus</i>.	57
4.1	Materiales y métodos	57
4.1.1	Cepas bacterianas	57
4.1.2	Generación de las cepas mutantes, complementadas y sobre-expresadas	57
4.1.3	Ensayos de formación de biofilm y hemólisis en <i>Staphylococcus aureus</i>	65
4.1.4	Producción de la proteína recombinante Qrp	67
4.2	Resultados	69
4.2.1	Evaluación del efecto de la delección del gen <i>qrp</i> en la formación de biofilm y hemólisis de eritrocitos en <i>Staphylococcus aureus</i>	69
4.2.2	Interacción Sonda-Proteína recombinante	72
4.3	Discusión.....	78
5.	Conclusiones y recomendaciones	85
5.1	Conclusiones.....	85
5.2	Recomendaciones.....	86
	Anexo A: Protocolos de las cromatografías utilizadas en la purificación de las proteínas SarX y Qrp.	87
	Anexo B: Péptidos utilizados para la identificación de la proteína Qrp.....	89
	Anexo C: Secuencia de los genes <i>qrp</i> y <i>yheB</i> y sus proteínas codificantes.....	90
	Anexo D: Diagrama de los plásmidos pJET y pMAD.....	91
	Anexo E: soporte de la divulgación de los resultados.....	92
	Bibliografía	93

Lista de figuras

Pág.

- Figura 1-1: Esquema de los 4 pasos propuestos para la formación de biofilm en bacterias.** 1. Las bacterias plantónicas se adhieren a la superficie (inerte o viviente), 2. Comienza la interacción célula-célula y la producción de la matriz extracelular (Exopolisacárido-EPS), 3. Proliferación y multiplicación bacteriana con una maduración del biofilm y 4. Disrupción del biofilm y dispersión de las bacterias a otros sitios. Imagen tomada de MicrobeWiki y modificada de acuerdo a Otto M, 2009. [41].....7
- Figura 1-2: Estructura del operón *ica* y asociación de la función de las proteínas codificadas por sus genes *icaA*, *icaB*, *icaC* e *icaD*.** Modificado de Otto M, 2009 [41]..9
- Figura 1-3: Esquema de la compleja red de señalización que regula la formación y disrupción del biofilm en *Staphylococcus aureus*.** Tomado de Arciola y colaboradores [23]..... 13
- Figura 2-1: Alineamiento múltiple de la región 5' del gen *icaA* en varias especies de *Staphylococcus spp.* (*S. aureus* USA300, *S. epidermidis*, *S. capitis*, *S. haemolyticus*, *S. caprae*, *S. lugdunensis*).** El recuadro rojo indica la región conservada donde se encuentra la secuencia palindrómica.....21
- Figura 2-2: Ensayos de retardamiento en gel de la sonda de DNA que simula la secuencia palindrómica del gen *icaA*.** Retardamiento a diferentes cantidades de proteína en mg. El carril con 0,00 corresponde a la sonda sola, sin proteínas. Dos replicas adicionales se muestran a la derecha para extractos de proteínas obtenidos en crecimiento planctónico (EP1 y EP2) y formando biofilm (EB1 y EB2) de *Staphylococcus aureus*. Se observa la formación de dos complejos denotados por C1 y C2.....22
- Figura 2-3: Separación por CET y detección de proteínas de unión a la sonda de DNA a partir de un extracto total de proteínas. A.** Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los recuadros indican la localización de las fracciones y los recuadros sombreados las fracciones donde se detectó retardamiento de la sonda. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente.....23

Figura 2-4. Separación por CIA y detección de proteínas de unión a la sonda de DNA a partir de un extracto total de proteínas. **A.** Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente.....25

Figura 2-5: Separación por CET y detección de proteínas de unión a la sonda de DNA a partir del sobrenadante tratado con $(\text{NH}_4)_2\text{SO}_4$. **A.** Cromatograma con el perfil de separación de las proteínas en crecimiento planctónico. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la CET. +PL y +BL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente. **C.** Cromatograma con el perfil de separación de las proteínas formando en biofilm. **D.** EMSA de las fracciones obtenidas de la CET.....26

Figura 2-6: Separación por CIA y detección de proteínas de unión a la sonda de DNA a partir de las fracciones positivas de la CET (planctónico). **A.** Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. La línea negra indica el gradiente de fuerza iónica (Cl^-), la línea discontinua indica el porcentaje del gradiente al cual fueron eluidas las proteínas de interés. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo. C1 y C2: complejos DNA-proteína 1 y 2, respectivamente. **C.** SDS-PAGE de las fracciones positivas en el EMSA, las flechas rojas indican las bandas diferenciales que fueron secuenciadas. MP: Marcador de masa molecular.....27

¡Error! No se encuentra el origen de la referencia.....29

Figura 2-8: Péptidos usados en la identificación de las proteínas SarX y P1/P9 (proteína hipotética) **A.** Secuencia de la proteína SarX (staphylococcal accessory protein X) y el único péptido usado para su identificación (Subrayado y sombreado). **B.** Secuencia de la proteína hipotética P1/P9 (línea inferior) y la localización de los 8 péptidos usados para su identificación, los cuales tienen una cobertura de 71%. Los aminoácidos arginina y lisina se muestran en rojo en las dos secuencias.....31

Figura 3-1: Análisis de la proteína SarX. **A.** Estructura terciaria del homodímero de SarX donde se resalta el motivo “Helix-Turn-Helix” (HTH) de unión a DNA (Naranja) y su superficie de accesibilidad al solvente. Modificación realizada de la proteína 5HS5 reportada en PDB (Protein Data Bank). **B.** Alineamiento múltiple de las secuencias de las proteínas SarX identificadas en *S. aureus* (SarX_{Sa}), en *S. epidermidis* (SarX_{Se1} y SarX_{Se2}), en *S. lugdunensis* (SarX_{Sl}) y *S. capitis* (SarX_{Sc}). En rojo se muestra el dominio HTH.....36

Figura 3-2: Proteínas involucradas en la formación de biofilm, competencia y esporulación de *Bacillus subtilis*. Las flechas continuas significan activación, las líneas

sin flecha significan inhibición. Las flechas discontinuas representan la producción de la proteína desde el gen correspondiente. Los círculos representan las proteínas. Los genes y proteínas conservados en *S. aureus* se muestran delineados en rojo.....41

Figura 3-3: Comparación de los entornos genéticos de los genes *yheA*, *yibF*, *ymcA* y *yaat* en *Staphylococcus aureus* y *Bacillus subtilis*. En rojo los genes de interés y en negro los genes conservados; se mantienen los códigos usados para cada especie. Las cajas Pribnow y las secuencias Shine-Dalgarno se simbolizan por las paletas rojas. Las secuencias de terminación independientes de Rho identificadas por medio del programa Transterm se simbolizan por los loops en negro.....44

Figura 3-4: Estructura 3D de la proteína YheA de *Staphylococcus aureus*. Estructuras 3D de los dos únicos modelos generados para YheA por medio del programa I-Tasser y validados “*ab initio*” por el programa Quark con una resolución de 1.5Å. También se muestra la superficie de accesibilidad al solvente. **A.** Estructura del modelo 1 y su respectivo homodímero (vista frontal y lateral). **B.** Estructura del modelo 2 y su respectivo homodímero (vista frontal y lateral). **C.** Superposición de los homodímeros de los modelos 1 y 2 (vista frontal y lateral).....47

Figura 3-5: Estructura terciaria de las proteínas YheA, YmcA y YibF de *Bacillus subtilis*. Homodímeros de las 3 proteínas con su superficie de accesibilidad al solvente. **A.** Estructura de la proteína YheA modelada a partir de la estructura parcial reportada en PDB (2OEE). **B.** Estructura de la proteína YmcA reportada en PDB (2PIH). **C.** Estructura parcial de la proteína YibF modelada por el programa I-tasser y Quark. **D.** Superposición de la estructura terciaria de las 3 proteínas usando como referencia la proteína YmcA..49

Figura 3-6: Estructura terciaria de las proteínas YmcA y YibF y comparación con YheA en *Staphylococcus aureus*. **A.** Estructura de la proteína YibF. **B.** Estructura de la proteína YmcA. **C.** Superposición de la estructura terciaria de las proteínas YheA (modelo 1), YibF y YmcA. **D.** Superposición de la estructura terciaria de las proteínas YheA (modelo 2), YibF y YmcA.....50

Figura 3-7: Ortólogos de las proteínas YheA, YmcA y YibF en diferentes bacterias Gram positivas. Porcentajes de identidad de las proteínas YheA, YmcA y YibF en bacterias Gram positivas con y sin capacidad de formar esporas respecto a las proteínas identificadas en *Staphylococcus aureus*. Modificado de [83].....52

Figura 3-8: Alineamiento múltiple de las proteínas YheA, YibF y YmcA en bacterias Gram positivas. Alineamiento de las tres proteínas identificadas en las diferentes bacterias Gram positivas con y sin capacidad para formar esporas, realizado con Clustal Omega y editado con BioEdit para resaltar los aminoácidos conservados. **A.** Proteína YheA, **B.** Proteína YmcA, **C.** Proteína YibF, con su motivo rico en cisteínas en su extremo C-terminal después de la posición 130. En rojo se muestran los aminoácidos conservados en >90% de las secuencias. Los números de acceso para cada una de las secuencias están descritos en materiales y métodos.....54

Figura 3-9: Localización de los aminoácidos conservados en las proteínas YheA, YmcA y YlbF en bacterias Gram positivas. A y B. Proteína YheA, vista frontal y vista lateral, respectivamente. **C y D.** Proteína YmcA, vista frontal y vista lateral, respectivamente **E y F.** Proteína YlbF, vista frontal y vista lateral, respectivamente. Las flechas indican los aminoácidos conservados consecutivos en los alineamientos múltiples de las tres proteínas..... 56

Figura 4-1: Proceso para la generación de las cepas mutantes con la delección del gen *grp*. 1. Amplificación por PCR de los fragmentos fAB y fCD localizados corriente arriba y abajo del gen *grp*, respectivamente, 2. Generación de un fragmento de DNA que contiene los dos fragmentos anteriores por medio de PCR de solapamiento (“*overlap extension PCR*” [98]), (fAD), 3. Clonación del fragmento de DNA (fAD) en el plásmido pJET, 4. Transformación de pJET-AD en *Escherichia coli* XL1 blue, 5. Liberación y clonación del fragmento de DNA (fAD) en el plásmido pMAD, 6. Transformación de pMAD-AD en *Escherichia coli* XL1 blue, 7. Purificación y transformación del plásmido pMAD-AD en la cepa RN4220 de *Staphylococcus aureus*, 8. Purificación y transformación del plásmido pMAD-AD en las cepas NCTC8325-4, 15981 y 132, 9. Delección del gen *grp* en las tres cepas de interés y confirmación por PCR..... 60

Figura 4-2: Proceso para la generación de las cepas complementadas y sobre-expresadas. 1. Amplificación por PCR del ORF completo del gen *grp*, 2. Clonación del gen *grp* en el plásmido pJET, 3. Transformación del plásmido recombinante pJET-*grp* en *Escherichia coli* XL1 blue, 4. Liberación y clonación del gen *grp* en el plásmido pCN51, 5. Transformación del plásmido recombinante pCN51-*grp* en *Escherichia coli* XL1 blue, 6. Purificación y transformación del plásmido pCN51-*grp* en la cepa RN4220 de *Staphylococcus aureus*, 7. Purificación y transformación del plásmido pCN51-*grp* en las cepas 8325-4 *grp*⁻ (mutante) y 8325-4 (silvestre) y confirmación por PCR..... 64

Figura 4-3: Confirmación molecular de la delección del gen *grp* en las cepas de *Staphylococcus aureus*. **A.** colonias de la cepa NCTC8325 usando los primers *grp*_E y *grp*_F (Tabla 4-2), el tamaño esperado para las colonias mutantes (números en rojo) es de 1315 pb y para las colonias silvestres 1611 pb; **B.** colonias de la cepa 15981 usando los primers *grp*_E y *grp*_F y **C.** confirmación de las colonias mutantes de las tres cepas usando los primers *grp*_E y *grp*_AS, no se espera amplificación en las colonias mutantes (números en rojo) y el tamaño esperado para las colonias silvestres es de 935 pb..... 70

Figura 4-4: Efecto de la delección del gen *grp* en la formación de biofilm en *Staphylococcus aureus*. Evaluación de la capacidad de formar biofilm de las cepas silvestres NCTC8325 y 15981, y sus respectivas cepas mutantes y complementadas. Adicionalmente se muestra el resultado obtenido en dos colonias donde el proceso de mutagénesis se realizó de manera independiente para cada cepa. El asterisco denota que hay una significancia estadística entre los valores de las medias de los dos grupos de datos usando la prueba t-Student, a un intervalo de confianza del 95% y dos colas (n=6)..... 71

Figura 4-5: Efecto de la delección del gen *grp* en la hemólisis de eritrocitos por *Staphylococcus aureus*. Evaluación de la capacidad de hemólisis de la cepa silvestre NCTC8325, y su respectiva cepa mutante, complementada y sobre-expresada. A. Ensayo en placa de agar de la cepa silvestre (WT), mutada (Δgrp), sobre-expresada (WT+*grp*) y complementada (Δgrp +*grp*). C+: Control positivo (cepa de *Staphylococcus aureus* ATCC29213), C-: Control negativo (cepa de *Klebsiella pneumoniae*). B. Ensayo de microplaca. No se encontró una significancia estadística entre los valores de las medias de los dos grupos de datos usando la prueba t-Student, a un intervalo de confianza del 95%..... 72

Figura 4-6: Confirmación molecular de los plásmidos recombinantes de los genes *grp* y *sarX*. Amplificación por PCR de genes específicos de los diferentes plásmidos recombinantes generados mostrando la dirección del gen. **A y B.** Plásmidos pGEM-T1 y pET-CT1 (gen *grp* clonado dentro del vector pGEM-T y pET-CT, respectivamente), se confirmaron dos plásmidos purificados desde dos colonias transformadas diferentes. Carril 1 y 2: PCR con los primers T7-GP538 (antisentido del gen), y Carril 3 y 4: PCR con los primers T7-GP545 (sentido del gen). **C y D.** Plásmidos pGEM-T2 y pET-CT2 (gen *sarX* clonado dentro del vector pGEM-T y pET-CT, respectivamente), se confirmaron dos plásmidos purificados desde dos colonias transformadas diferentes. Carril 1 y 2: PCR con los primers T7-GP510 (antisentido del gen), y Carril 3 y 4: PCR con los primers T7-GP509 (sentido del gen). MP: Marcador de masa molecular..... 73

Figura 4-7: Seguimiento de la purificación de la proteína recombinante rQrp. A. Cromatograma del proceso de purificación de la proteína rQrp por medio de cromatografía de afinidad con la resina Ni-NTA. La línea negra indica el inicio del buffer de elución suplementado con imidazol. **B.** SDS-PAGE de las fracciones obtenidas de la cromatografía de afinidad. ET: Extracto total de proteínas con inducción con IPTG, SI: extracto total de proteínas sin inducción. MP: marcador de masa molecular.....74

Figura 4-8: Ensayos de retardamiento en gel de la proteína rQrp y la sonda de DNA de la secuencia palindrómica. Evaluación de la interacción de la proteína rQrp (80 ng) con la sonda de DNA (1500 fmoles) en presencia de diferentes cantidades de extracto total de proteínas (Fracción 40, Figura 4-7).....75

Figura 4-9: Comportamiento de la proteína recombinante rQrp en la CET. A. Cromatograma de la proteína rQrp en la CET. Los primeros números indican las fracciones recolectadas (0,5 mL) y su relación respecto a los valores que tendrían las fracciones de 1 mL (valores en negrilla). **B.** SDS-PAGE de las fracciones obtenidas de la CET de la proteína rQrp. MP: marcador de masa molecular.....76

Figura 4-10: Ensayos de retardamiento en gel de la proteína rQrp y la sonda de DNA de la secuencia palindrómica. Evaluación de la interacción de la proteína rQrp (80 ng) con la sonda de DNA en presencia de diferentes cantidades de extracto total de proteínas después de la CET (Fracción 30, Figura 4-4)..... 77

Lista de tablas

	Pág.
Tabla 2-1: Proteínas identificadas en las fracciones finales del proceso de purificación.....	30
<p>regulador transcripcional, ya que regula varios genes en <i>Staphylococcus</i> spp. dentro de los cuales los más destacados son el operón SarA y agr, dos súper reguladores que controlan la expresión de cientos de genes en la bacteria. Dos proteínas ortólogas de SarX han sido descritas en <i>S. epidermidis</i> [45], y un análisis de bioinformática nos permitió identificar otros ortólogos en <i>Staphylococcus lugdunensis</i> y <i>Staphylococcus capitis</i> (Figura 3-1), con porcentajes de identidad que varían desde el 64% al 67%. SarX pertenece a la familia de proteínas SarA, conformada por 11 proteínas pequeñas, con uno o máximo dos dominios, que actúan como reguladores transcripcionales (activadores o inhibidores) de múltiples genes en <i>Staphylococcus</i> spp (Tabla 3-1).</p>	
Tabla 3-1. Familia de proteínas SarA. Comparación de las características de la proteína SarX respecto a otras proteínas pertenecientes a la familia SarA, modificado de [77]...	37
Tabla 3-2: Comparación de la secuencia de la proteína hipotética (P1/P9) de <i>Staphylococcus aureus</i> con las proteínas con dominio Com_YlbF en <i>Bacillus subtilis</i> , YheA, YmcA y YlbF. %Id: porcentaje de identidad, %Si: porcentaje de similitud.....	39
Tabla 3-3: Valores de los parámetros de calidad de los modelos 3D generados y evaluados por medio de los programas PROCHECK, ERRAT, ProSA-web y Rampage.	47
Tabla 4-1. Cepas bacterianas y plásmidos usados en este capítulo.....	58
Tabla 4-2. Oligonucleótidos usados para la producción de las cepas mutantes, cepas complementadas y proteínas recombinantes.....	61

Introducción

Staphylococcus aureus es un patógeno oportunista de gran impacto clínico debido a que es responsable de una amplia gama de infecciones, desde leves como impétigo, celulitis, intoxicación alimentaria; hasta graves como choque tóxico, neumonía necrosante, endocarditis y sepsis [1-4]. *Staphylococcus aureus* se encuentra generalmente asociado al microbioma humano, donde su principal nicho ecológico son las fosas nasales (entre 20-25% de la población se encuentra colonizada permanentemente), pero también puede encontrarse en faringe, axila, ingle y manos [5, 6]. Como patógeno, *S. aureus* es un problema de salud pública nivel mundial [2, 7, 8], ya que tiene la capacidad de generar un amplio rango de infecciones, algunas de ellas mortales, en pacientes hospitalizados y en personas aparentemente sanas de la comunidad [2-4, 7-9]. Es el cuarto microorganismo causante de infecciones hospitalarias en el mundo [10] y Colombia no es la excepción [11-17]. En Estados Unidos las infecciones asociadas con *S. aureus* generan una tasa de mortalidad aproximadamente del 25% y un aumento en el tiempo de hospitalización y costos médicos para su tratamiento [18]. Estudios previos han demostrado una fuerte relación entre la colonización nasal y el incremento en el riesgo de la infección [6, 18-20]. Un tercio de la población de Estados Unidos, el Reino Unido, Japón y otros países se encuentra colonizado con este patógeno de forma asintomática [8].

El aumento en el uso de dispositivos, biomateriales e implantes ha resultado en un aumento de las infecciones bacterianas, principalmente causadas por *Staphylococcus spp.* [18, 21-23]. Las principales infecciones generadas por este género bacteriano son las asociadas a implantes ortopédicos, ventiladores, catéteres intravenosos, bombas de infusión, desfibriladores cardiacos, válvulas cardiacas, aspiradores, entre otros [18, 24]. La principal causa de asociación de *S. aureus* con estas infecciones es por su capacidad de adherirse a la superficie de éstos dispositivos y permanecer por largos periodos de tiempo en ellos a través de la formación de biofilm o biopelículas. El biofilm es un

agregado de bacterias que crecen embebidas en una matriz extracelular que ellas mismas producen. Este biofilm se produce debido a la adhesión del microorganismo a la superficie del implante y posterior acumulación de múltiples capas de células sésiles por la adhesión de una célula con otra por medio de la producción de una matriz extracelular conformada por diversas macromoléculas (polisacáridos, proteínas, lípidos, aminoácidos y ácidos nucleicos). Las bacterias formadoras de biofilm son muy difíciles de erradicar clínicamente debido a que ésta estructura le confiere protección contra la respuesta inmune del hospedero, y la acción de antibióticos y detergentes, por lo cual en muchos casos es necesario una intervención quirúrgica [18, 25-27]. Se estima que sólo en Estados Unidos, 10 millones de personas al año sufren infecciones asociadas a biofilm, debidas a implantes u otros procedimientos médicos invasivos. A pesar de que *S. aureus* es uno de los patógenos más significativos en términos de morbilidad y mortalidad anual, no se han obtenido grandes avances en la terapia o prevención en las infecciones asociadas a biofilm [18].

Se han reportado hasta el momento dos mecanismos independientes por los cuales *S. aureus* puede realizar la formación del biofilm. El primero, el más importante y estudiado, está mediado por el operón *ica*, conformado por los genes *icaA*, *icaB*, *icaC* e *icaD*, los cuales codifican las enzimas responsables de la síntesis del polisacárido de adhesión intercelular (PIA). Este polisacárido es el principal constituyente de la matriz extracelular del biofilm de *S. aureus*. [27-30]. El segundo, es un mecanismo independiente de *ica* que involucra la sobre-expresión de varias proteínas de adhesión como las proteínas de unión a fibronectina A y B (FnBPA, FnBPB) [28]. La formación del biofilm en *S. aureus* es aún un tema de reciente desarrollo, pocos estudios se han orientado a la búsqueda de nuevos factores genéticos relacionados con la formación de biofilm. Teniendo en cuenta la importancia clínica que tiene el biofilm en la persistencia y recurrencia de las infecciones causadas por *S. aureus*, es de gran importancia entender todos los factores moleculares involucrados en su formación y regulación. Comprender molecularmente cómo esta bacteria forma el biofilm, permitirá el desarrollo de mejores estrategias para su inhibición, erradicación y un mejor manejo de las infecciones recurrentes [18, 31].

En estudios de caracterización del operón *ica* en aislamientos colombianos de *S. aureus* realizados en el Laboratorio de Genética Molecular Bacteriana (LGMB) se identificó una secuencia palindrómica dentro del gen *icaA*, la cual es altamente conservada en cuatro especies de *Staphylococcus spp.* formadores de biofilm. A través de ensayos de

retardamiento en gel se determinó la existencia de proteínas que reconocen y se unen a esta secuencia palindrómica de forma específica. Debido a la importancia que tiene el operón *ica* en la formación de biofilm en *S. aureus* y al hallazgo de estas proteínas de unión al gen *icaA* se plantea la siguiente hipótesis de estudio:

Staphylococcus aureus posee proteínas que se unen al gen *icaA* a través del reconocimiento de una secuencia palindrómica, las cuales posiblemente pueden participar en la regulación de la formación o disrupción del biofilm

A partir de esta hipótesis se plantean las siguientes preguntas de investigación:

¿Qué proteínas se unen a esta secuencia palindrómica en el gen *icaA*?, ¿Estas proteínas podrían participar en la regulación de la formación o disrupción del biofilm en *Staphylococcus aureus*?

Para responder estas preguntas se emprendió un proceso robusto de purificación para el enriquecimiento de las proteínas que interactúan con la secuencia palindrómica del gen *icaA*, las cuales fueron identificadas por medio de espectrometría de masas. Después de su identificación, se realizó una detallada caracterización “*in silico*” y se evaluó la participación de una de estas proteínas en la formación de biofilm a través de la delección del gen que la codifica. A continuación, esta información será presentada en 4 capítulos, el primero contiene los conceptos fundamentales del modelo de estudio *S. aureus* y los mecanismos moleculares involucrados en la formación de biofilm. El segundo muestra cómo fueron identificadas las proteínas de unión a la secuencia palindrómica del gen *icaA*. El tercero abarca la caracterización “*in silico*” de las proteínas identificadas y finalmente en el cuarto se muestra el impacto en la formación del biofilm (y hemólisis) de la bacteria cuando los genes que codifican las proteínas identificadas son suprimidos de su genoma.

OBJETIVOS

Objetivo general

Identificar y caracterizar una proteína de unión a la secuencia palindrómica del gen *icaA* y determinar su posible participación en la formación del biofilm en *Staphylococcus aureus*.

Objetivos específicos

- Purificar e identificar una proteína de unión a la secuencia palindrómica del gen *icaA* de *Staphylococcus aureus*.
- Describir las características genéticas, moleculares y estructurales del gen que codifica para la proteína de unión a la secuencia palindrómica del gen *icaA*, así como las características genéticas de la proteína de unión.
- Evaluar la participación de la proteína de unión a la secuencia palindrómica del gen *icaA* en la formación del biofilm en *Staphylococcus aureus*.

1. Capítulo 1

1.1 Generalidades de *Staphylococcus aureus*

Staphylococcus aureus es una bacteria Gram positiva, con forma de coco, séstil, no formador de esporas, aerobio y anaerobio facultativo y productor de las enzimas coagulasa y catalasa. Este microorganismo puede resistir temperaturas de hasta 50°C, altas concentraciones de sal y la desecación. Sus colonias en agar son usualmente largas (0.5 a 1.5 μm de diámetro) y pueden presentar pigmentación de color amarillo crema a naranja debido a la producción de β -carotenos [9]. En promedio el tamaño de su genoma es de 2.900.000 pares de bases con un bajo porcentaje de guanina y citosina (~30%) y un contenido aproximado de 3000 genes [17, 32].

Staphylococcus aureus es considerado como una bacteria del microbioma humano, pero puede comportarse como un patógeno oportunista, ya que puede estar colonizando varios sitios del cuerpo sin causar enfermedad, pero también puede causar infecciones leves y severas. El principal nicho anatómico humano son las fosas nasales, en donde se encuentra con mayor frecuencia, pero también se puede encontrar en faringe, cuello, axila, periné y manos. La condición de portador es referida como colonización para distinguirla de la infección. La colonización por *S. aureus* puede darse en tres patrones distintos en la población: portador persistente (30% de frecuencia), personas que pueden portar la bacteria por periodos de tiempo de hasta 4 años; portador intermitente (30%), con una duración media de 14 días; y portador transitorio (40%), personas con la bacteria por lapsos muy cortos de tiempo. Existe evidencia que ser portador es un importante factor de riesgo para adquirir infecciones invasivas por *S. aureus* pero se conoce poco sobre los factores de la bacteria y del hospedero que promueven la colonización. Varios reportes han demostrado que por lo menos el 80% de las infecciones hospitalarias tienen su origen en portadores nasales [5, 6].

1.2 Formación, estructura y adherencia del Biofilm o biopelícula en Gram positivos.

El concepto de biofilm o biopelícula ha cambiado a través del tiempo a medida que los avances tecnológicos han permitido estudiarlo con mayor profundidad. La definición más aceptada actualmente es: el biofilm o biopelícula es una comunidad microbiana altamente estructurada, compleja, compuesta ya sea por células bacterianas de la misma especie o de diferentes especies, inmersas dentro de una matriz de biopolímeros y adheridas a un sustrato o interfase inerte o viviente [33, 34]. Aunque *Bacillus subtilis* y *Bacillus cereus* pueden formar biofilm en la interfase de soluciones acuosas con el ambiente, facilitando su esporulación [35]. Las bacterias cuando están formando biofilm, exhiben un fenotipo alterado en cuanto a su metabolismo, crecimiento, replicación, síntesis de ácidos nucleicos, expresión génica, producción de proteínas y biosíntesis de la pared celular [18, 25, 33, 36-40].

En el biofilm hay una considerable disminución del crecimiento bacteriano y una abundante secreción de polímeros orgánicos extracelulares, los cuales inducen a la formación de capas mucosas sobre las superficies. Éstas biopelículas se desarrollan prácticamente en todas las superficies inmersas en ambientes acuosos naturales, tanto biológicos (plantas acuáticas y animales) como abióticos (cemento, metal, plásticos, piedras, biomateriales, entre otros). El crecimiento del biofilm aumenta en sistemas acuosos que fluyen, donde se provee al microorganismo de un aporte regular de nutrientes (como es el caso por ejemplo de implantes ortopédicos). Algunos microorganismos pueden crear sus propios microambientes y nichos mediante la formación de biofilm, incluso sin disponer de entornos físicos estructurados [33, 36]. Muchas bacterias que se encuentran en ambientes naturales están organizadas en biofilm, formando poblaciones multicelulares altruistas, comunidades sésiles que constituyen un gran componente de su biomasa global [36, 41].

En términos generales, se ha propuesto que la formación del biofilm ocurre en cuatro etapas (Figura 1-1): 1. **Unión** (inespecífica) inicial de las bacterias a la superficie a través de interacciones hidrofóbicas, electrostáticas y fuerzas de Lifshitz - van der Waals (utilizando componentes de la pared bacteriana, como los ácidos teicóicos y el peptidoglicano). Esta unión puede ser reversible o irreversible. 2. **Formación de**

monocapas de bacterias mediante la unión célula-célula, generalmente a través de proteínas de adhesión transmembrana. 3. **Maduración del biofilm** mediante la formación de agregados celulares y la producción de sustancias poliméricas extracelulares (SPE) constituyentes principales de la matriz y 4. **Desensamblaje parcial** del biofilm, liberación y migración de algunas células, las cuales pueden colonizar otro sitio e iniciar nuevamente el ciclo [37-40, 42] (Figura 1-1).

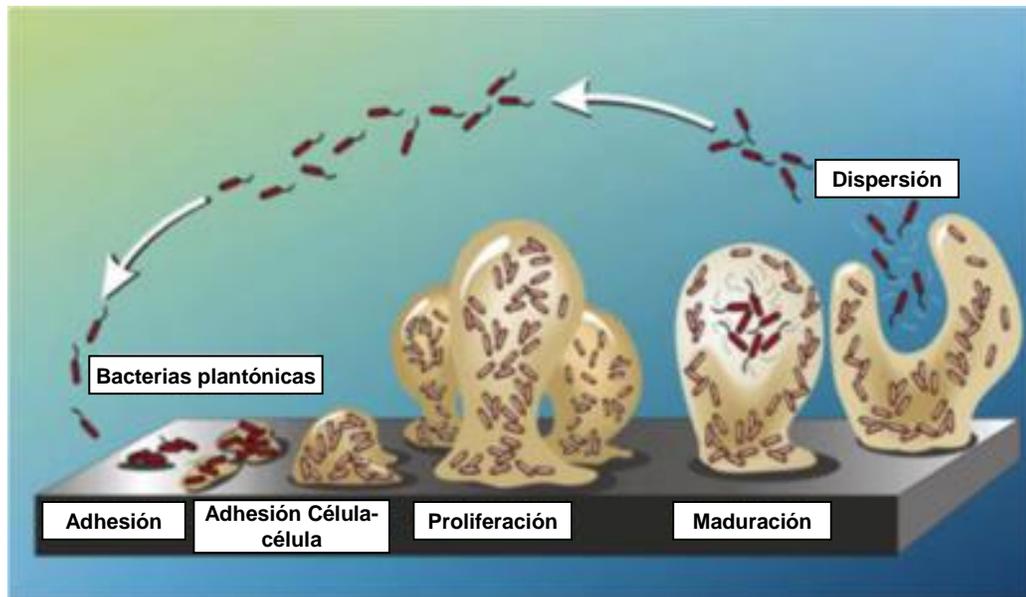


Figura 1-1: Esquema de los 4 pasos propuestos para la formación de biofilm en bacterias. 1. Las bacterias planctónicas se adhieren a la superficie (inerte o viviente), 2. Comienza la interacción célula-célula y la producción de la matriz extracelular (Exopolisacárido-EPS), 3. Proliferación y multiplicación bacteriana con una maduración del biofilm y 4. Disrupción del biofilm y dispersión de las bacterias a otros sitios. Imagen tomada de MicrobeWiki y modificada de acuerdo a Otto M, 2009. [41].

La unión inicial a la superficie se realiza a través de componentes de la pared celular como los ácidos teicóicos, en colaboración con proteínas de superficie. Estas proteínas facilitan la formación de agregados celulares (unión célula-célula) y luego se inicia la producción de SPE. *Staphylococcus epidermidis* y *Staphylococcus aureus* producen PIA, el cual es el constituyente principal de la matriz extracelular del biofilm maduro [37, 41-43]. Finalmente, si las condiciones del medio son favorables o los estímulos medio-ambientales cambian, algunas células dentro del biofilm activan un proceso de dispersión

y ruptura de la matriz extracelular mediado principalmente por la producción de proteasas y modulinas solubles en fenol (PSM). Estas últimas actúan como agentes dispersores por su capacidad de bloquear las interacciones hidrofóbicas entre los diferentes componentes del biofilm [23, 41]. La activación de la dispersión del biofilm esta mediada principalmente por un sistema de *quorum sensing* o de comunicación bacteriana, que en bacterias Gram positivas se realiza a través del operón Agr.

1.2.1 Biofilm en *Staphylococcus aureus*

Staphylococcus aureus tiene una gran adaptación para la formación del biofilm por medio de la disminución de procesos celulares básicos como la biosíntesis de ácidos nucleicos, proteínas y pared celular. Estos cambios regulatorios podrían explicar la limitada actividad de varios antibióticos como penicilinas, aminoglicósidos y quinolonas, cuyo blanco son células en crecimiento. Los componentes de la matriz tienen la capacidad de disminuir la penetración de varios antibióticos a través del biofilm, reduciendo así su concentración dentro de la bacteria. Estas características fisiológicas y arquitectónicas del biofilm son la base de la resistencia a los antibióticos y a los mecanismos de defensa del hospedero [33, 41]. El biofilm en *S. aureus* permite que estas bacterias se adhieran a superficies como dispositivos protésicos, catéteres e instrumentos médicos. En estas biopelículas las células se encuentran incluidas en una matriz de exopolisacáridos o proteínas, las cuales se forman sobre los biomateriales y se asocian con infecciones recurrentes. El exopolisacárido parece inhibir la quimiotaxis de los neutrófilos, la fagocitosis y la acción de antimicrobianos como vancomicina y teicoplanina [18, 22, 33, 44]. El implante, se cubre con las proteínas de la matriz extracelular del hospedero y provee una superficie adecuada para la unión bacteriana y posterior formación del biofilm [18].

Actualmente se han reportado dos mecanismos por los cuales *S. aureus* puede realizar la formación del biofilm. El primero y más importante, esta mediado por los genes codificados en el operón *ica* que codifica las enzimas responsables de la síntesis y transporte de PIA (Figura 1-2). Este operón se mantiene y se expresa en casi todos los aislamientos de *S. aureus* [25, 28, 45, 46]. El segundo y menos frecuente, es un mecanismo independiente de *ica* que involucra la sobreexpresión de varias proteínas de adhesión como las proteínas de unión a fibronectina A y B (FnBPA, FnBPB) [25, 28, 45,

46]. Estas proteínas no se encuentran en todos los aislamientos de *S. aureus*, lo cual sugiere que es un mecanismo adquirido a través del tiempo y no es constitutivo de esta bacteria. Vergara-Irigaray y colaboradores en el 2009 reportaron una cepa de *S. aureus* con la capacidad de formar dos tipos diferentes de biofilm dependiendo del estímulo externo al que se encontraba sometida: uno inducido por estrés osmótico (altas concentraciones de NaCl), conformado por una matriz densa de polisacáridos, la cual está constituida principalmente por PIA y es dependiente de la expresión del operón *ica*; y el segundo inducido por condiciones de acidez (dado por concentraciones moderadas de glucosa), compuesto principalmente por proteínas formando un agregado celular denso. La existencia de estos dos mecanismos alternos podría implicar que algunos aislamientos de *S. aureus*, con los dos mecanismos, puede desarrollar un tipo de matriz particular de biofilm dependiendo del ambiente donde crece [34].

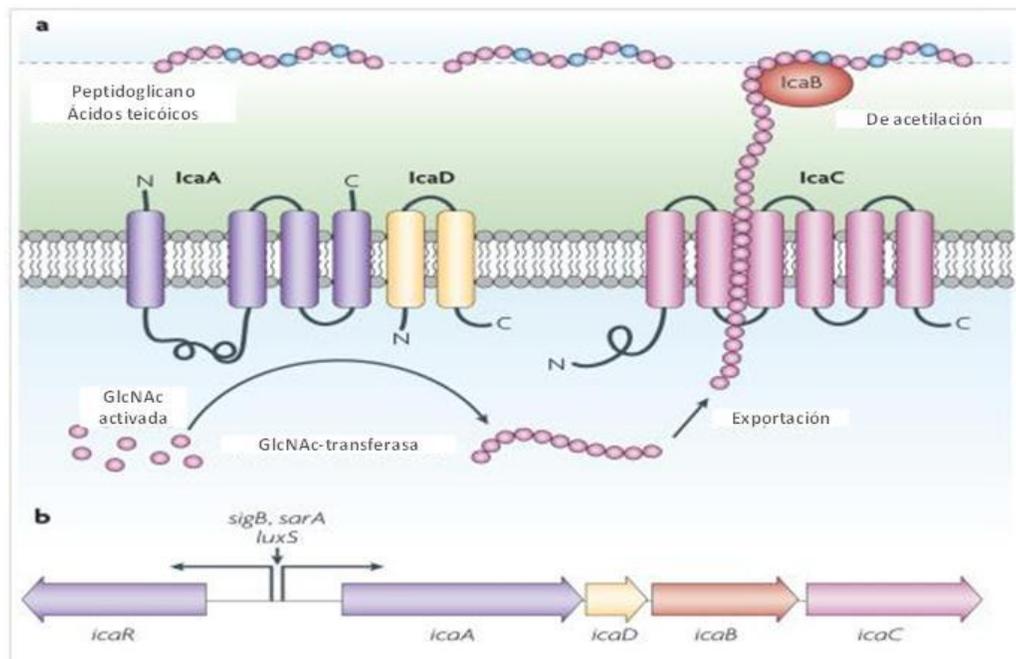


Figura 1-2: Estructura del operón *ica* y asociación de la función de las proteínas codificadas por sus genes *icaA*, *icaB*, *icaC* e *icaD*. Modificado de Otto M, 2009 [41].

PIA es un glicosaminoglicano unido por enlaces β -1,6 de al menos 130 residuos de 2-deoxi-2-amino-D-glucopiranosil, de los cuales 80% al 85% están N-acetilados y el resto

se encuentran cargados positivamente [25, 27, 28, 45] [29]. Este polímero no se asemeja a otra poli-N-acetilglucosamina encontrada en la naturaleza debido a su enlace β -1-6. Sin embargo, ha sido encontrado en otros microorganismos como *Yersinia pestis* y *Escherichia coli* [47]. El operón *ica* está conformado por los genes *icaA*, *icaD*, *icaB*, *icaC* e *icaR*; fue identificado inicialmente en *S. epidermidis* y ha sido ampliamente estudiado en esta especie; su mecanismo de acción es muy similar al que se presenta en *S. aureus* [29]. Aunque este operón ha sido identificado en otras especies de *Staphylococcus spp.*, su mecanismo de acción no ha sido demostrado experimentalmente. La comparación de la secuencia de aminoácidos de las proteínas codificadas por este operón en estas dos especies muestra que tienen entre 59% a 78% de identidad y un porcentaje de similitud de 79% a 89% [27]. Los genes *icaA*, *icaD*, *icaB*, *icaC* son co-transcritos a partir del promotor del gen *icaA*; la inactivación de cualquier gen del operón *ica* inactiva la formación de biofilm. El gen *icaA* codifica para una proteína de 412 aminoácidos con una masa molecular calculada de 47,7 kDa y un punto isoeléctrico de 8.01 [48], contiene 43.6% de aminoácidos hidrofóbicos, una región hidrofóbica cerca al extremo N-terminal y tres cerca al C-terminal. La presencia de las cuatro hélices transmembranales es un indicativo de su localización transmembranal [29]. La secuencia de IcaA muestra similitud con las enzimas que polimerizan polisacáridos, como la N-acetilglucosaminiltransferasa de *Rhizobium meiloti* que está involucrada en la síntesis de lipo-quitina [30]. Gerke y colaboradores en 1998 demostraron la actividad N-acetilglucosaminiltransferasa de IcaA y el uso de UDP-N-acetilglucosamina como sustrato para formar el polisacárido PIA, el cual se encuentra localizado en la superficie celular, de acuerdo a estudios realizados con anticuerpos específicos para PIA [29, 30].

La co-expresión de *icaA* con *icaD* incrementa significativamente su actividad enzimática. El gen *icaD* posee 306 nucleótidos y se transcribe en la misma dirección que los genes *icaABC*. El polipéptido que se deduce de *icaD* consiste de 101 aminoácidos con un peso molecular calculado de 11,9 kDa y un *pI* de 9.88, contiene 54.5% de aminoácidos hidrofóbicos y se predicen dos hélices transmembranales en su conformación [30]. No se han encontrado similitudes con proteínas conocidas en las bases de datos; la inactivación de *icaD* causa pérdida de la agregación celular y producción de PIA en *S. carnosus*. El gen *icaD* se encuentra localizado entre *icaA* e *icaB* y se solapa entre los dos genes, 37 nucleótidos al final de *icaA* y 4 nucleótidos al inicio de *icaB*. Sólo cuando *icaA*

e *icaD* son expresados con *icaC*, la cadena de oligómero es reconocida con un anticuerpo específico para PIA, por lo cual se asume que ésta última proteína es requerida para la correcta formación y secreción del polisacárido [30].

IcaC es una proteína hidrofóbica de 355 aminoácidos con una masa de 42,0 kDa y un pl de 8.54, contiene 56.5% de aminoácidos hidrofóbicos y seis hélices transmembranales distribuidas en su secuencia. No se ha encontrado homología de IcaC con otras proteínas conocidas. Al parecer es una proteína integral de membrana y puede estar involucrada en la translocación de PIA a través de la membrana citoplasmática [29]. De acuerdo a esto, se propone una cercana interacción entre IcaA, IcaD e IcaC y representan una nueva combinación de proteínas entre las β -glicosiltransferasas [30]. El codón de inicio de *icaC* se sobrelapa con el final de *icaB* por 14 nucleótidos. Corriente abajo de *icaC* hay una secuencia independiente de rho, esta es la única secuencia de terminación en el operón y parece ser el único sitio de terminación lo que indica que *icaA*, *icaB*, *icaC* y *icaD* son co-transcritos a partir del promotor de *icaA* [29].

IcaB está codificada por el gen *icaB* y consiste de 289 aminoácidos con una masa molecular calculada de 33,3 kDa y un pl de 9.59; el 33.1% de los aminoácidos son hidrofóbicos; contiene una secuencia señal típica con un motivo AXA (Ala-Gln-Ala) para peptidasa señal I. Se ha propuesto que la molécula se secreta, ya que se encuentra en el sobrenadante del cultivo celular. Se propone que actúa como una deacetilasa debido a que en el polisacárido PIA un 15% a 20% de los residuos de azúcar se encuentran desacetilados. En contraste, se sugiere que puede actuar como una lectina extracelular involucrada en el entrecruzamiento de PIA y en la interacción célula-célula [29]. La expresión del operón *ica* está regulada por el gen *icaR*, el cual se encuentra localizado corriente arriba del gen *icaA* y tiene homología con las proteínas TetR de reguladores transcripcionales [49]. El gen *icaR* tiene un papel central en la regulación ambiental de la expresión del operón *ica*. Al modificar el ambiente de la bacteria por la adición de NaCl o etanol se activa este operón por un mecanismo dependiente de *icaR* [50]. Jefferson y colaboradores (2004) demostraron que la proteína IcaR de *S. aureus* se une a la región promotora del operón *ica*, y actúa como un regulador negativo de su transcripción [50]. La expresión del gen *icaR* es finamente regulada en varios niveles, incluyendo nuevos mecanismos a nivel de traducción de su mRNA, involucrando interacciones de sus regiones UTR (regiones no traducibles) 5' y 3' [51]. Actualmente, se han identificado

múltiples reguladores transcripcionales del operón *ica*, como por ejemplo TcaR, miembro de la familia MarR de reguladores transcripcionales que controla la expresión del locus asociado a teicoplanina (*tca*) [48, 52].

La expresión temporal de varios factores de virulencia en *S. aureus* ha sido demostrado estar bajo el control de varios loci genéticos tales como *agr* y *sarA* [46]. SarA (Staphylococcal accessory regulator A) es un regulador global en *S. aureus* y se ha reportado que su activación es requerida para la transcripción del operón *ica*, la producción del PIA y la formación del biofilm [53]. El locus SarA, codifica para una proteína de unión de 14.5 kDa que se une al promotor del operón *ica* en *S. aureus* y *S. epidermidis* y actúa como un regulador positivo de la expresión del operón *ica* por un mecanismo independiente de *icaR* [43, 46] (Figura 1-3). *Agr* (accessory gene regulator) es un operón bifuncional conformado por dos promotores divergentes, P2 y P3, los cuales activan la transcripción de dos RNA, denominados como RNAII y RNAPIII, el primero contiene los genes *agrA*, *agrB*, *agrC* y *agrD*, encargados de la síntesis y respuesta de un péptido autoinductor; y el segundo cumple dos funciones, sirve de RNA antisentido para la regulación de varios genes en la bacteria y además alberga el gen para la δ -hemolisina. En contraste con SarA, *agr* inhibe la formación del biofilm [54, 55].

Como se mencionó anteriormente, *S. aureus* puede formar biofilm por un mecanismo independiente del operón *ica* utilizando varias proteínas de superficie pertenecientes a la familia MSCRAMMs (Microbial Surface Components Recognizing Adhesive Matrix Molecules) que han sido asociadas a la unión de proteínas del hospedero [34, 56]. En 2008 O'Neil y colaboradores reportaron la formación de biofilm en un aislamiento de *S. aureus* por un mecanismo independiente del operón *ica*, y se evidenció que las proteínas de unión a fibronectina FnBPA y FnBPB tenían la capacidad de promover la acumulación intercelular y el desarrollo del biofilm [28]. En este caso la presencia de PIA no fue esencial para el desarrollo de biofilm. Estas proteínas FnBPA y FnBPB median la unión de la bacteria con las superficies y con los tejidos del hospedero [28, 57]. La formación de biofilm por la producción de estas proteínas ha sido asociada a la acidificación del medio (por ejemplo en presencia de altas concentraciones de glucosa) [34]. Estas proteínas han sido relacionadas con la capacidad de *S. aureus* para adherirse a válvulas cardíacas, producir endocarditis y en la internalización en células epiteliales [58].

Otras proteínas de adhesión que han sido relacionadas con la formación de biofilm en *S. aureus* son las proteínas Bap (Biofilm-associated proteins), las cuales tiene un alto peso molecular, múltiples dominios, algunos de ellos con repetidos en tándem, y asociadas a la superficie bacteriana [59]. Sin embargo, estas proteínas han sido únicamente encontradas en aislamientos asociados a mastitis en bovinos. Las proteínas Bap promueven la unión primaria a las superficies inertes, independientemente de PIA, y la adhesión intercelular por su capacidad de formar dímeros y fibras amiloides. Dos bacterias pueden unirse entre sí a través de una fibra de Bap [46, 57, 60]. Además, se han reportado las proteínas Aap/SasG (proteínas asociadas a acumulación de *S. epidermidis*/proteína G de superficie de *S. aureus*) [61], las cuales promueven la adherencia bacteriana a las células nasales epiteliales y la formación del biofilm independiente de *ica* [61]. Por lo tanto, éstas proteínas son importantes factores de virulencia en las infecciones relacionadas con instrumentos biomédicos particularmente en las cepas resistentes a los antibióticos y pueden ser importantes blancos terapéuticos [28, 34].

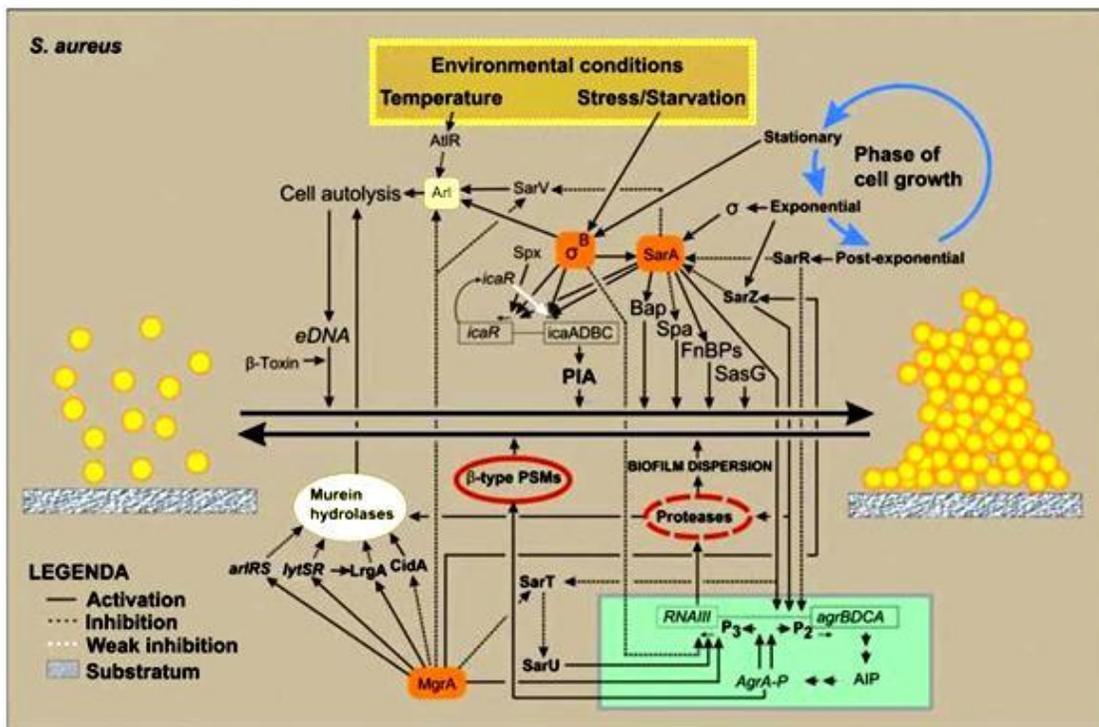


Figura 1-3: Esquema de la compleja red de señalización que regula la formación y disrupción del biofilm en *Staphylococcus aureus*. Tomado de Arciola y colaboradores [23].

De acuerdo a lo descrito por Archer y colaboradores, otro componente importante en la formación de biofilm en *S. aureus*, es el DNA extracelular (eDNA) (Figura 1-3). Se ha observado que la lisis celular por parte de varias proteasas, libera el eDNA y permite la agregación celular en etapas tempranas de la formación de biofilm [18].

1.2.2 Dispersión del biofilm

La bacteria debe alternar entre la formación de biofilm y un estado de crecimiento planctónico, con el fin de contribuir a la dispersión y colonización de nuevos órganos u hospederos e iniciar un nuevo proceso infeccioso [18, 33]. Para esto debe utilizar diferentes mecanismos para dispersarse del biofilm de acuerdo al tipo de moléculas que lo compongan; ya sea degradación enzimática de las proteínas que componen el biofilm (p.ej. FnBPs) por la acción de proteasas o degradación enzimática de exopolímeros (PIA) o interrupción de enlaces no covalentes por moléculas con actividad detergente [18, 41]. En el proceso de dispersión del biofilm en *S. aureus*, formado principalmente por proteínas, actúa el sistema de *quorum-sensing*, el cual está codificado por el operón *agr* (Figura 1-3). La molécula de comunicación que lo produce es un péptido autoinductor (AIP-Auto inducing peptide), el cual cuando alcanza una concentración determinada, se une a un receptor de histidin quinasa, lo que causa el inicio de una cascada de regulación que controla la expresión de varios genes, particularmente factores de virulencia como proteasas, hemolisinas y toxinas. El mecanismo de dispersión es mediado entonces por proteasas extracelulares que se secretan cuando se activa el *quorum sensing*. Cuando la bacteria se desprende del biofilm presenta altos niveles de *agr*, y cuando la bacteria se encuentra en biofilm, tiene este sistema reprimido [62]. En cuanto a la dispersión del biofilm formado por el exopolisacárido PIA, no se tiene información clara al respecto; ya que no se ha encontrado una enzima dedicada a la hidrólisis de PIA; sin embargo, se propone que en *S. epidermidis* algunas moléculas con características de detergentes pueden interrumpir los enlaces no-covalentes como las interacciones electrostáticas e hidrofóbicas, en los enlaces catiónicos del PIA y en la superficie aniónica de los polímeros, o entre las partes hidrofóbicas de la superficie bacteriana. Las modulinas anfipáticas solubles en fenol (PSMs) que poseen las δ -toxinas en *S. epidermidis* y *S. aureus* parecen tener esta acción [63].

Estos progresos en elucidar el papel del polisacárido de adhesión intercelular PIA codificado en el operón *ica* en la formación del biofilm en *Staphylococcus spp.* han contribuido en gran manera al entendimiento de la patogénesis de las infecciones relacionadas con biofilm. Sin embargo, el conocimiento actual sobre cómo el operón *ica* y la biosíntesis de PIA son regulados no es suficiente, y aún permanecen muchas preguntas sin resolver [43]. Es por esta razón que es de gran interés encontrar nuevos determinantes genéticos en la formación de biofilm en *S. aureus*. El control de biofilms ha sido una preocupación constante en centros hospitalarios; debido a que las bacterias lo usan como mecanismo de supervivencia y diseminación, y son causa frecuente de infecciones recurrentes, de difícil tratamiento y aumento de la morbi-mortalidad de pacientes. Las estrategias que se proponen para su erradicación es el desarrollo de nuevos antibióticos capaces de penetrarlos, moléculas que interfieran con su formación a través de diversas vías como la inhibición de la comunicación celular, degradación de la matriz extracelular y bloqueo de las proteínas de adhesión [18, 25, 46, 64, 65].

1.2.3 Secuencias palindrómicas y regulación génica

Las secuencias palindrómicas son fragmentos de DNA que tienen la misma secuencia en su hebras sentido y antisentido, leídas desde sus extremos 5' a 3'. Estas secuencias han sido asociadas a varios procesos celulares fundamentales dentro de las cuales se encuentran: 1. formación de estructuras secundarias como "hairpins" o cruciformes, los cuales pueden actuar como represores de la transcripción. 2. Sitios de reconocimiento de varios tipos de proteínas como endonucleasas de restricción que hidrolizan el enlace fosfodiéster en uniones de nucleótidos específicas; metilasas que adicionan grupos metilos en las cadenas de DNA recién sintetizadas (proceso reportado en bacterias), y unión de factores de transcripción involucrados en la regulación de la expresión génica. A nivel bacteriano, se han identificado varias secuencias palindrómicas relacionadas con la regulación de la expresión de varios operones, generalmente estas secuencias se encuentran localizadas en regiones específicas de regulación corriente arriba del operón y corriente abajo de una proteína reguladora (la cual frecuentemente actúa como represora), la cual posee dominios de reconocimiento que le permiten unirse a estas secuencias evitando la transcripción del operón. El papel de las secuencias palindrómicas ha sido ampliamente estudiado en varios operones como por ejemplo en el

operón *lac* y en el operón *arc*, cuyos represores reconocen secuencias palindrómicas de 22 pb y 20 pb, respectivamente [66]. Aunque, las secuencias palindrómicas han sido encontradas principalmente corriente arriba de los operones, existen algunos reportes de la identificación de secuencias palindrómicas corriente abajo, ejerciendo un papel de reconocimiento para proteínas represoras pero también activadoras del gen donde se encuentra. Por ejemplo Dattananda y colaboradores en 1991 reportaron la identificación de un elemento regulatorio dentro del operón *proU* relacionado con la repuesta a cambios osmóticos de *Escherichia coli* [67]. Mullin y Newton en 1993 reportaron la existencia de secuencias insertadas dentro de elementos *ftr*, los cuales fueron asociados con la transcripción diferencial de los genes *flaN* y *flaG* en *Caulobacter crescentus* [68]. Finalmente Mitra y colaboradores in 2005 analizando la regulación del operón *nifLA* en *Azotobacter vinelandii*, identificaron una secuencia regulatoria dentro del primer gen, la cual estaba asociada con un aumento en su transcripción [69].

2. Capítulo 2. Identificación de proteínas de unión al gen *icaA* en *Staphylococcus aureus*

2.1 Materiales y métodos.

2.1.1 Cepas bacterianas.

El proceso de identificación de las proteínas de unión al gen *icaA* se realizó en la cepa de *Staphylococcus aureus* USA300-0114 por las siguientes razones: tiene alta capacidad de formación de biofilm, es resistente a múltiples familias de antibióticos, es un clon pandémico que se originó en la comunidad (generalmente las cepas resistentes a los antibióticos se generan en el ambiente hospitalario), ha tenido una amplia diseminación ya que ha sido identificado en varios continentes, tiene un alto impacto clínico y su genoma ya fue secuenciado y es utilizado como referencia en esta bacteria. En Colombia circula un variante estrechamente relacionada con este clon USA300, la cual es responsable de casi la totalidad de infecciones por *Staphylococcus aureus* resistente a metilicina en la comunidad y en pacientes hospitalizados.

2.1.2 Obtención de bacterias en crecimiento planctónico y formando biofilm.

Para obtener las bacterias en crecimiento planctónico, una colonia de la cepa USA300 fue crecida en 5 mL de caldo BHI a 37°C durante 16 horas para alcanzar su fase estacionaria de crecimiento; posteriormente esta suspensión bacteriana fue centrifugada a 1100 g y 4°C durante 7 minutos, el sobrenadante fue desechado y las bacterias usadas inmediatamente para obtener el extracto de proteínas. Las bacterias formando biofilm se obtuvieron de acuerdo al protocolo estandarizado en el Laboratorio de Genética Molecular Bacteriana de la siguiente manera: una colonia fue crecida en 5 mL de caldo

BHI a 37°C durante 3 horas; a partir de esta suspensión se realizó una dilución 1:50 (equivalente a un estándar de turbidez 0.5 de McFarland) en 20 mL de caldo BHI suplementado con glucosa al 1% en una caja de poliestireno de 9 cm de diámetro e incubada a 37°C por 24 horas [70]. El sobrenadante fue descartado y con cuidado las bacterias adheridas fueron lavadas tres veces con agua destilada estéril. Luego, las bacterias formando biofilm fueron desprendidas mecánicamente y recolectadas en un tubo de 2 mL, nuevamente lavadas y usadas para obtener el extracto de proteínas.

2.1.3 Obtención del extracto de proteínas totales de *S. aureus*.

Las bacterias en crecimiento planctónico y formando biofilm fueron tratadas con lisozima (1 mg/mL) y lisostafina (50 µg/mL sigma aldrich®) a 37°C durante 1 hora, luego se pusieron en hielo, se adicionaron 500 µL de buffer de sonicación (Tris-HCl 0,1mM, pH=7,0) y se sometieron a sonicación en un homogeneizador ultrasónico (Socincs, vibra cell VCX130®), con ciclos de 2 minutos a una amplitud de 30% con un pulso inicial de 30 segundos y final de 15 segundos cada uno; este ciclo se repitió tres veces para las bacterias en crecimiento planctónico y 5 veces para las bacterias formando biofilm. El extracto fue centrifugado a 10.000 g a 4°C por 10 minutos, el sobrenadante fue transferido a otro tubo y 10 µL de un coctel de inhibidor de proteasas fue adicionado (Thermo Scientific™ Halt™ Protease Inhibitor-100X). La cuantificación de cada extracto fue realizado por su absorción a 280 nm en NanoDrop, por el método de Bradford y SDS-PAGE.

2.1.4 Detección y seguimiento de la proteína de unión al gen *icaA* por retardamiento en gel.

La detección de las proteínas de unión al gen *icaA* se realizó por medio del ensayo en retardamiento en gel (EMSA por sus siglas en inglés) usando una sonda de DNA de doble cadena que simula la región conservada de este gen. Esta sonda tiene una longitud de 22 nucleótidos (AAATTATTATCATTAATGATGG), con una biotina unida covalentemente a su extremo 5' para permitir su detección. Los extractos de proteínas a evaluar se incubaron con 1500 fmoles (valor obtenido a través de un proceso de estandarización) de esta sonda a temperatura ambiente durante 30 min, luego fueron sometidos a electroforesis en un gel de poliacrilamida al 7% en condiciones no

denaturantes, transferidos a una membrana de Nylon Hybond N⁺ (Amersham®) y detectados por quimioluminiscencia utilizando el Kit “Chemiluminescent Nucleic Acid Detection Module” (Thermo Scientific) de acuerdo a las especificaciones del fabricante y el foto documentador ChemiDoc® (Biorad).

2.1.5 Identificación de las proteínas de unión al gen *icaA*.

La purificación de proteínas es un paso fundamental y necesario para poder identificarlas. El proceso de purificación de la proteína de unión al gen *icaA* se realizó por medio de un tratamiento con (NH₄)₂SO₄ y luego cromatografía rápida líquida de proteínas (FPLC BioLogic DuoFlow™ de Biorad). El seguimiento de las proteínas de unión a la secuencia palindrómica en el extracto total y en las fracciones obtenidas de las diferentes cromatografías se determinó por EMSA y SDS-PAGE [71].

Para obtener una cantidad detectable de las proteínas de interés, se partió de 25 mL de una suspensión bacteriana de la cepa USA300, que generaron 5 mL de extracto total de proteínas. Esta cantidad de extracto fue tratada con una solución de (NH₄)₂SO₄ hasta alcanzar un porcentaje de saturación del 37% y centrifugado a 13.000 g durante 15 minutos a 4°C; posteriormente, el volumen total del sobrenadante fue separado por medio de 10 ensayos de cromatografía de exclusión por tamaño (CET) usando la columna Superose 6 HR10/30 (Pharmacia®) (500 µL de volumen de inyección por ensayo) y de acuerdo a las condiciones óptimas de separación y parámetros operacionales previamente optimizados (Anexo A). Las fracciones positivas para las proteínas de interés identificadas por EMSA en cada uno de los 10 ensayos fueron inyectadas en un solo ensayo de cromatografía de intercambio aniónico (CIA) a través de múltiples inyecciones utilizando la columna Mono Q HR 5/5 (Pharmacia®) y UNO Q1 (Biorad®), la elución de las proteínas retenidas se realizó mediante un gradiente lineal con una solución de Tris-HCl 20mM pH 7.5 suplementado con NaCl 1M (Anexo A).

De las fracciones positivas de la CIA y los controles (fracciones negativas), 100 µL fueron concentrados por vacío hasta un décimo de su volumen y sus proteínas separadas por SDS-PAGE. La identificación de las proteínas de interés en las fracciones positivas se realizó de dos formas, la primera por medio del método de “secuenciación total de la muestra”, para lo cual la muestra completa de proteínas fue corrida poca distancia dentro del gel de poliacrilamida, teñida con Coomassie R-250, cortada y secuenciada. En el

segundo caso, las proteínas fueron separadas totalmente por SDS-PAGE al 12%, teñidas con Coomassie R-250, cortadas individualmente y secuenciadas. En los dos casos, las proteínas fueron posteriormente tratadas con tripsina y los péptidos generados identificados por espectrometría de masas MALDI-TOF-TOF. A partir de los espectros de fragmentación obtenidos se obtuvieron las secuencias de cada uno de los péptidos y la identificación de sus proteínas correspondientes en *S. aureus* por medio del programa MASCOT (Matrixscience, UK). Finalmente las proteínas fueron identificadas en el genoma de la cepa USA300 por medio del programa BLAST (disponible en www.ncbi.nlm.nih.gov).

2.2 Resultados

2.2.1 Diseño de la sonda de DNA.

Estudios previos en el LGMB de la Universidad El Bosque permitieron identificar una región conservada dentro del gen *icaA* en tres especies de *Staphylococcus* spp. (*Staphylococcus aureus*, *Staphylococcus epidermidis* y *Staphylococcus lugdunensis*), ubicada a 240 nucleótidos corriente abajo del inicio de este gen. Interesantemente, dentro de esta región también se identificó una secuencia palindrómica de 12 nucleótidos (ATCATTAAATGAT). En los últimos años han sido liberadas las secuencias de los genomas de más especies de *Staphylococcus* spp., permitiendo realizar un análisis más amplio de la conservación de esta región del gen *icaA*. Como se muestra en la figura 2-1, en las 3 secuencias adicionales incluidas puede observarse la conservación de esta región (*S. aureus*-USA300, *S. epidermidis*, *S. capitis*, *S. haemolyticus*, *S. caprae*, *S. lugdunensis*). A partir de este análisis se sintetizó un oligonucleótido marcado con biotina (AAATTATTATCATTAAATGATGG) y otro complementario, con el cual se generó la sonda de doble hebra para determinar si existen proteínas que la reconozcan y se unan a ella en *S. aureus*. Esta sonda, con su respectivo oligonucleótido complementario, fue utilizada para el seguimiento de las proteínas en los diferentes pasos de purificación por medio de ensayos de retardamiento en gel (EMSA).

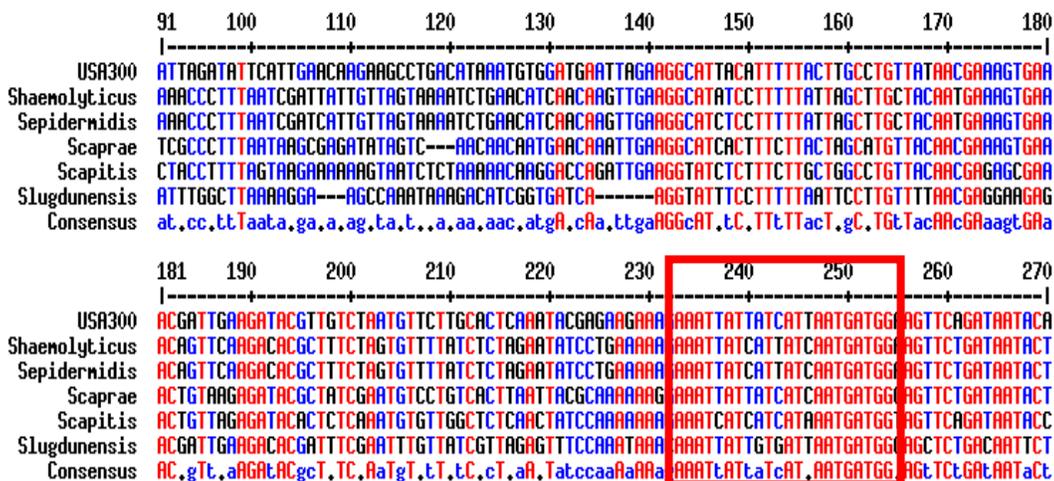


Figura 2-1: Alineamiento múltiple de la región 5' del gen *icaA* en varias especies de *Staphylococcus* spp. (*S. aureus* USA300, *S. epidermidis*, *S. capitis*, *S. haemolyticus*, *S. caprae*, *S. lugdunensis*). El recuadro rojo indica la región conservada donde se encuentra la secuencia palindrómica.

2.2.2 Detección y purificación parcial de las proteínas de unión al gen *icaA* por cromatografía líquida rápida de proteínas (FPLC).

Los ensayos de retardamiento en gel realizados con el extracto de proteínas totales de *S. aureus* muestran un retardamiento en la señal de la sonda (1500 fmoles), con la generación de dos bandas (marcadas como C1 y C2) respecto a la migración de la sonda sola sin proteínas, y su intensidad aumenta proporcionalmente al aumentar la cantidad de proteína adicionada (Figura 2-2). Pueden observarse una tercera banda cuya migración fue mayor respecto a las bandas de los 2 complejos, sin embargo, la intensidad y aparición de esta banda fue variable entre cada experimento realizado, por lo cual se decidió no tenerla en cuenta en este estudio. Los ensayos de retardamiento en gel con los extractos de bacterias formando biofilm mostraron la generación de dos complejos muy similares (Figura 2-2). Después optimizar las condiciones del EMSA y encontrar señales claras y repetibles del retardamiento de la sonda con los extractos totales en las dos condiciones, se procedió a iniciar el proceso de purificación o enriquecimiento de las proteínas responsables de este retardamiento de la sonda para

facilitar su identificación. Como se mencionó en materiales y métodos, esta purificación incluyó un primer paso de tratamiento con $(\text{NH}_4)_2\text{SO}_4$ y luego una separación por cromatografía rápida de mediana presión (FPLC).

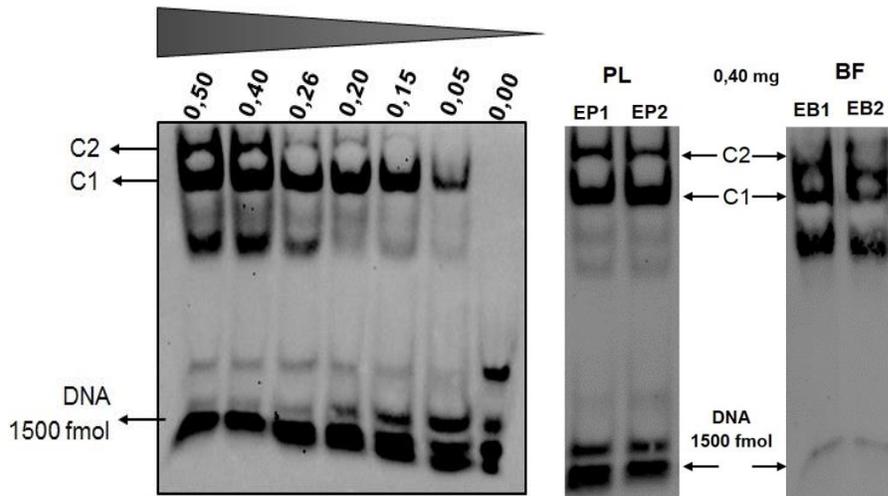


Figura 2-2: Ensayos de retardamiento en gel de la sonda de DNA que simula la secuencia palindrómica del gen *icaA*. Retardamiento a diferentes cantidades de proteína en mg. El carril con 0,00 corresponde a la sonda sola, sin proteínas. Dos replicas adicionales se muestran a la derecha para extractos de proteínas obtenidos en crecimiento planctónico (EP1 y EP2) y formando biofilm (EB1 y EB2) de *Staphylococcus aureus*. Se observa la formación de dos complejos denotados por C1 y C2.

Inicialmente, se optimizaron las condiciones de corrido para las cromatografías de exclusión por tamaño (CET) e intercambio aniónico (CIA) con extractos totales de proteína teniendo en cuenta variables como: cantidad de proteína, volúmenes de inyección, tipos de columna en la CET (superosa 6 o 12), parámetros de corrida, entre otras. Las condiciones óptimas encontradas para cada cromatografía fueron descritas en materiales y métodos, y los parámetros finales usados en el equipo están detallados en el Anexo A. Como se muestra en la figura 2-3, la CET del extracto total de proteínas generó 6 picos con diferentes tiempos de retención (picos A-G), lo cual indica una adecuada separación por tamaño de las proteínas, resultado confirmado por SDS-PAGE. Al realizar el EMSA en las 40 fracciones obtenidas, se encontró un retardamiento en las

fracciones 14, 15 y 16 (pico B, recuadros sombreados), con una mejor intensidad de la señal en la fracción 15 y menor en la fracción 16 (Figura 2-3).

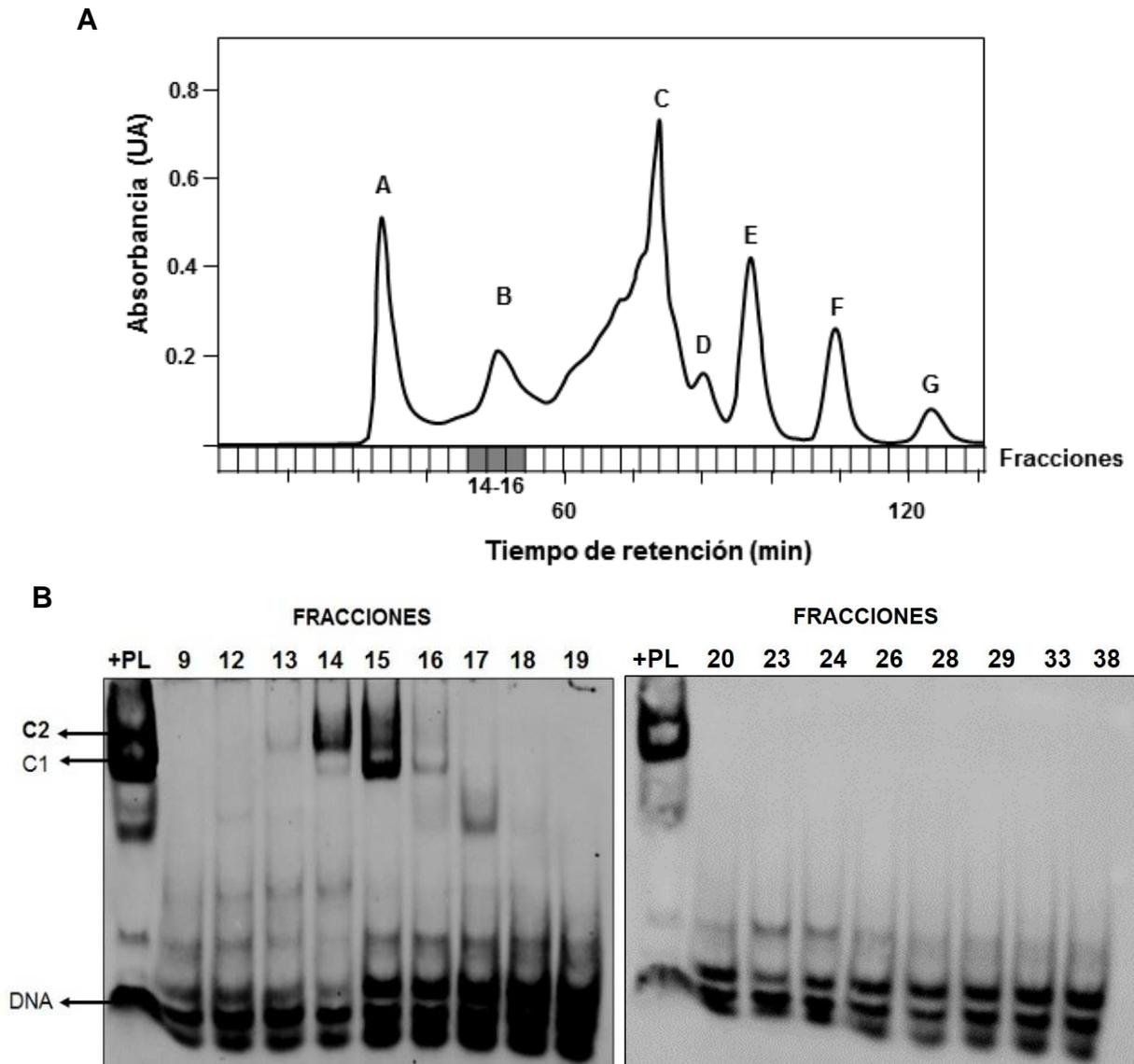


Figura 2-3: Separación por CET y detección de proteínas de unión a la sonda de DNA a partir de un extracto total de proteínas. A. Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los recuadros indican la localización de las fracciones y los recuadros sombreados las fracciones donde se detectó retardamiento de la sonda. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente.

Un extracto total de proteínas también fue separado por CIA usando las condiciones óptimas establecidas (Anexo A), y 8 picos fueron obtenidos (Figura 2-4), el pico A corresponde a las proteínas que no se unen a la fase estacionaria y probablemente tienen una carga neta positiva al pH de trabajo (7,5). Los demás picos fueron obtenidos a diferentes fuerzas iónicas proporcionadas por el anión Cl⁻. El EMSA permitió identificar los complejos sonda-proteína en las fracciones 22 a 25 (Figura 2-4), sugiriendo que estas proteínas tienen una carga neta negativa a pH 7,5.

Luego de identificar las condiciones y parámetros óptimos para la separación de las proteínas de unión a la sonda en las dos cromatografías, se dio inicio al proceso de purificación completo. Como primer paso, se realizó una precipitación de proteínas no deseadas por medio de (NH₄)₂SO₄, al 37%, las proteínas solubles (sobrenadante) fueron separadas por medio de CET, a través de 10 corridas cromatográficas independientes (500 µL de volumen de inyección por ensayo). En la figura 2-5 se muestra el perfil de separación de las proteínas, así como la identificación de proteínas de unión a la sonda en el pico B, como fue determinado con el extracto total (Figura 2-3), en las fracciones 14, 15 y 16. Adicionalmente, un comportamiento muy similar fue encontrado en el proceso de purificación a partir de proteínas formando biofilm, con señales de retardamiento en las fracciones 15 y 16 (Figura 2-5 C, D).

Teniendo identificadas las fracciones resultantes de la CET donde están las proteínas de interés, se procedió a realizar la CIA; para garantizar una cantidad adecuada de las proteínas para su detección e identificación, se reunieron las fracciones positivas de 10 cromatografías de exclusión por tamaño (20 mL), las cuales fueron inyectadas y separadas en un solo ensayo de CIA, de acuerdo con las condiciones óptimas previamente encontradas (Anexo A). Este procedimiento también se realizó para el extracto de proteínas en la condición de biofilm. Las proteínas contenidas en las fracciones fueron separadas en por lo menos 5 picos visibles (A-E), donde a diferencia de los resultados obtenidos con el extracto total, no se detectaron proteínas con carga positiva, es decir, que no se unieron a la fase estacionaria (Figura 2-6). Los EMSA revelaron señales en la fracción 52 (y ligeramente en la 51) que corresponden al pico B. Las proteínas contenidas en estas fracciones de la CIA fueron separadas por SDS-PGE, generando un número de bandas diferencial en cada una de ellas, al comparar las bandas de las fracciones 51, 52, 53 y 54 pueden observarse por lo menos 5 bandas

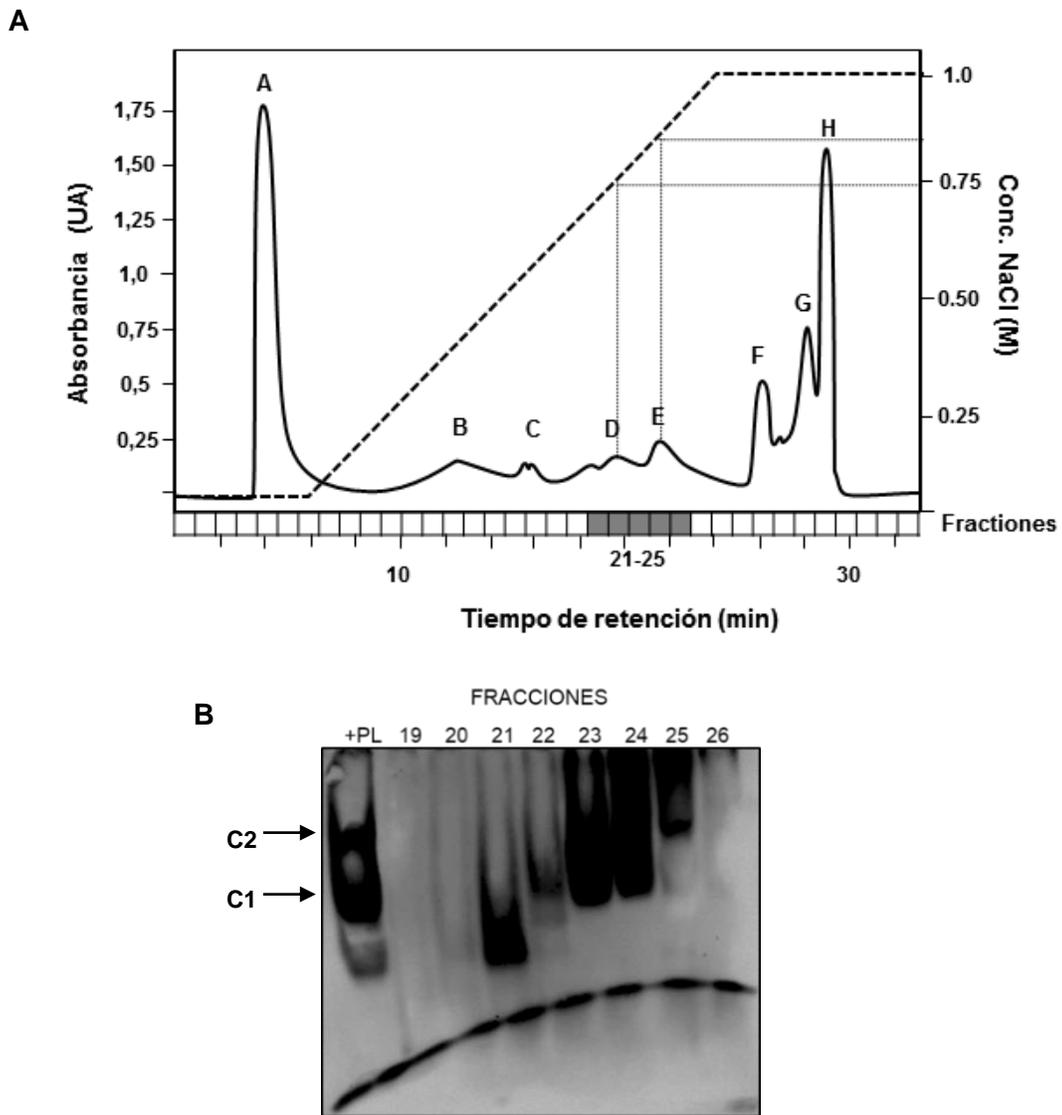


Figura 2-4. Separación por CIA y detección de proteínas de unión a la sonda de DNA a partir de un extracto total de proteínas. A. Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente.

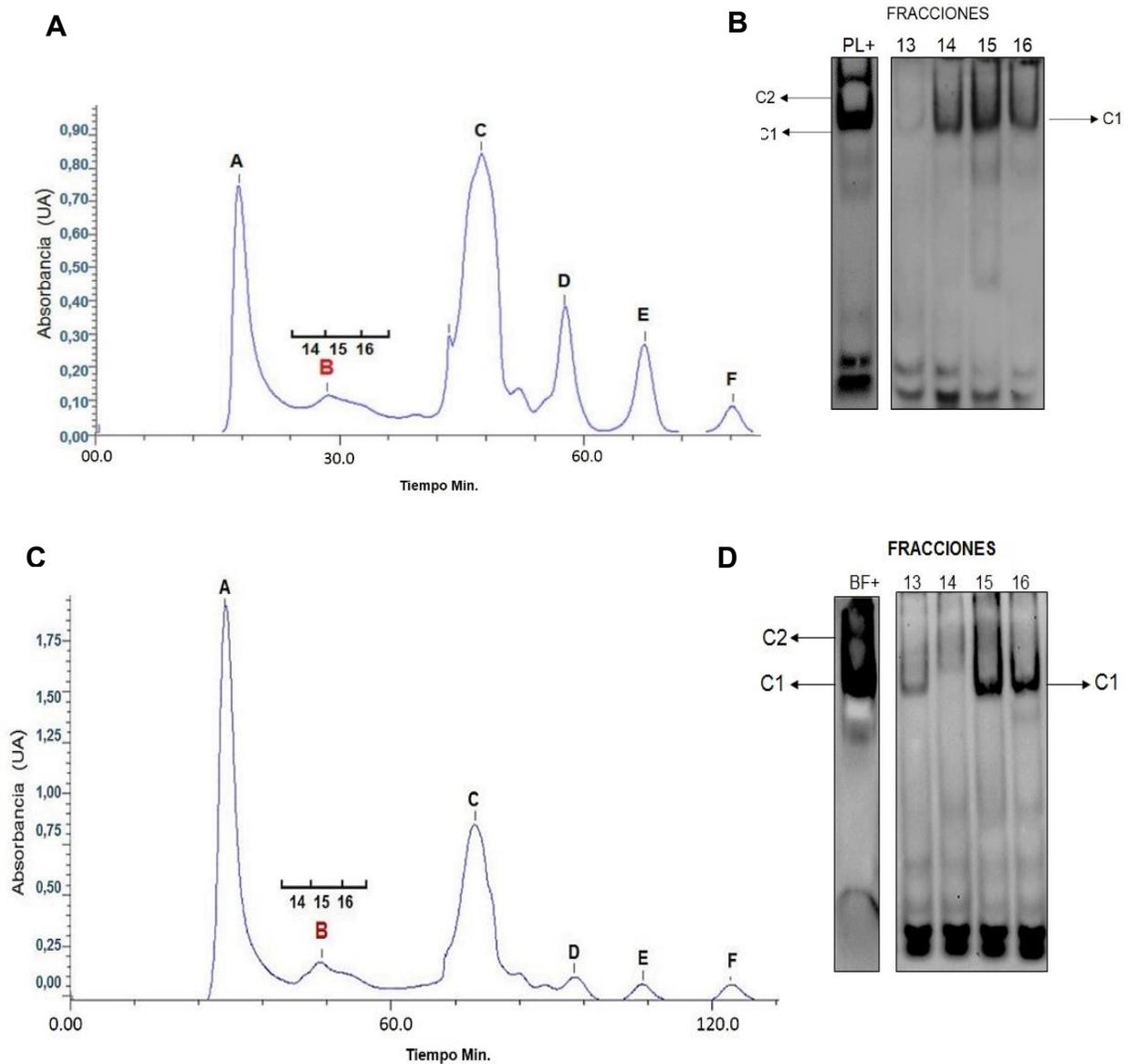


Figura 2-5: Separación por CET y detección de proteínas de unión a la sonda de DNA a partir del sobrenadante tratado con $(\text{NH}_4)_2\text{SO}_4$. **A.** Cromatograma con el perfil de separación de las proteínas en crecimiento planctónico. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. **B.** EMSA de las fracciones obtenidas de la CET. +PL y +BL: control positivo (extracto de proteínas totales). C1 y C2: complejos DNA-proteína 1 y 2, respectivamente. **C.** Cromatograma con el perfil de separación de las proteínas formando en biofilm. **D.** EMSA de las fracciones obtenidas de la CET.

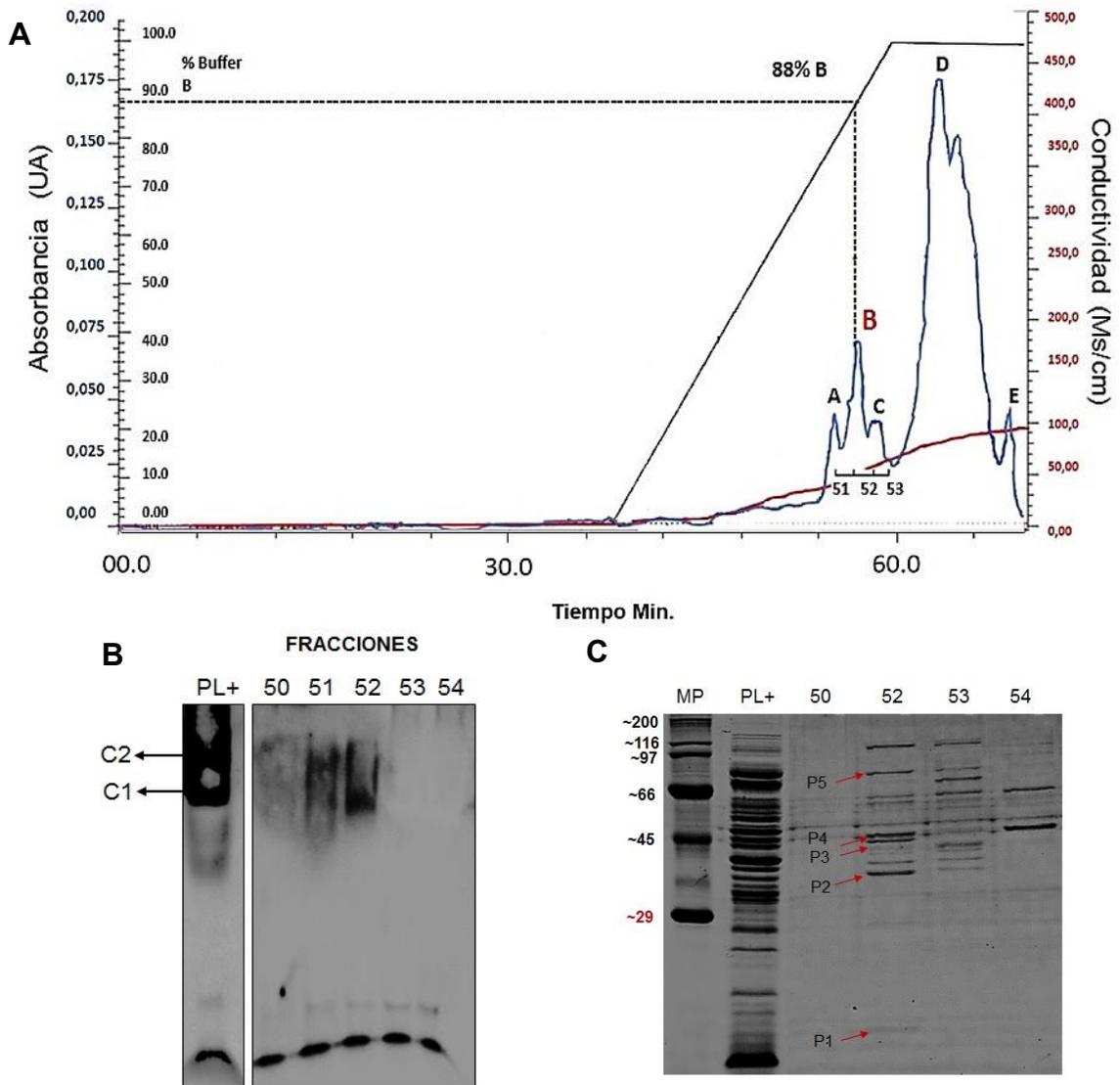


Figura 2-6: Separación por CIA y detección de proteínas de unión a la sonda de DNA a partir de las fracciones positivas de la CET (planctónico). **A.** Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. La línea negra indica el gradiente de fuerza iónica (Cl), la línea discontinua indica el porcentaje del gradiente al cual fueron eluidas las proteínas de interés. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo. C1 y C2: complejos DNA-proteína 1 y 2, respectivamente. **C.** SDS-PAGE de las fracciones positivas en el EMSA, las flechas rojas indican las bandas diferenciales que fueron secuenciadas. MP: Marcador de masa molecular.

específicas de la fracción 52 (Figura 2-6 C), que parecen no estar en las fracciones aledañas. Estas bandas fueron cortadas del gel y secuenciadas para su identificación. Adicionalmente, la fracción 52 fue enviada a secuenciar por medio del método de “secuenciación total de la muestra”, donde 15 μ L de la fracción fue ingresada en un gel de SDS-PAGE por 15 minutos y la banda con todas las proteínas sin separar fue secuenciada.

En el caso de los extractos de la bacteria formando biofilm, la CIA generó 5 picos principales (A-E), y el ensayo de retardamiento en gel identificó señales en las fracciones 48 y 49 correspondientes a los picos B y C, respectivamente (Figura 2-7). La separación por SDS-PAGE generó un patrón de bandas diferencial entre las fracciones positivas y sus fracciones más cercanas; 10 bandas específicas de las fracciones 48 y 49 fueron cortadas del gel y secuenciadas para su identificación (Figura 2-7 C). El proceso de purificación se partió de 250 mg de proteína total, en las fracciones de la CET se obtuvieron 25 mg de proteína y después de la CIA la cantidad de proteína total fue de 1,5 mg.

2.2.3 Identificación de las proteínas por espectrometría de masas MALDI-TOF

En la tabla 2-1 se muestran las proteínas que fueron identificadas en cada una de las condiciones. En los experimentos de “secuenciación total de la muestra” se identificaron más de 15 proteínas, pero el péptido mayoritario fue RAQMKGK, el cual de acuerdo al programa MASCOT corresponde a la proteína ribosomal L6 de *S. aureus*. Esta proteína tiene 178 aminoácidos, está relacionada con la síntesis de proteínas y no contiene dominios reportados de unión a DNA. También se encontraron proteínas relacionadas con el metabolismo de la bacteria como la enzima lactato deshidrogenasa, alanina deshidrogenasa, fosfogluconato deshidrogenasa, entre otras. Sin embargo, un análisis detallado de los demás péptidos minoritarios mostró la presencia de un péptido pequeño de secuencia KLLVSIR, el cual corresponde a la proteína SarX (staphylococcal accessory protein X), una proteína pequeña que ha sido relacionada con la formación de biofilm en *Staphylococcus* spp (Figura 2-8) [45, 72].

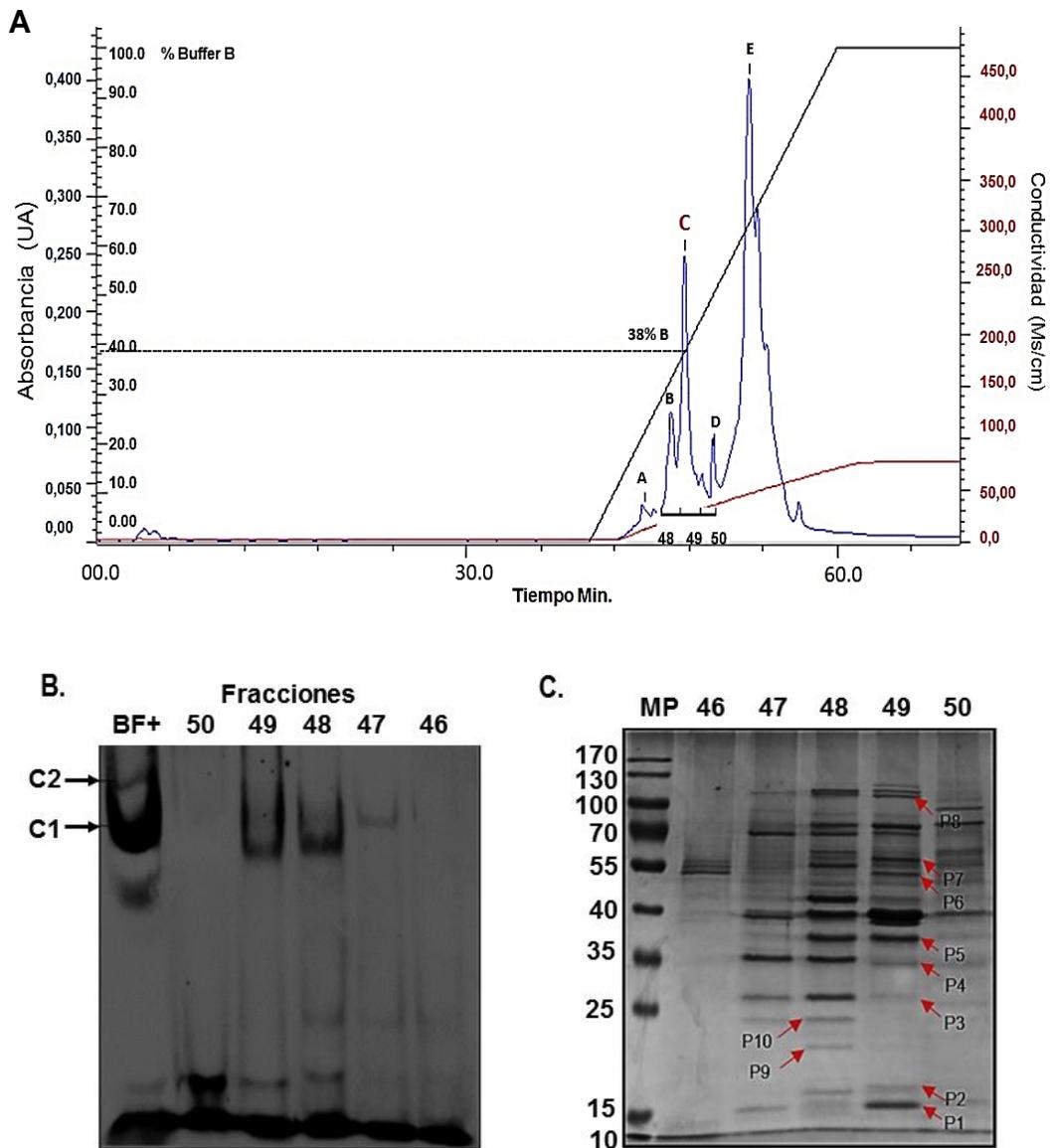


Figura 2-7: Separación por CIA y detección de proteínas de unión a la sonda de DNA a partir de las fracciones positivas de la CET (biofilm). **A.** Cromatograma con el perfil de separación de las proteínas. Las letras indican los picos más representativos obtenidos. Los números indican la localización de las fracciones de cada pico. UA: Unidades de absorbancia. La línea negra indica el gradiente de fuerza iónica (Cl^-), la línea discontinua indica el porcentaje del gradiente al cual fueron eluidas las proteínas de interés. **B.** EMSA de las fracciones obtenidas de la cromatografía. +PL: control positivo. C1 y C2: complejos DNA-proteína 1 y 2, respectivamente. **C.** SDS-PAGE de las fracciones positivas en el EMSA, las flechas indican las bandas que fueron secuenciadas. MP: Marcador de masa molecular.

Tabla 2-1: Proteínas identificadas en las fracciones finales del proceso de purificación.

(Código) Proteína	N° acceso GenBank	Masa Molecular (kDa)	N° de amino-ácidos	Función
Planctónico				
(PD ¹) Ribosomal L6-50S	KII20457	19,8	178	Síntesis de proteínas
(PD ¹) Lactato deshidrogenasa	OUL42360	34,6	330	Metabolismo
(PD ¹) Proteína SarX	ABD21615	16,7	141	Regulador transcripcional
(P1) Proteína hipotética	ABD21907	13,3	114	Desconocida
(P2) Lactato deshidrogenasa	OUL42360	34,6	330	Metabolismo
(P3) Alanina deshidrogenasa	EZU08965	40,1	324	Metabolismo
(P4) Fosfogluconato deshidrogenasa	EHS18694	51,7	428	Metabolismo
(P5) Metionil tRNA sintetasa	KII19459	71,8	657	Síntesis de proteínas
Biofilm				
(P6) Ribosomal L7/L12-50S	ORO08660	12,7	122	Síntesis de proteínas
(P7) Proteína Dps ²	AQR52476	16,7	147	Protección DNA
(P8) Ribosomal L10-50S	EHS20632	17,7	108	Síntesis de proteínas
(P9) Proteína hipotética	ABD21907	13,3	114	Desconocida
(P10) Peroxidasa de unión al grupo hemo	WP_054189156	27,9	250	Sistema Redox
(P11) Factor GrpE/Chaperona	EFW32388	23,9	187	Síntesis de proteína/Chaperona
(P12) LSU Ribosomal L25-50S/proteína de estrés general	EIA13239	23,7	221	Síntesis de proteínas
(P13) Lactato deshidrogenasa	OUL42360	34,6	330	Metabolismo
(P14) Ribosomal S1-30S	EFW34524	43,2	398	Síntesis de proteínas
(P15) GMP sintasa	EUV11019	58,2	515	Metabolismo
(P16) Valin tRNA ligasa	ORO13457	101,6	876	Síntesis de proteínas
(P17) Proteína Clp (proteasa)	KII21969	21,1	195	Proteasa
(P18) Superóxido dismutasa	KII20963	22,7	199	Sistema Redox

¹“secuenciación total”. Proteínas principales encontradas en este ensayo.

²Proteína de protección del DNA en fase estacionaria/inanición.

En la identificación individual de proteínas, en los dos estados de crecimiento (planctónico y biofilm) se identificaron proteínas en su mayoría relacionadas con algunas vías metabólicas y síntesis de proteínas de la bacteria. Además, en las fracciones positivas para la bacteria formando biofilm se identificaron proteínas que no fueron

encontradas en las fracciones en crecimiento planctónico como la proteína Dps, L25 ribosomal, Clp, superóxido dismutasa y peroxidasa, algunas de ellas relacionadas con condiciones de estrés. Sin embargo, dentro de los resultados de la secuenciación sobresale una proteína hipotética de 13,3 kDa, constituida por 114 aminoácidos y con función desconocida en *S. aureus* que fue identificada en los dos estados de crecimiento (códigos P1 y P9 en crecimiento planctónico y biofilm, respectivamente). Un análisis detallado del proceso de secuenciación por medio del software MASCOT permitió establecer que en los dos casos esta proteína fue identificada a partir de 8 péptidos diferentes, los cuales tienen una cobertura de 71% del total de la proteína (Figura 2-8) (Anexo B). Este resultado sugiere que la banda enviada a secuenciar (en los dos estados de crecimiento) se trata con alto grado de certeza de esta proteína hipotética.

A

MAE**KFKI**IMTEALSLYIWGCNINL**TEKLE**TLLGFY**KQYK**ALSEYID**KKYKLS**LNDL
 AVL**DLTMKHCKDEK**VLMQSF**LKTAMDELDLSRTKLLVSI**RR**LIEKERLSKVRSSKD**
 E**RKI**YIYLNDDI**SK**FNALFEDVEQFLNI

B

KLDFEFR

EAFANVKANEESK ETQINFQ**QK** AQEQQAIEK**DENISALMNAEQK**

MAVNLYD**YANQLEQALR** EAFAN**VK** L**DFEFR** QM**QGEETAEEDLQK**

1 MAVNLYD**YANQLEQALRESEYKAIKEAFANVKANEESK**KL**DFEFR**ETQINFQ**QK**QM**QGEETAEEDLQK**AQEQQAIE**KDENISALMNAEQKMSQV**FQ**ETINQIIVKPLDEIYAD** 114

Figura 2-8: Péptidos usados en la identificación de las proteínas SarX y P1/P9 (proteína hipotética) **A.** Secuencia de la proteína SarX (staphylococcal accessory protein X) y el único péptido usado para su identificación (Subrayado y sombreado). **B.** Secuencia de la proteína hipotética P1/P9 (línea inferior) y la localización de los 8 péptidos usados para su identificación, los cuales tienen una cobertura de 71%. Los aminoácidos arginina y lisina se muestran en rojo en las dos secuencias.

3. Capítulo 3. Características moleculares y estructurales de las proteínas SarX y YheA.

3.1 Materiales y métodos

3.1.1 Predicción de las características fisicoquímicas de las proteínas.

Las características fisicoquímicas como hidrofobicidad, punto isoelectrico, masa molecular, contenido de aminoácidos, entre otras, fueron calculadas usando el programa "ProtParam" de ExPASy y mediante comparación con información de proteínas ortólogas en la base de datos de UniProt <http://www.uniprot.org/>. Los dominios, motivos y familias fueron obtenidos con la ayuda de herramientas como BLAST-P, Pfam y prosite <http://pfam.xfam.org/>. El análisis de la existencia de dominios trans-membrana fue realizado con los programas TMpred (ExPASy), TMHMM disponible en <http://www.cbs.dtu.dk/services/TMHMM/> y CCTOP disponible en <http://cctop.enzim.ttk.mta.hu/>. Las secuencias putativas de terminación de la transcripción independientes de Rho fueron identificadas por medio del programa Transterm disponible en <http://transterm.cbcb.umd.edu/> [73].

3.1.2 Alineamientos pareados y múltiples

La comparación de las diferentes proteínas fue realizada a través de alineamientos múltiples con el programa Clustal omega 1.2.1, disponible en <http://www.ebi.ac.uk/Tools/msa/clustalo/> [74] y graficados con Multalin. Los porcentajes de identidad y similitud de las proteínas fueron determinados a través de alineamientos pareados con Clustal omega con EMMBOS Water (local) y EMMBOS Needle (global). La búsqueda de proteínas ortólogas se realizó a través de la herramienta BLAST-P y tBLASn disponible en www.ncbi.nlm.nih.gov/; en el caso de los ortólogos de YheA

encontrados tienen los siguientes números de acceso del GenBank: ABD21907.1 (*Staphylococcus aureus* USA300 FPR3757), NP_388861.1 (*Bacillus subtilis*), WP_002833772.1 (*Pediococcus pentosaceus*), AQY71882.1 (*Lactobacillus plantarum*), ARB01090.1 (*Lactobacillus fermentum*), WP_068223581.1 (*Lactobacillus backii*), ARR89743.1 (*Leuconostoc mesenteroides* subsp. *Mesenteroides*), ASX15462.1 (*Lactobacillus acidophilus*), ABN44859.1 (*Streptococcus sanguinis*), AOG58335.1 (*Streptococcus pneumoniae*), ALR79572.1 (*Streptococcus salivarius*), AMF85656.1 (*Streptococcus mutans*), APG66667.1 (*Lactobacillus delbrueckii* subsp. *Lactis*), EPH95372.1 (*Enterococcus faecalis*), WP_003729622.1 (*Listeria monocytogenes*), ASF17835.1 (*Staphylococcus saprophyticus*), AJH34205.1 (*Bacillus anthracis*), BAB04868.1 (*Bacillus halodurans*), BAC13092.1 (*Oceanobacillus iheyensis*), ACB60162.1 (*Exiguobacterium sibiricum*), AOV06873.1 (*Lysinibacillus sphaericus*) y BAD74925.1 (*Geobacillus kaustophilus*).

Para YmcA: ABD21673.1 (*Staphylococcus aureus* USA300 FPR3757), NP_389584.1 (*Bacillus subtilis*), ARS63477.1 (*Streptococcus mutans*), WP_011836121.1 (*Lactobacillus delbrueckii* subsp. *Lactis*), WP_016623786.1 (*Enterococcus faecalis*), EFR84635.1 (*Listeria monocytogenes*), WP_002483425.1 (*Staphylococcus saprophyticus*), WP_000870462.1 (*Bacillus anthracis*), WP_010898525.1 (*Bacillus halodurans*), WP_011066029.1 (*Oceanobacillus iheyensis*), ACB60523.1 (*Exiguobacterium sibiricum*), WP_010858552.1 (*Lysinibacillus sphaericus*) y WP_044731661.1 (*Geobacillus kaustophilus*).

Para YlbF: ABD20633.1 (*Staphylococcus aureus* USA300 FPR3757), NP_389382.1 (*Bacillus subtilis*), WP_002298033.1 (*Streptococcus mutans*), WP_064973459.1 (*Lactobacillus delbrueckii* subsp. *Lactis*), WP_061101306.1 (*Enterococcus faecalis*), WP_031674748.1 (*Listeria monocytogenes*), WP_069822494.1 (*Staphylococcus saprophyticus*), WP_047423599.1 (*Bacillus anthracis*), WP_010898747.1 (*Bacillus halodurans*), WP_011065849.1 (*Oceanobacillus iheyensis*), WP_012370848.1 (*Exiguobacterium sibiricum*), WP_010858307.1 (*Lysinibacillus sphaericus*) y WP_044733207.1 (*Geobacillus kaustophilus*).

3.1.3 Modelamiento de la estructura tridimensional (3D)

La predicción de las estructuras 3D fue realizada por medio de los programas I-TASSER (Iterative Threading ASSEmbly Refinement) y Quark disponibles “on line” en (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) [75]. En el primero, la secuencia de la proteína es usada para encontrar pequeños motivos estructurales en la base de datos del Protein Data Bank (PDB) (www.pdb.org), los cuales luego son utilizados para construir la primera versión de la posible estructura 3D completa de la proteína (*threading method*), refinada luego por un proceso iterativo hasta llegar a los 5 modelos más confiables. El segundo, construye la estructura 3D “*ab initio*” a partir únicamente de la secuencia de aminoácidos de la proteína; Quark genera la estructura 3D de la proteína completa a partir de modelos tridimensionales de pequeños fragmentos (1 a 20 aminoácidos) teniendo en cuenta las restricciones moleculares a nivel atómico. Los modelos 3D obtenidos de todas las proteínas fueron validados usando los programas PROCHECK (<http://servicesn.mbi.ucla.edu/PROCHECK/>), ERRAT (<http://services.mbi.ucla.edu/ERRAT/>) ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>). Las gráficas de Ramachandran fueron construidas usando los programas Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) and PROCHECK. La visualización de las estructuras 3D se realizó por medio del programa PyMOL (PyMOL Molecular Graphics System, 1.8 Schrodinger, LLC). Las comparaciones de las estructuras 3D fueron realizadas por superposición de las mismas usando la aplicación MatchMaker del programa UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>) [76], teniendo como referencia las estructuras 3D de las proteínas YmcA (2PIH) y YheA (2OEE) de *Bacillus subtilis* reportadas en la base de datos del Protein Data Bank (PDB).

3.2 Resultados

3.2.1 Análisis de la proteína SarX

La proteína SarX está conformada por 119 aminoácidos, es codificada por el gen *sarX* (426 nucleótidos), contiene un único dominio HTH (Helix-turn-Helix), el cual tiene la capacidad de unirse a DNA (Figura 3-1). Esta proteína es considerada como un gran

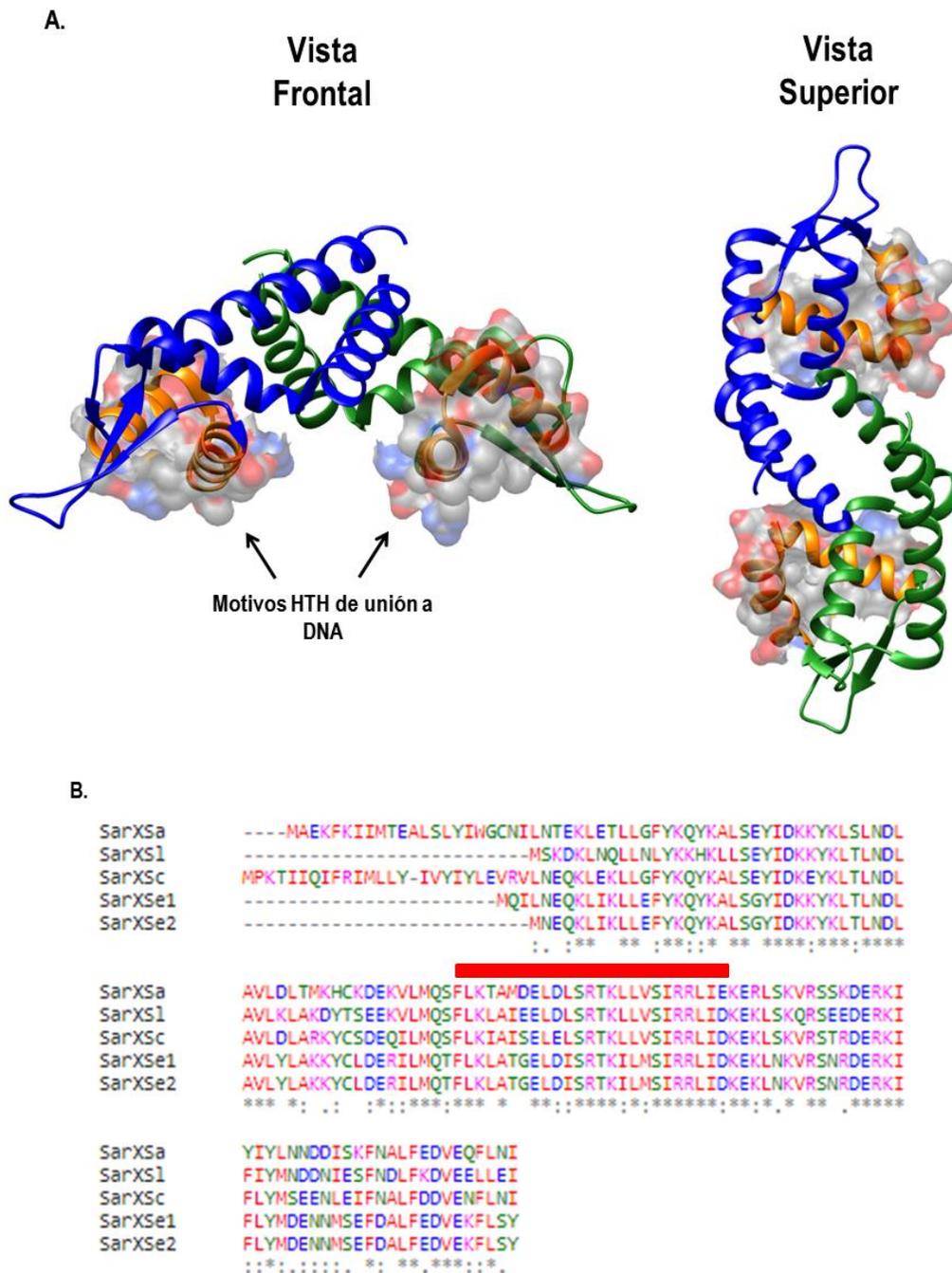


Figura 3-1: Análisis de la proteína SarX. **A.** Estructura terciaria del homodímero de SarX donde se resalta el motivo “Helix-Turn-Helix” (HTH) de unión a DNA (Naranja) y su superficie de accesibilidad al solvente. Modificación realizada de la proteína 5HS5 reportada en PDB (Protein Data Bank). **B.** Alineamiento múltiple de las secuencias de las proteínas SarX identificadas en *S. aureus* (SarXSa), en *S. epidermidis* (SarXSe1 y SarXSe2), en *S. lugdunensis* (SarXS1) y *S. capitis* (SarXSc). En rojo se muestra el dominio HTH.

regulador transcripcional, ya que regula varios genes en *Staphylococcus* spp. dentro de los cuales los más destacados son el operón *SarA* y *agr*, dos súper reguladores que controlan la expresión de cientos de genes en la bacteria. Dos proteínas ortólogas de SarX han sido descritas en *S. epidermidis* [45], y un análisis de bioinformática nos permitió identificar otros ortólogos en *Staphylococcus lugdunensis* y *Staphylococcus capitis* (Figura 3-1), con porcentajes de identidad que varían desde el 64% al 67%. SarX pertenece a la familia de proteínas SarA, conformada por 11 proteínas pequeñas, con uno o máximo dos dominios, que actúan como reguladores transcripcionales (activadores o inhibidores) de múltiples genes en *Staphylococcus* spp (Tabla 3-1).

Tabla 3-1. Familia de proteínas SarA. Comparación de las características de la proteína SarX respecto a otras proteínas pertenecientes a la familia SarA, modificado de [77].

Proteína	N° aminoácidos	N° dominios	Función
SarA	124	Un dominio	Regulón, activación de varios genes
SarR	115	Un dominio	Regulador negativo de SarA
SarS	250	Dos dominios	Activador de la síntesis de proteína A
SarT	118	Un dominio	Activador de SarS y represor de alfa hemolisina
SarU	247	Dos dominios	Regulador positivo de <i>agr</i>
Rot	133	Un dominio	Represor de síntesis de toxinas
SarX	141	Un dominio	Regulador negativo de <i>agr</i> y activado por MgrA
MgrA	147	Un dominio homólogo a MarR	Regulador de la autólisis y <i>agr</i>
SarZ	148	Un dominio homólogo a MarR	Regulador positivo de alfa hemolisina
SarV	116	Un único dominio	Regulador de la autólisis, inhibido por SarA y MgrA
SarY	247	Dos dominios	Desconocida

La proteína SarX ya ha sido asociada con la regulación positiva de la formación del biofilm en *Staphylococcus epidermidis*, ya que su delección reduce significativamente la capacidad de formar biofilm y la expresión del operón *ica*. Esta disminución del biofilm es dada a través de la unión de SarX al promotor del operón *agr*, inhibiendo su expresión, y al promotor del operón *ica*, activándolo [45]. En el caso de *S. aureus*, SarX también tiene un efecto positivo en la formación del biofilm por medio de la activación del operón *ica*, la cual se realiza a través de la unión de esta proteína a una región de 300 pb dentro del marco abierto de lectura del gen *icaA* [72]. En esta región se encuentra la secuencia palindrómica identificada en este estudio. Estos resultados demuestran que SarX podría estar participando en una cascada regulatoria que promueve la formación del biofilm en *Staphylococcus* spp., en la cual podría estar involucrada la secuencia palindrómica dentro del gen *icaA*. Debido a que esta proteína SarX ya ha sido estudiada ampliamente, se decidió continuar solo con la caracterización de la proteína hipotética P1/P9.

3.2.2 Análisis de la proteína YheA (proteína hipotética P1/P9)

Dentro de las múltiples proteínas identificadas en las fracciones finales del proceso de purificación, se encontró una proteína hipotética con función desconocida dentro del genoma de la cepa USA300 y los demás genomas secuenciados de esta bacteria. Esta proteína fue identificada en los dos estados de crecimiento, planctónico y biofilm (como proteína P1 y P9, respectivamente). A partir de esto, se emprendió un completo análisis “*in silico*” de esta proteína para determinar su posible función y ortólogos en otras bacterias. La búsqueda de proteínas ortólogas por medio de BLAST-P, teniendo como referencia la base de datos del National Center of Bio**te**chnology Information (NCBI), permitió identificar que contenía el dominio **com_YlbF**, el cual fue confirmado al utilizar los programas Prosite y Pfam (código del dominio PF06133) y aparece depositado en la base de datos de UniProt (<https://www.ebi.ac.uk/interpro/entry/IPR023378> y <https://www.ebi.ac.uk/interpro/entry/IPR010368>). De acuerdo a estos programas, este dominio ha sido relacionado con la formación de biofilm en la bacteria Gram positiva *Bacillus subtilis*, como aparece en su descripción: “*YlbF is a family of short Gram-positive and archaeal proteins that includes both YlbF and YmcA which may interact synergistically. The family is necessary for correct biofilm formation, as null mutants of ymcA and ylbF fail to form pellicles at air-liquid interfaces and grow on solid media as smooth, undifferentiated colonies. During development, YmcA, YlbF and YaaT, family*

PSPI, PF04468 interact directly with one another forming a stable ternary complex, in vitro. All three proteins are required for competence, sporulation and the formation of biofilms.” (tomado de <http://pfam.xfam.org/family/PF06133>). Este dominio fue identificado por primera vez en la proteína YlbF (de donde proviene su nombre), la cual actúa como un regulador de la formación del biofilm, esporas y competencia en *Bacillus subtilis*; sin embargo, también ha sido encontrado en las proteínas YheA y YmcA en esta bacteria [78-83].

Alineamientos múltiples pareados de la proteína hipotética (P1/P9) de *S. aureus* con cada una de las proteínas YlbF, YheA y YmcA de *B. subtilis* permitieron establecer que ésta correspondía a la proteína YheA, con una identidad de 47% y una similitud de 79% (Tabla 3-2). Aunque existe información de la posible función de las proteínas YlbF y YmcA en *B. subtilis*, hasta el día de hoy no existía información sobre la funcionalidad de la proteína YheA en *S. aureus* ni en *B. subtilis*.

Tabla 3-2: Comparación de la secuencia de la proteína hipotética (P1/P9) de *Staphylococcus aureus* con las proteínas con dominio *Com_YlbF* en *Bacillus subtilis*, YheA, YmcA y YlbF. %Id: porcentaje de identidad, %Si: porcentaje de similitud.

	P1		YheA		YmcA		YlbF	
	%Id	%Si	%Id	%Si	%Id	%Si	%Id	%Si
P1	100	100	47	79	13	40	16	49
YheA			100	100	9	35	18	44
YmcA					100	100	16	41
YlbF							100	100

Participación de las proteínas con dominio com_YlbF en la formación del biofilm en Bacillus subtilis.

Dentro del género *Bacillus spp.* las especies *Bacillus subtilis* y *Bacillus cereus* quizás han sido las más estudiadas por sus llamativos proceso genéticos y moleculares. En *B. subtilis* ha llamado la atención su “elevada” capacidad de volverse competente, formar esporas y biofilm bajo ciertas condiciones medioambientales [38-40, 84]. A diferencia de

Staphylococcus aureus, *B. subtilis* generalmente forma biofilm en la interfase aire-líquido (aunque algunas cepas pueden formar biofilm sumergido en medio líquido). Este biofilm también presenta una estructura diferente y particular, después de migrar a la interfase, las bacterias pierden su movilidad (eliminación de los flagelos) y forman largas cadenas paralelas. Luego se forma una gran “arruga” (traducido de la palabra inglesa *wrinkle*) y en la punta grupos de células conforman proyecciones aéreas propicias para la esporulación [38-40]. Como sucede en *Listeria monocytogenes*, durante la formación de biofilm, *B. subtilis* debe cambiar de un crecimiento planctónico, unicelular móvil, a un crecimiento “multicelular” no móvil. En la formación del biofilm en *B. subtilis* están involucrados múltiples genes, algunos encargados de la síntesis de moléculas y componentes básicos y otros de la regulación (positiva o negativa) y la transducción de señales externas. La matriz extracelular del biofilm de *B. subtilis* está conformada por un exopolisacárido, sintetizado por el mega operón *epsA-O* conformado por 15 genes, así como por las proteínas TasA, TapA, ubicadas en un operón junto con SipW, una proteasa relacionada con su procesamiento; y la proteína BslA, la cual recubre el biofilm y le confiere características hidrofóbicas (Figura 3-2) [39].

Cuando la bacteria está en crecimiento planctónico, los operones *epsA-O*, *tapA* y el gen *bslA* no son expresados por la acción independiente de dos reguladores negativos del biofilm, el regulador SinR y AbrB, los cuales se unen a sus regiones promotoras impidiendo su transcripción (Figura 3-2). Sin embargo, cuando las señales ambientales favorecen la formación del biofilm, hay una activación del regulador maestro Spo0A a través de su fosforilación (Spo0A~P) por la transferencia del grupo fosfato desde la proteína Spo0B, que previamente ha sido fosforilada por la proteína Spo0F, la cual a su vez ha sido fosforilada por una serie de kinasas (A-E) específicas. Este proceso es conocido como “phosphorelay process” [38, 39]. La proteína Spo0A~P (fosforilada) directamente reprime la transcripción del gen *abrB* y además promueve la expresión del gen *abba* que codifica la proteína Abba, un anti-represor que se une a AbrB y lo secuestra lejos de su sitio de acción [85]. Paralelamente, Spo0A~P activa la transcripción del gen *sinI* que codifica el anti-represor SinI, el cual se une de manera irreversible a SinR, permitiendo la transcripción de algunos genes relacionados con la formación de biofilm (por ejemplo el operón *epsA-O*) [86]. In 2008, otro anti-represor de SinR fue

identificado, la proteína SlrR, codificada por el gen *slrR*, el cual está regulado por el represor YwcC a través de una vía independiente de Spo0A [87].

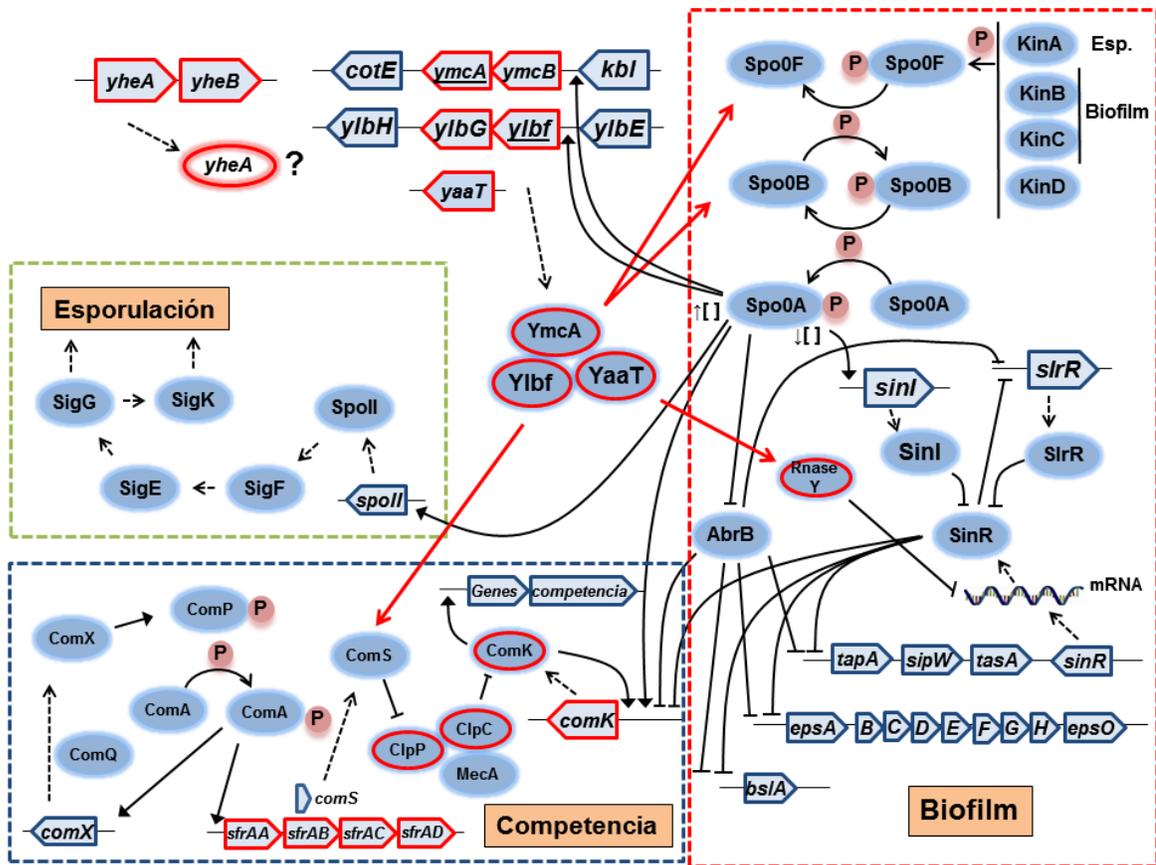


Figura 3-2: Proteínas involucradas en la formación de biofilm, competencia y esporulación de *Bacillus subtilis*. Las flechas continuas significan activación, las líneas sin flecha significan inhibición. Las flechas discontinuas representan la producción de la proteína desde el gen correspondiente. Los círculos representan las proteínas. Los genes y proteínas conservados en *S. aureus* se muestran delineados en rojo.

Respecto a las proteínas YlbF y YmcA, las dos fueron identificadas en experimentos de mutagénesis aleatoria, la primera a partir de mutantes defectuosas en su competencia (capacidad de adquirir y estabilizar DNA foráneo) [78] y la segunda en mutantes incapaces de formar biofilm [80]. Más tarde se logró demostrar que cepas con el gen *ylbF* mutado tampoco podían formar biofilm [80]. Experimentos “*in vitro*” han sugerido que YlbF y YmcA interactúan directamente y forman un complejo ternario estable de aproximadamente 80 kDa con otra proteína, YaaT. Se ha logrado establecer que mutantes individuales del gen *yaaT*, así como la triple mutante, generan cepas deficientes

en la capacidad de formar biofilm, generar esporas y ser competentes [79]. Estos tres genes son codificados en operones diferentes (Figura 3-2). Se ha demostrado que este complejo YlbF-YmcA-YaaT podría estar actuando en diferentes pasos de la formación del biofilm, de la esporulación y la competencia. Por ejemplo, la disrupción de YlbF ocasiona un descenso en la expresión de la proteína ComK (Figura 3-2), un regulador maestro que activa (directa e indirectamente) aproximadamente 100 genes, 20 de los cuales están relacionados con la competencia en *B. subtilis* [88-91]. ComK es degradado por un complejo multimérico de proteínas compuesto por las proteasas ClpP y ClpC, y la proteína adaptadora MecA; esta proteólisis es inhibida por el péptido ComS, codificado dentro del operón *sfrA*, exactamente dentro del gen *sfrAB* (Figura 3-2) [91]. Actualmente hay evidencia experimental que sugiere dos roles, posiblemente contradictorios, del complejo YlbF-YmcA-YaaT en la formación del biofilm en *B. subtilis*; el primero sugiere que este complejo podría estar actuando corriente arriba del regulador maestro Spo0A acelerando su fosforilación por medio de la interacción con las proteínas Spo0F y Spo0B. Se ha determinado que el complejo tripartita tiene la capacidad de formar dos “clusters” de $(4\text{Fe-4S})^{+2}$, gracias al alto contenido de cisteínas de las proteínas YaaT y YlbF, los cuales son muy sensibles a cambios de estado de óxido-reducción dentro de la bacteria y se ha sugerido que esta podría ser la forma de modular la interacción del complejo con Spo0B y Spo0F [79, 82, 83]. En contraste, el otro rol propuesto para el complejo YlbF-YmcA-YaaT estaría corriente abajo de Spo0A, actuando sobre la degradación del mRNA del gen *sinR* a través de una posible interacción funcional con la endoribonucleasa RNasa Y [81, 92]. La proteína RNasa Y es codificada por el gen *rny*, y está involucrada en la degradación del mRNA en diferentes bacterias Gram-positivas [93]. El complejo YlbF-YmcA-YaaT estaría formando un complejo mucho más grande con RNasa Y y estaría involucrado en la degradación de mRNA en *B. subtilis*, incluyendo a *sinR* (Figura 3-2) [81, 92]. En cepas mutantes para los genes *ylbF*, *ymcA* y *yaat*, se ha observado una vida media más larga para el mRNA de SinR, lo cual favorece una mayor cantidad del represor, inhibiendo la formación del biofilm. Existe evidencia que muestra que pequeños cambios en los niveles de SinR tienen efectos dramáticos en la formación de biofilm [94, 95]. Hasta el momento, no hay evidencia experimental que sugiera alguna función de la proteína YheA en *B. subtilis*.

Basados en estos antecedentes, se emprendió un análisis para determinar cuáles de estos genes relacionados con la formación de biofilm en *B. subtilis* están conservados en *S. aureus*. Los genes *yIbF* y *ymcA* están dentro del genoma de *S. aureus*, aunque aparecen como proteínas hipotéticas con función desconocida; de la misma manera otros genes como *yaat*, *rny* (que codifica para la RNasa Y), *comK*, *clpP*, *clpC* y el operón *sfrA* también fueron identificados (Figura 3-2). Curiosamente no se identificaron en *S. aureus* genes relacionados con el proceso de “phosphorelay”. Debido a la falta de información sobre la posible función de la proteína YheA, se realizó un análisis de las características comunes entre las tres proteínas con dominio *com_YIbF* con el fin de establecer si YheA tiene el potencial para ser una proteína de unión a DNA. Una comparación de los entornos genéticos de los genes *yheA*, *yIbf*, *ymcA* (dominio *com_YIbF*) y *yaat* (por su interacción con YIbF y YmcA) muestra un mayor nivel de conservación en el entorno de éste último gen (Figura 3-3), sugiriendo que la proteína YaaT podría estar involucrada en procesos básicos conservados en estas dos bacterias.

En los otros tres genes existe una mayor variabilidad de los entornos genéticos, sin embargo en cada uno de ellos se conserva un gen corriente arriba; en el caso de *yheA*, se conserva el gen *yheB*. Un análisis de las características fisicoquímicas y estructura secundaria de estas dos proteínas muestran que YheA tiene una masa molecular teórica de 13,3 kDa, un punto isoelectrico putativo de 4,3, está conformada únicamente por alfa hélices, regiones ricas en aminoácidos polares y sin regiones trans-membrana, sugiriendo que probablemente es una proteína globular citoplasmática (como YIbF y YmcA). Para YheB se encontraron dos dominios trans-membrana, al inicio y al final de la proteína, sugiriendo que podría ser una proteína anclada a membrana. Resultados similares fueron encontrados para las proteínas YheA y YheB de *B. subtilis*.

Teniendo en cuenta la proximidad de los dos genes *yheA* y *yheB* (separados por 78 nt) y sus características diferentes, es posible que puedan estar conformando un único operón; un análisis detallado de sus secuencias permitió identificar una caja Pribnow y la secuencia AAGGAGT (de unión a ribosoma) corriente arriba del gen *yheB*, en contraste corriente arriba del gen *yheA* solo fue posible encontrar la secuencia de unión a ribosoma (AAGGAGT) y una caja Pribnow incompleta (figura 3-3 y anexo C). Además dos secuencias de terminación independientes de Rho fueron identificadas corriente abajo del gen *yheA*. Sin embargo, en algunos experimentos realizados con la cepa NCTC8325, el gen *yheA* es transcrito independientemente del gen *yheB* de acuerdo a resultados

obtenidos por RNA-seq (resultados obtenidos por el Dr Iñigo Lasa líder del Laboratory of Microbial Biofilm de la Universidad Pública de Navarra, en Pamplona, España; y compartidos amablemente mediante comunicación personal).

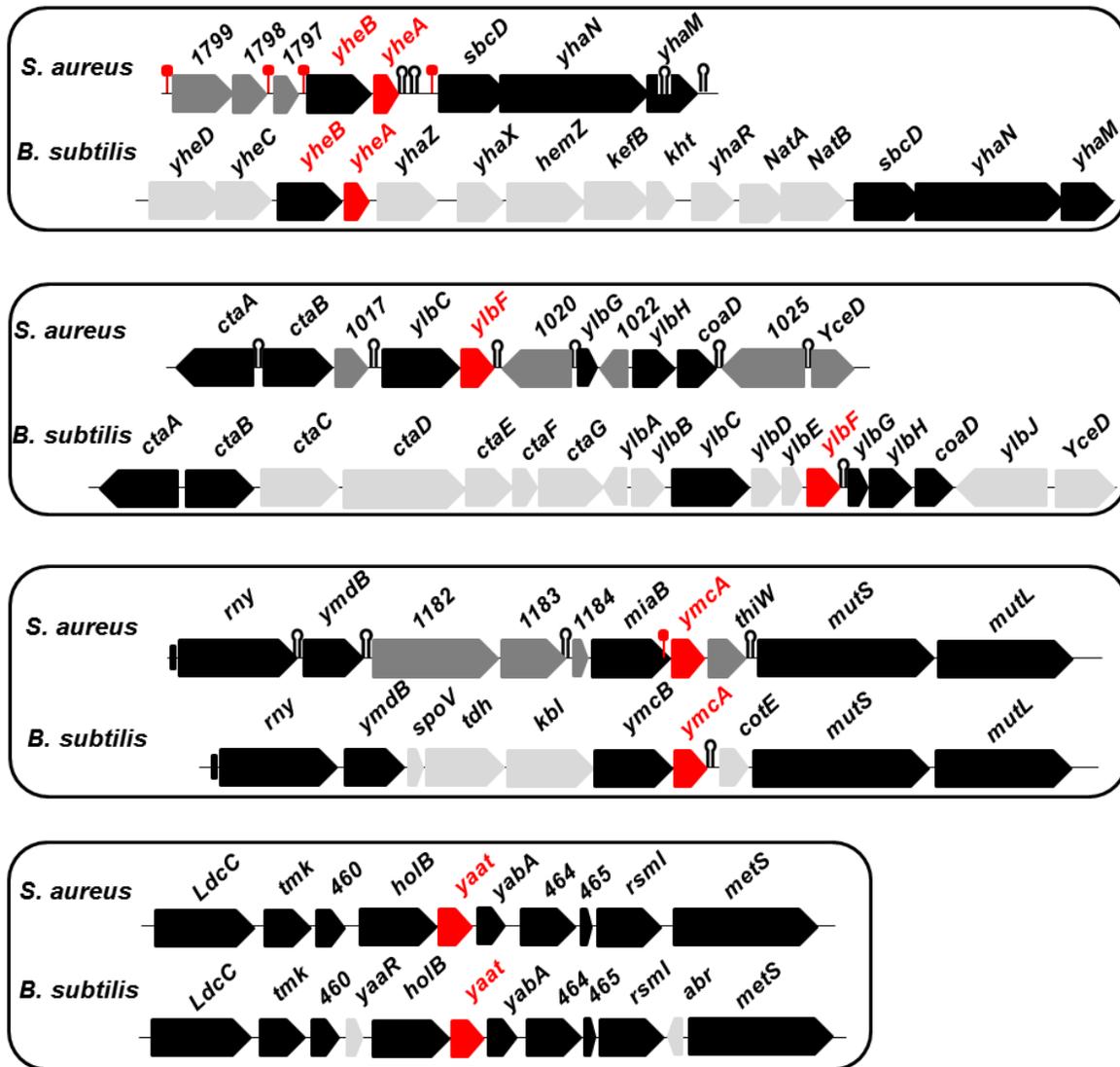


Figura 3-3: Comparación de los entornos genéticos de los genes *yheA*, *ylbF*, *ymca* y *yaat* en *Staphylococcus aureus* y *Bacillus subtilis*. En rojo los genes de interés y en negro los genes conservados; se mantienen los códigos usados para cada especie. Las cajas Pribnow y las secuencias Shine-Dalgarno se simbolizan por las paletas rojas. Las secuencias de terminación independientes de Rho identificadas por medio del programa Transterm se simbolizan por los loops en negro.

Modelamiento de la estructura 3D de las proteínas YheA, YlbF y YmcA de *Bacillus subtilis* y *Staphylococcus aureus*.

Para determinar el potencial de unión a DNA de la proteína YheA, se realizó el modelamiento de su estructura 3D por medio de dos programas, I-TASSER y Quark. Los dos programas generan modelos de las estructuras 3D “más” confiables de la proteína de interés, I-TASSER muestra máximo 5 modelos, mientras que Quark muestra 10. En el caso de YheA, el programa I-TASSER solo generó dos modelos para la estructura 3D, lo cual sugiere un mayor grado de confiabilidad de la estructura, ya que todas las predicciones realizadas confluyeron a esos dos modelos. Adicionalmente, Quark generó 10 modelos probables, dentro de los cuales estaban los 2 modelos obtenidos por I-TASSER. En la figura 3-4 se muestra la estructura 3D de YheA y la comparación de los dos modelos generados. La estructura 3D de YheA está conformada exclusivamente por 5 alfas hélices, dos de ellas (3 y 4) formando un ángulo agudo con las restantes; la diferencia más notoria entre los dos modelos está precisamente en la apertura de este ángulo (Figura 3-4). En la tabla 3-3 se muestran los valores obtenidos en la validación de los modelos generados. De acuerdo a la base de datos del PDB, hay un reporte de una estructura terciaria parcial para la proteína YheA (2OEE) y una estructura completa de la proteína YmcA (2PIH) de *B. subtilis* (Figura 3-5) determinadas experimentalmente por difracción de rayos X. Es importante resaltar que las dos estructuras están conformadas exclusivamente por alfa hélices (5) al igual que la estructura modelada para YheA de *S. aureus*. Adicionalmente las dos proteínas en *B. subtilis* forman homodímeros, por lo cual también se modeló el homodímero de los modelos 1 y 2 de YheA en *S. aureus* (Figura 3-4). El homodímero del modelo 1 está más asociado a la proteína YheA de *B. subtilis*, mientras que el modelo 2 es más conservado respecto a la proteína YmcA. Observando el alto grado de conservación de la estructura terciaria de las proteínas YheA y YmcA, surgió la inquietud de analizar si esto también sucedía con la proteína YlbF, sin embargo, a la fecha no existen reportes de la estructura terciaria de esta proteína ni en *B. subtilis* ni en *S. aureus*

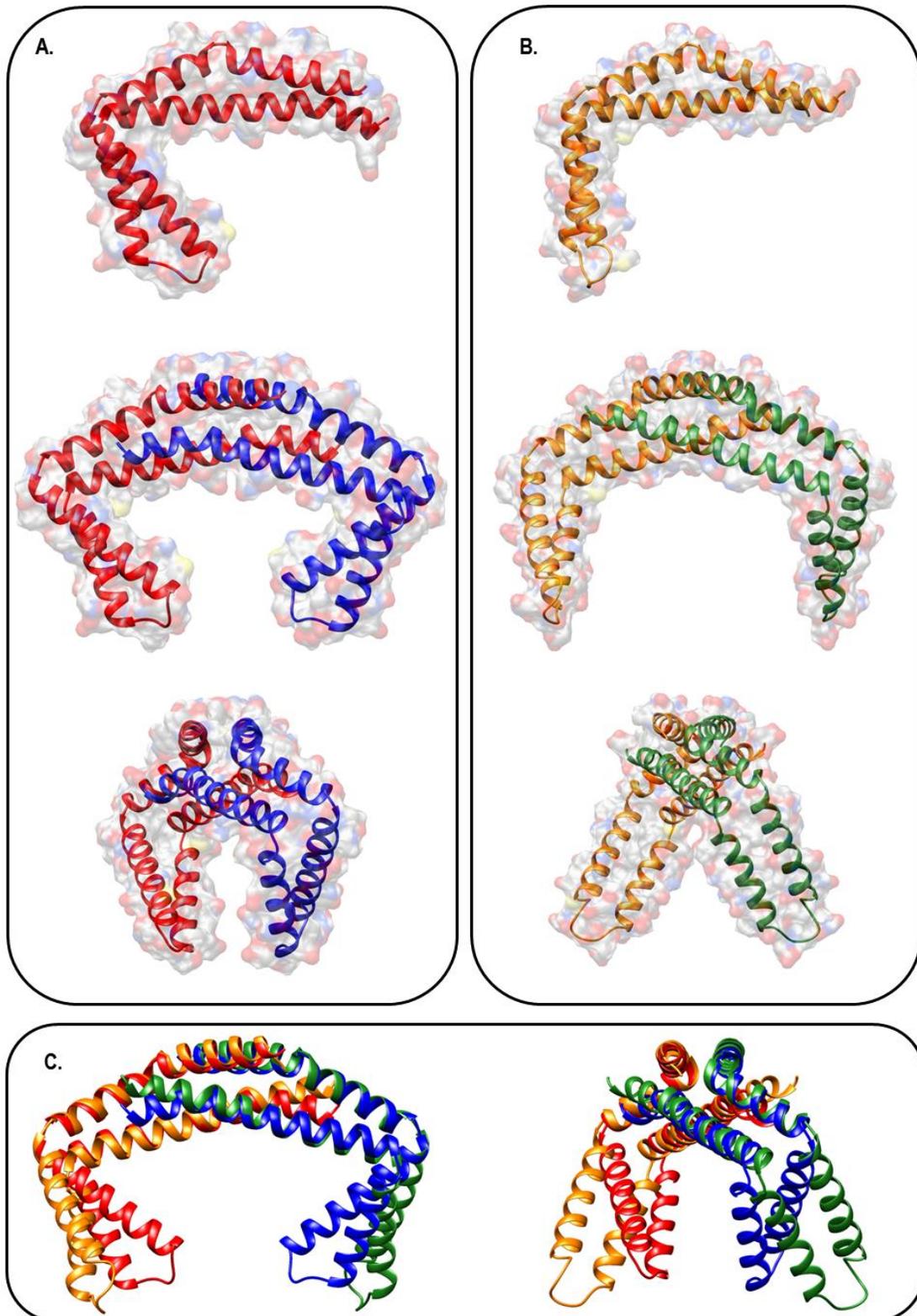


Figura 3-4: Estructura 3D de la proteína YheA de *Staphylococcus aureus*. Estructuras 3D de los dos únicos modelos generados para YheA por medio del programa I-Tasser y validados “*ab initio*” por el programa Quark con una resolución de 1.5Å. También se muestra la superficie de accesibilidad al solvente. **A.** Estructura del modelo 1 y su respectivo homodímero (vista frontal y lateral). **B.** Estructura del modelo 2 y su respectivo homodímero (vista frontal y lateral). **C.** Superposición de los homodímeros de los modelos 1 y 2 (vista frontal y lateral).

Tabla 3-3: Valores de los parámetros de calidad de los modelos 3D generados y evaluados por medio de los programas PROCHECK, ERRAT, ProSA-web y Rampage.

	<i>S. aureus</i>			<i>B. subtilis</i>		
	YheA	YmcA	YlbF**	YheA	YmcA*	YlbF**
Gráficas de Ramachandran						
Residuos en regiones favorecidas	98.2%	92.3%	93.1%	96.4%	97.5%	91.4%
Residuos en regiones adicionales permitidas	0.9%	5.1%	6.0%	3.6%	2.5%	7.8%
Residuos en otras regiones permitidas	0.9%	1.7%	0.9%	0%	0%	0%
Residuos en regiones no permitidas (“ <i>outlier</i> ”)	0%	0.9%	0%	0%	0%	0.9%
Overall G-factor	0.01	- 0.07	- 0.08	- 0.08	0.56	- 0.11
Z-score	- 4.52	- 4.98	- 3.99	- 4.21	- 4.07	- 3.87
C-score	0.46	1.16	0.56	0.40	-	0.54
TM-score	0.77	0.87	0.79	0.77	-	0.79
RMSD	3.3	2.1	3.2	3.5	-	3.3
ERRAT	100	96.5	100	100	100	100

* Valores reportados para YmcA en PDB, código 2PIH.

**Valores calculados sobre las estructuras generadas desde las secuencias parciales, las cuales solamente comprenden el dominio *com_ylbF*.

Entonces, se procedió a modelar la estructura terciaria de YlbF en *B. subtilis* (Figura 3-5) y *S. aureus* (Figura 3-6). Los 5 modelos generados por I-TASSER para YlbF estuvieron conformados por 5 alfa hélices, distribuidas espacialmente en forma de “L” como las proteínas YheA y YmcA. Sin embargo, una porción del extremo C-terminal de la proteína tuvo una conformación aleatoria (incluso en los 5 modelos) y no se encontraron alfa-hélices u hojas beta. Justamente en esta región, la proteína YlbF posee un motivo rico en cisteínas (Figura 3-8), el cual se cree contribuye a la formación de los “clusters” (4Fe-4S)²⁺ junto a YaaT [83]. Las proteínas YheA y YmcA no contiene este motivo. Para facilitar la comparación de las estructuras terciarias de las 3 proteínas, se modelo nuevamente la estructura de YlbF, pero sin el motivo C-terminal. La delección de este motivo no generó cambios visibles en la estructura terciaria de la proteína. Una superposición de las estructuras de los homodímeros de las 3 proteínas tanto en *B. subtilis* (Figura 3-5) como en *S. aureus* (Figura 3-6) muestran su alto grado de conservación, aunque los porcentajes de identidad sean relativamente bajos entre ellas (Tabla 3-2). Interesantemente, estas proteínas asumen una estructura en forma de “tenaza”, donde la diferencia más notable entre YheA y sus homólogos YlbF y YmcA es la apertura de su “quijada”.

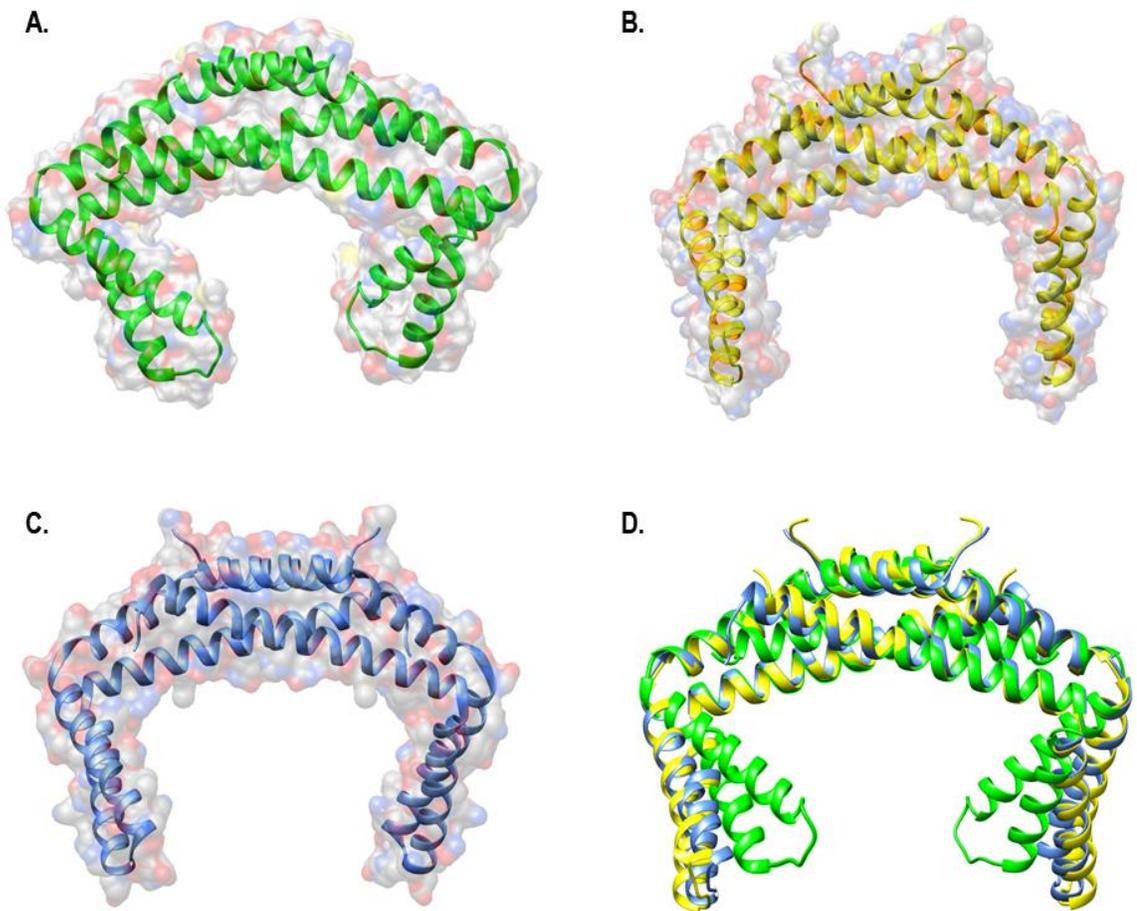


Figura 3-5: Estructura terciaria de las proteínas YheA, YmcA y YlbF de *Bacillus subtilis*. Homodímeros de las 3 proteínas con su superficie de accesibilidad al solvente. **A.** Estructura de la proteína YheA modelada a partir de la estructura parcial reportada en PDB (2OEE). **B.** Estructura de la proteína YmcA reportada en PDB (2PIH). **C.** Estructura parcial de la proteína YlbF modelada por el programa I-tasser y Quark. **D.** Superposición de la estructura terciaria de las 3 proteínas usando como referencia la proteína YmcA.

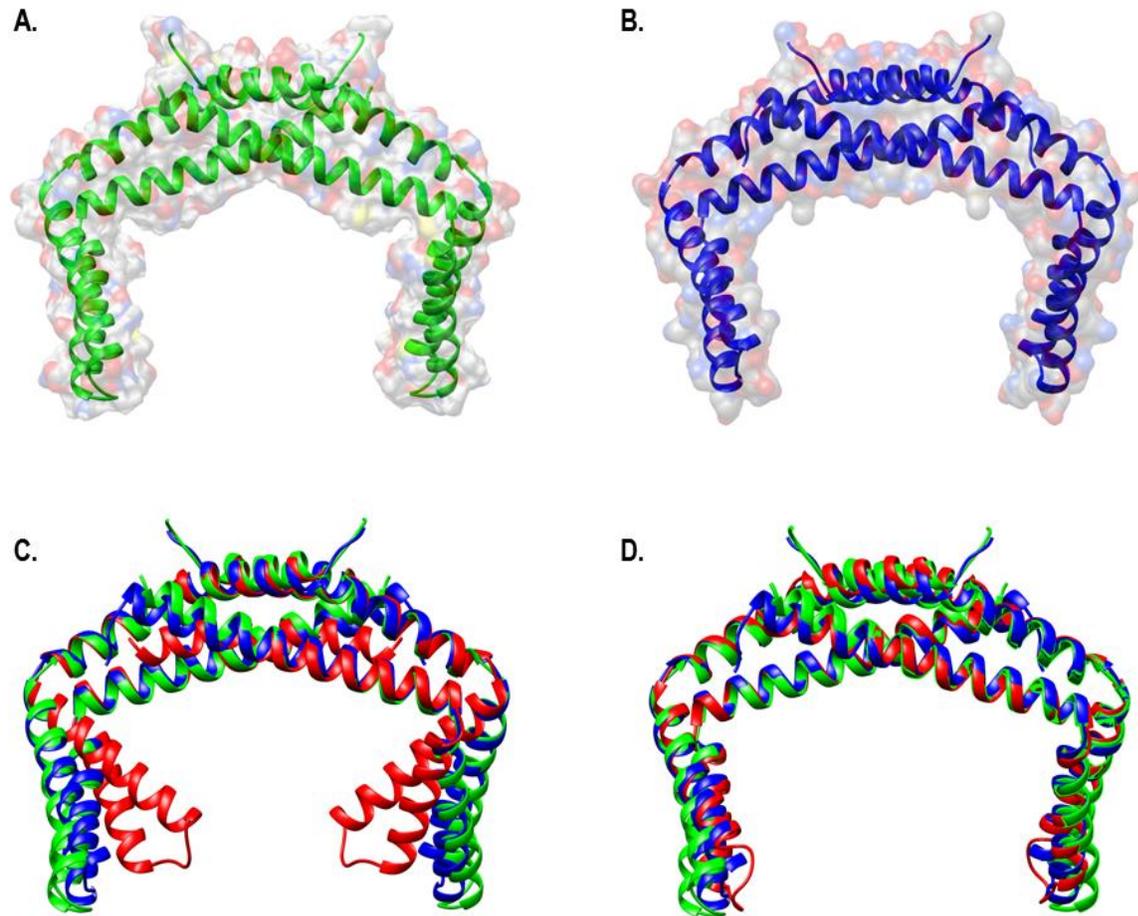


Figura 3-6: Estructura terciaria de las proteínas YmcA y YlbF y comparación con YheA en *Staphylococcus aureus*. **A.** Estructura de la proteína YlbF. **B.** Estructura de la proteína YmcA. **C.** Superposición de la estructura terciaria de las proteínas YheA (modelo 1), YlbF y YmcA. **D.** Superposición de la estructura terciaria de las proteínas YheA (modelo 2), YlbF y YmcA.

Análisis filogenético de YheA, YmcA y YlbF en bacterias Gram positivas.

En 2017, se realizó un análisis filogenético de las proteínas YlbF, YmcA y YaaT que incluyó 28 bacterias Gram positivas del *phylum* Firmicutes (Lactobacillales, bacillales y Clostridiales) algunas de ellas con capacidad de formar esporas (como *B. subtilis*) [83]. Este análisis permitió identificar que estas proteínas no están limitadas únicamente a bacterias que forman esporas, sino que también se encontraron de forma diferencial en bacillales y Clostridiales [83]. Para generar más evidencia sobre la posible función de la proteína YheA, se realizó este mismo análisis pero para la proteína YheA, YmcA y YlbF de *S. aureus* (Figura 3-7). Se encontraron homólogos de la proteína YheA en Lactobacillales (como *Lactobacillus acidophilus*) y en bacilos no esporulados (como *Listeria monocytogenes*), pero no en Clostridia. En contraste, ortólogos de las proteínas YlbF y YmcA no fueron encontrados en Lactobacillales ni Clostridia. Las identidades de las diferentes proteínas homólogas a YheA estuvieron en un rango entre 26% (*Streptococcus pneumoniae*) a 83% (*Staphylococcus saprophyticus*). Estos resultados pueden sugerir que YheA podría estar implicada en procesos básicos que han sido mantenidos en algunas “clases” del *phylum* Firmicutes, independientemente de la capacidad para formar esporas.

Después de determinar la existencia de ortólogos de YheA, YmcA y YlbF en otras bacterias con un rango amplio de porcentajes de identidad, se procedió a estudiar la localización y distribución de estos aminoácidos idénticos a través de un alineamiento múltiple (Figura 3-8). Interesantemente, en las proteínas YheA se identificó la secuencia consecutiva de aminoácidos **QKQMXG**, la cual está conservada en casi todas las secuencias y está localizada en su sección intermedia (Figura 3-8A). Analizando el alineamiento múltiple de las proteínas YmcA también se encontró una secuencia consecutiva con los aminoácidos **QKQAVN**, la cual está conservada en todas las bacterias y también en su sección intermedia, como fue encontrado en YheA (Figura 3-8B). Una característica especial de estas dos secuencias conservadas es la alta frecuencia del aminoácido glutamina (Q), 3 y 2 respectivamente; a partir de este resultado se determinó el porcentaje de glutamina en la secuencia total las proteínas YheA y YmcA, siendo inusualmente alto (13,2 % y 13,3%, respectivamente). Respecto a las proteínas YlbF, el alineamiento múltiple permitió identificar la secuencia de aminoácidos conservadas **FGXYHPDY**, la cual se encuentra localizada también en la

sección intermedia de las proteínas; sin embargo, no contiene glutamina; resultado que coincide con el bajo porcentaje de este aminoácido en la proteína YlbF (4,2%).

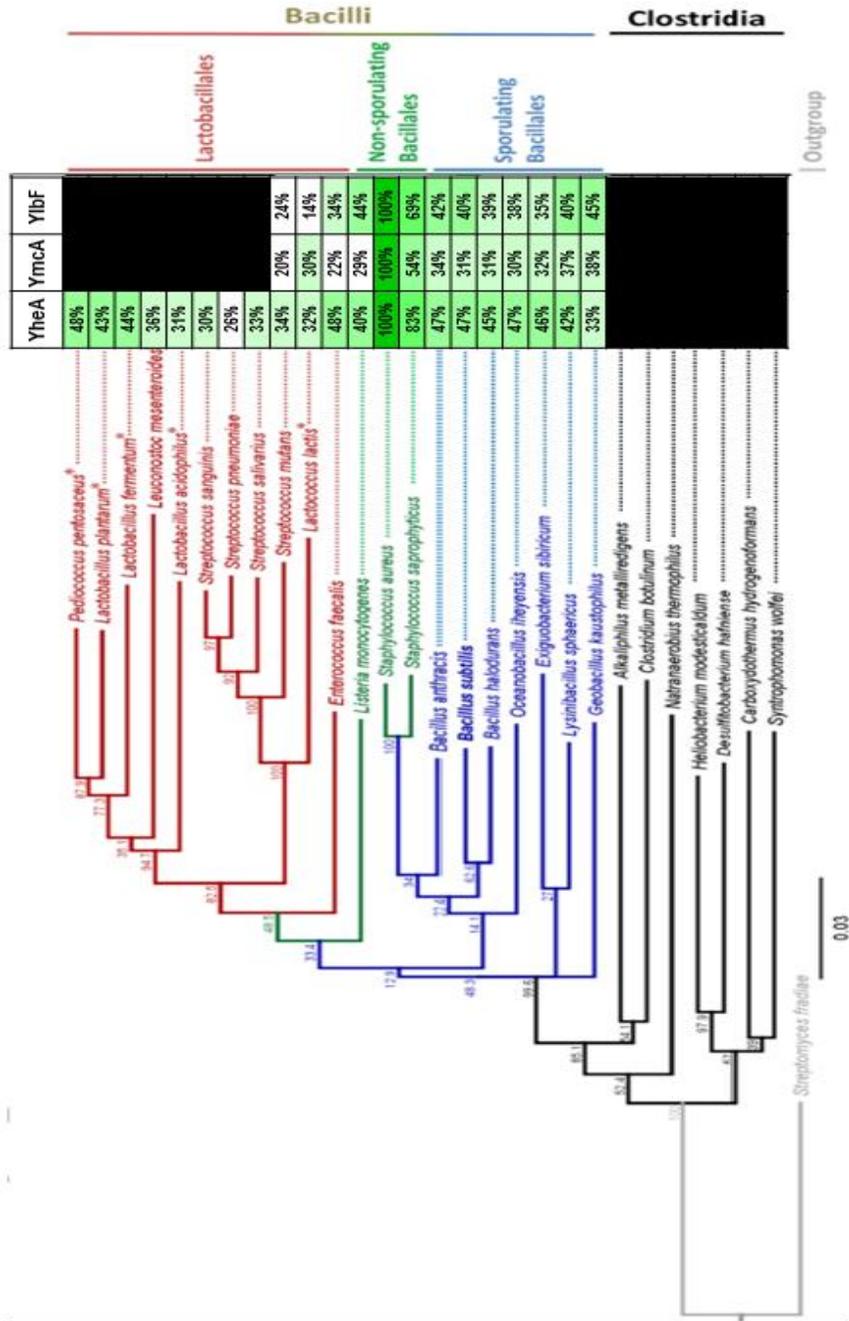


Figura 3-7: Ortólogos de las proteínas YheA, YmcA y YlbF en diferentes bacterias Gram positivas. Porcentajes de identidad de las proteínas YheA, YmcA y YlbF en bacterias Gram positivas con y sin capacidad de formar esporas respecto a las proteínas identificadas en *Staphylococcus aureus*. Modificado de [83].

A

```

>YheASa MAVNLYDYANQLEQALRESDEYKAIKEAFANVKANEESKLLFDEFRETQINFPQQKMQGEEIAEE
>YheASs MAVNLYDHDANQLEQALRESDEYQAIQNAAYAKVKNQESKDLFDEFRETQLSFPQQKMQGEEIGE
>YheAApp MAVNIYDTANQMEQEIRQTSSEYKSLQEAFAEVRKDEHASALYSKFDVQSNLQQKQMNQGEELTEE
>YheALp MAVNIYDTANQLEQELRQTKFELKVAIDYDMKNDYSAFSLFKDFQEVQMLSQKQMNQGEELTDD
>YheALb MAVKVDYDANQLEGLRKTDEFLALESAYTELKKNQKAYDLFKNFQQINMKLQQKMQGQENLEDS
>YheALf MVENIYDTANERLQMRQTQEFIGLKEAFDLDKADKEATDLFVKFQAKQAAAQQKMQGQEIISED
>YheAEf MANIYDTANQLERELRESEQFDLKGSDRLKANE EAYALFKEFQELQQLSQKQMMQGEEMTEE
>YheALm MAVNIYDLAHDLDKGIRETPEFISLQDAYREVNENADAKAKFERFRDVTI QEKQMTQGEITDE
>YheABs MAVNFYDVAYDLENALRGSEEFTRLNKLYDEVNADESARMPFNFRDVLRLQQKQMGAGEITQE
>YheABh MSNVYDKAQELKQALAESEFKALKSLHTQIDQDDIARLLENFRQLQLDLQQKMQGQVQITEE
>YheAOs MANTLDVANLEQALRESDEFKSLKEAYEAVMGEPTAKQMFENFRDQLSLQEKQMQGIEITEE
>YheABa MTKNIHDVAYELQKAIENDDFRTLKESYAAVQADAASKNLFDEFRTMQLSLQQKMMQGEITEE
>YheAEs MSETNLYDHAYTLERALETSEYTTLDKLYAQVNADPMSNELFASFRNVQLELQQKMQGQGEITPE
>YheALs MTVNIYDDLNRLEATFRKTAFAEVKEAVDVVAAD EALAMPKNFREVQMKMQEKQMQGEEISGD
>YheAGk MSEPLHALAKQLEQAIRASEPFQQLKRAYEDVRRDETAYRMPANVRDIQLQLHEKQMRGAAILPD
>YheALm MTVNIYDNANEMANILITETQQYISWQNAFNAINQD TDSKALFGFQEI QMAVQQMMSQQQPKPE
>YheALa MNIYDSANKLAEDLTQTDQYKALADAVKVVQEDAESAALFKKMDLQTKIMQSQQTGPLSE
>YheALd MVNIYDTANQLANDLRETQFLALKEVMDAVKADGSLALFKELDAAQMEIMEAQQTGKELTEE
>YheASa MSNIYDLANELSRNLRELPEYKVAESKKAVDADSEAKAIFTDYLAFAQELQQLAQTGQVPTQE
>YheASp MSNIYDSANELSRGLRGLPEYKTVKAAKDAIAADAESKIFTEYLAFAQEEIQKLAQTGQMPDAS
>YheASs MTNIYDLANELERGI RALPEYKTLVEKKEAIAADAESALFKFEFTDFQEDFYAKMQAGTMTPEE
>YheASm MTVNVYDLANELERGI RALPEYQALVQAKAKIDEDEARKLWTEFTEFQMKIQGLIQTGQAPTPE

```

```

>YheASa EDLQKAQEQAQAI EKDENISALMNAEQKMSQVFQEQINQIIVKPLDEIYAD
>YheASs EELQKAQEQAQKIENDSNISELMAAEQNMSQVFQEQINQIIVKPLDEIYAD
>YheAApp EEIKNVHEIADQIEKVDLIKDLMDKERNNMQLFSQVSIIVKPIQELYKN
>YheALp DEVQKAHDLADKVGNVDEIKSLMGKERNLNQLMNDLNQIITKPVQELYQN
>YheALb SEIDEAKKLADKVGKFKIEIQTLMEEKGLSQTMEDLNGIITKPVQELYAN
>YheALf DEIKEIQALAKDVTSKDVIQALMAKEQQVDQMIQQQLNQIITGPIQLYKQFGPQEG
>YheAEf EEAKQKQELSQRKIQQDGI VSDLMQKEQAFSLVANDLNRIMGVPVRDLYEA
>YheALm ETVDVAQKVAQEVQENELIVKLMKEEQAMSTIINDLNRIMTPLQDLYNVAND
>YheABs EEVTQAQKTVALVQQHEKISQLMEAEQRMSMLIGELNKIIMKPLEELYGSVEG
>YheABh EEAKQKQQLFELVQQHELISKLMEEQRLSVVIGDLNRIITPLEEYIGNPNE
>YheAOs EEVEKARQVVETVQQHEGISKLMEEQRLLNLINEISQIITKPLEDLYGTAEQN
>YheABa EDNQQAQEVVVRIQQDAKITKLMETEQRLNVVIGDVNKIIMKPLEELYSQQQV
>YheAEs EDMASAQAKMLEVQNNPLIGLQMQEBQRMHTMLTEVTQIIMKPLEELYAPMMEQNVQDDVQ
>YheALs DELEHAQKVAQLAQGNEKIMNMLQAEMLKSLGLEEVNRVLMKPVQEFYKEL
>YheAGk DEIEQAQKAMALAQQNEKLARLMALAEQQMSMTIAEVQQIIMKPLEELHRSFMGR----
>YheALm EQEKEWDAVAKVQKDEKINALMEAEQALNTLLTELNDIVTKPVAEAYSKLQK
>YheALa EDQQAQKDLNDQVQNNKQIVSLQTEQSLYELLNDIQRTYSKPINDLIEDLR
>YheALd EQQDHFKSLNERVSNQNTLQSMMLAEQAVYTLNDVQKNIGQPLSKAYEDLRKA
>YheASa EVQDKMTEFGEKIQGNVLSSEFFNKQQQLSIYLADIERIIFPVPVQDLLK
>YheASp SFQAKMEGFGKQIQGNLSLSEFFTKQQQLAIYLSDIEKIVFEPVSELLK
>YheASs EEQAAVQELGQKVEANALLKEYLAAQQGLSVYLNDIERIIFKPLQELNS
>YheASm EVQTEMQSFNQKIEANPLLKEYATAQQALGVYVNDIERIIFSPLODIAK

```

B

```

>YmcASa MYNKDDVLNQADNIANKIKNLDITKTYQQIEAQIHQNQTIKTKMDMLKHKQKQAVNFQNYGKQNALEQSEHTIQSIEAE
>YmcASs MYEKDDILAEERLSERIKSLDITVKEVHKVESQIHNNENLEQRMKELKKNKQKQSVNLQNYGKIHALQASEEKIQDIEED
MIKNGDEIVTVSKDEIMKKA TELRDALQQTEVFSYRVAEERINANSKVAAKVSKIKLLQKEAVNLEHYQKLEAMKQNTENQIDNVRAD
>YmcALm MTEVMYTKEDLIIKKSHEIAHMIANTPEVEFFKKAQAQINENQLVRERIASLSLQKQAVNFQHLGKREALKMIEDKIAKIEEE
>YmcAEs MYMYTDEQIIEKANEIAKMISETSEVERFKAAEANINGNEKVQKMMKQIKFLQKEAVNQCQHYGKTEALARVEAKLDKLF AE
>YmcAOi MTEYTRKEVLDEAKLANMLASTEEIDRFKQVEAKLNENKQVSLIQKIKTLQKQAVNFQAYEKREALKRVEEIIDRLQAE
>YmcABh MTAKYTRDDIKQRARELALMIAETEEVDFFKRAEKQINEHLRVQEIIEAIKKYQKEAVNLQHYGKKEALKEVEAKIDELQNE
>YmcABs MTLYSKDKIVQQRNLAKMIS ETEEVDFFKRAEAQINENDKQVSTIVNQIKALQKQAVNLKHYEKHEALKRVEAKIDALQEE
>YmcABa MKVYSKDEIVEQAKELAKMIS ETEEVDFFKRAEAQIHNENKRAIDEIKALQKQAVNLQHYGKWEALKRVEAKIDALQDK
>YmcAGk MAKYTRDDILVQAKQLAKMIAETEEVDFFKRAEKIHQNEKVRTLINELKSLQKQAVNLQHYGKHEALKRVEAKIDAIYEE
>YmcAEf MALQKEPEIQHAVDELLSLKENETVCFKLEEKVQVQNSLDQLVEDIKQAQKDAVQFAHYGKPAEQEAIKRADRLTKE
>YmcASm MSPFEEALNQLIELLQEHDSVQAFQAVEKKIKSLPALNKVAHDMKGYQDQAVLFRQIEKSKAQKEADQKQKMGES
>YmcALL MHYDDLVDLLVKKIKNLDITVKEFQQAEEALMENQELFKSQDEMKAQKEAVLYQKIKIQAYKRTSQAQQMIEKR

```

```

>YmcASa INTLPIVEQFQTSQYEANQLLKMVFVSTMETRLNDHNKAKHSD
>YmcASs INILPIVEEFRESQAEANDLLQMMIHTMSDRLNKHKQE
>YmcALm IDSLPIVTEFRRAQEDANDLLQISITSEITTKVTELEKES
>YmcALs INAIPIVQQFKDSQVDVNDLLQLVSNNTIANNVTNEIIRSTGGDLLRGETGSYVANSQPGSCS
>YmcAEs LDEMPIVQQFQESQEEVNDLLQYVTVTISNGITDQIEATDGDVLAGTGSQMEENKSGGCGY
>YmcAOi IDDIPIVQEFKQIVVNDVLQVLSKTIAREVTNEVESTGGDILAGETGSKAKNKSGCSH
>YmcABh IDAIPIVQEFKQSQMDVNELLQLVSMTISKTVTDHITSMGGDLLRGTTKNPLIDDGCGKEKC
>YmcABs LEEIPIVQEFKQVMEVNDLLQLVAHTISNQVTNEIITSTGGDLLKGETGSKVKHSNNSCSL
>YmcABA LDSIPIVQEFKSSQTYVNDLLQLVASTISNNVTDEILISTNGDVLKGETGAAVESSKGNCGC
>YmcAGk LSQIPIVSEFQSQSDVNDLLQLVASTISNTVTDEILASTGGDVLRGETGAALRYNKHSGCH
>YmcAEf FEEHPIVIAYREQLAEANDLLHHLTNLIQQQLNEKIEEEDYASKN
>YmcASm LDKLPIVQDYRAKMQDASDLLQYVTKTLEEKINEELGND
>YmcALL IKHHPIVEDYATKLEDVNDLVQYITGELEEKVNLLLTGTE

```

C

```

>YlbfLl  MLIIDQNLEIDNLEKIMDEFLLKFPVEAYQKAKADFMADENLQSQLKTLQ-----DNSEYIAFRPELRALQ----
>YlbfSs  MLIINEELLAIDQAI DCLVSYFLKLEVENYRLRQKRFENDRELQQLFAFQELKDSYDDAKSYEAFRPDVRELKRQVL
>YlbfSm  MLKINSQLFAIEDTLDDLAATFLNLEVVRTYKAAQKAFADKHLQKEISNFQHYNEDYNDQKTFIKYRPEVKQLRRKMF
>YlbfEf  MIVTEHLFAIEDQTECLISALLESDAVQNYKETKKAMYASTEVAQLQKAFLESKSAFERVEAYGIHAPDFREKQRALR
>YlbfSa  MINEASLAILDDIDE LADMIVASDIYASFEQAKQALENNDEAHLLYQSFLKSKKEYDEVMRFGRYHPDYKVMLETR
>YlbfSs  MITEETLTVLDEIEMLSDKILQSRLYKAYKTAEQTLANDDEAHLLYQAFLLKSKKEYDEIMRFGRYHPDYQSVMLETR
>YlbfEs  MIYSDKTIDLINAAEDLAYSLSQSETGQAYRLAKQARANSRQAQDVIRD FNRIKEEYELVQRFGRYHPDYQITKKVH
>YlbfOi  MIATMEVNIIDSEKLGKMIQSDIMQDYQTRSTMETDDEAQKLIREFSAMKEHYDDVQRFGRYHPDYDIMKKVR
>YlbfBh  MLTSDITSLIEQADQLGAMIKDSEPFHHYIEAKRALQQDKEAQQLIAEFNKMKDMHEEVQRFGRYHPDYDRVSRIR
>YlbfLm  MLATMENMALLDLSDLEASMILNSEEAQNYKRAKEALLNDANSQKNIRQFIRIKEQYEEVQRFGRYHPDYKEVTRKTR
>YlbfLs  MMTSEWAILDEADALCEMILSSEPANTLEEAYKAVYSDAHLVGAIHAFGRMKEQYEDVQRFGRYHPDYHTIMKSIR
>YlbfGk  MKIATLERIEILDKAEALAKMIVESDVADEYRRFAFWRLKQDRRAQQLIARFVKLKERVEEVQRFGRYHPDYRNVMKVVR
>YlbfBa  MIVATLESVLIIDKAEQLAKAIIICSDIAEDYRKYKELQEDMEVQTLIQQFTAMKERVEEVQRFGRYHPDYTFVSTKMR
>YlbfBs  MYATMESVRLQSEAQQLAEMILQSETAENYRNCYKRLQDEEAGRIIRSFIKIKEQYEDVQRFGRYHPDYREISRRMR

>YlbfLl  ---HEINLNEKVYDFRLAENDLQQIILTALTKKITNSISEQIYVDENLPLKGGQHGRHHGKH
>YlbfSs  RMKRQIDLNEVIGYRQTEFDLQSI LANLGEETIAQAVSDQIFIDTGLPLAPHKPHHKKGDTNIKERMSH---D
>YlbfSm  EKKRQLDLNEKVIALRRAEVDLQEVLAQKIAEVSVDVFDVDTGLPLAPHKPHHKKGYGNNIKER
>YlbfEf  KAKRALDLNEIVANYRFAETNVQTLTLDITGLKIAQLISEDIKVDAGNPFPERGKKH--SGC--GGSCHAS
>YlbfSa  QRKRAYEMLDVVMHYKAREMALQHLIDEVVTKIAAYVSEHVKIETGNPFFQTS---HSGCATGGSCNCSL
>YlbfSs  KRKRTYEMLDVVMYKQKEVALQELIDQVIVKIAAYVSENVKIEAGNPFQKE---SGGCATGGSCGCSL
>YlbfEs  EIKRALDLQEEVAQFKKAEKQLETLGQVSLVLAGQVSPQIKVPTGNPFFDQ--GCA-GGCSTGGSCSCSG
>YlbfOi  STKREMDMNDKVAAFKIAERNLQKLLDEISSFVANSVSEQVKAPREGAALSD-SGC---GCGSGGACGCAS
>YlbfBh  MLKRTLHLPSVVAFKAEAELESLNELSVMIARAVPSIKVPTGNPFFDS-IGCG-GGCGSGGSCSCSA
>YlbfLm  AYKREVMDQNVAAFRRAEMDLQSLDEISLLLASAVSENIKVPTGNPFFETKSACSTGGCGSGGGCGCSA
>YlbfLs  QQKRELDLNEKVAALKIAENDFQDLDEISLLIGKTVSEAVKVPVSNPFFASNSSCG-SGCGSGGSCSCSA
>YlbfGk  EAKRELDLHETVAAFKAEKAMQELDEISVLIQKAVSEHVKVPTGNPFFLS-AGCS-GGCRAGGSCGCRT
>YlbfBa  ELKRSVDLHDKIAAFKRVETALQKLLDEVSVAIGSEVSSIKVPTGNPFFDAGGGC---GCGTGGGCGCKKTG
>YlbfBs  EIKRELDLNDKVADFKRAENELQSI LDEVSVIEGTAVSEHVKVPTGNPYFDGLSSCG-GGCGSGGSCGCKVS

```

Figura 3-8: Alineamiento múltiple de las proteínas YheA, YlbF y YmcA en bacterias Gram positivas. Alineamiento de las tres proteínas identificadas en las diferentes bacterias Gram positivas con y sin capacidad para formar esporas, realizado con Clustal Omega y editado con BioEdit para resaltar los aminoácidos conservados. **A.** Proteína YheA, **B.** Proteína YmcA, **C.** Proteína YlbF, con su motivo rico en cisteínas en su extremo C-terminal después de la posición 130. En rojo se muestran los aminoácidos conservados en >90% de las secuencias. Los números de acceso para cada una de las secuencias están descritos en materiales y métodos.

Actualmente, se han propuesto dos nombres diferentes para las proteínas YlbF, YmcA y YaaT en *B. subtilis*, ya que estos provienen de una nomenclatura genérica usada en los procesos de secuenciación del genoma de esta bacteria, el primero sugiere nombrarlas como proteínas RcsA, RcsB y RcsC, respectivamente (RNase Y-containing complex subunitsA, B and C) [81]; el segundo sugiere nombrar a YmcA, YlbF y YaaT como RicA, RicF y RicT, respectivamente (Regulatory iron-sulfur containing complex subunits A, F and T) [83]. Basados en nuestros resultados proponemos renombrar a la proteína YheA de *Staphylococcus aureus* como proteína Grp (Glutamine-rich protein).

Posteriormente, se procedió a determinar la ubicación de las secuencias conservadas previamente identificadas en las estructuras terciarias de las proteínas YheA, YmcA y YlbF (Figura 3-9). Interesantemente, las secuencias conservadas se localizan en la tercera hélice dentro de las tres proteínas, específicamente en el loop que conecta las hélices (excepto para YmcA) y además orientadas hacia el interior, es decir en las “quijadas” de la tenaza. Estos resultados sugieren la presencia de 3 posibles motivos putativos específicos para las proteínas YheA, YmcA y YlbF.

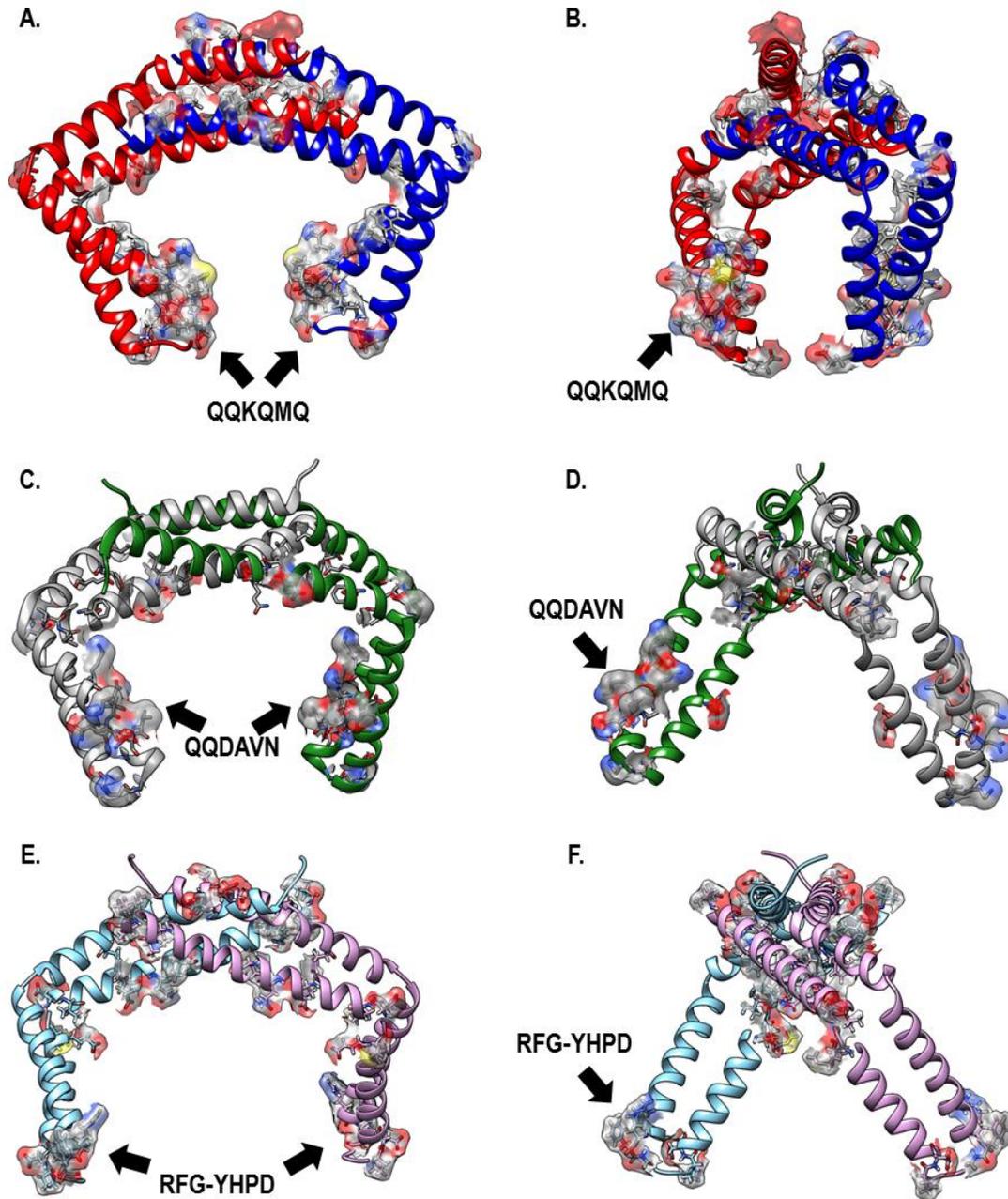


Figura 3-9: Localización de los aminoácidos conservados en las proteínas YheA, YmcA y YlbF en bacterias Gram positivas. A y B. Proteína YheA, vista frontal y vista lateral, respectivamente. **C y D.** Proteína YmcA, vista frontal y vista lateral, respectivamente. **E y F.** Proteína YlbF, vista frontal y vista lateral, respectivamente. Las flechas indican los aminoácidos conservados consecutivos en los alineamientos múltiples de las tres proteínas.

4. Capítulo 4. Evaluación de la participación del gen *qrp* en la formación del biofilm y hemólisis en *Staphylococcus aureus*.

4.1 Materiales y métodos

4.1.1 Cepas bacterianas

Las cepas bacterianas usadas en este capítulo son descritas en la tabla 4-1. Brevemente, los ensayos de mutagénesis fueron realizados en las cepas NCTC8325, 15981 (cepa de origen clínico con alta capacidad de formar de biofilm) y 132. Para los ensayos de unión de la proteína recombinante Qrp (rQrp) con la sonda se utilizaron las cepas USA300-0114, a partir del cual se amplificó el gen *qrp*, *Escherichia coli* TOP10, para almacenamiento y propagación de los plásmidos recombinantes y *Escherichia coli* BL21 para la expresión y purificación de la proteína recombinante.

4.1.2 Generación de las cepas mutantes, complementadas y sobre-expresadas

Obtención de cepas mutantes. La delección del gen *qrp* del genoma se realizó en tres cepas de *Staphylococcus aureus* genéticamente diferentes: NCTC8325, 15981 y 132 usando el plásmido pMAD [96]. Los pasos más importantes de la metodología son mostrados en la figura 4-1. Se diseñaron primers específicos para la amplificación de dos fragmentos corriente arriba y corriente abajo del gen *qrp* (Tabla 4-2). Para el fragmento corriente arriba (fAB) que tiene un producto esperado de 579 pb se emplearon los primers

Tabla 4-1. Cepas bacterianas y plásmidos usados en este capítulo

Plásmidos y cepas bacterianas	Características principales
Plásmidos	
pMAD	Plásmido termosensible para recombinación homóloga en <i>Staphylococcus aureus</i> [96].
pCN51	Plásmido para la sobreexpresión de genes en <i>Staphylococcus aureus</i> [97].
pJET	Vector de clonación/Amp ^r /
pJET-AD	Plásmido pJET con el fragmento AD corriente arriba y abajo del gen <i>qrp</i>
pJET- <i>qrp</i>	Plásmido pJET con el gen <i>qrp</i> completo.
pMAD-AD	Plásmido pMAD con el fragmento AD corriente arriba y abajo del gen <i>qrp</i>
pCN51- <i>qrp</i>	Plásmido pCN51 con el gen <i>qrp</i>
pGEM-T	Vector de clonación/Amp ^r /
pET302-C	Vector de expresión/Amp ^r /adición del His6x tag en el C-terminal
pET-K	Plásmido pET-Kinase para control de las transformaciones
pET- <i>arcR</i>	Plásmido pET-C recombinante con el gen <i>arcR</i> . Control.
pGEM-T1	Plásmido pGEM-T recombinante con el gen <i>qrp</i>
pET-CT1	Plásmido pET302-C recombinante con el gen <i>qrp</i>
Cepas bacterianas	
8325-4	Cepa de referencia NCTC8325 de <i>Staphylococcus aureus</i> .
15981	Cepa de <i>Staphylococcus aureus</i> fuertemente formadora de biofilm (amablemente regalada por el Dr Iñigo Lasa)
132 <i>icaC</i> -3xFLAG	Cepa de <i>Staphylococcus aureus</i> 132 modificada genéticamente, el gen <i>icaC</i> tiene 3 secuencias FLAG (amablemente regalada por el Dr Iñigo Lasa)
RN4220	Cepa de <i>Staphylococcus aureus</i> derivada de la 8325-4 pero con una mutación en el sistema de modificación-restricción, apta para experimentos de manipulación genética.
<i>Escherichia coli</i> TOP10	<i>Escherichia coli</i> TOP10
<i>Escherichia coli</i> BL21	<i>Escherichia coli</i> BL21 con el gen que codifica para la RNA polimerasa T7.
8325-4 <i>qrp</i> ⁻	<i>S. aureus</i> 8325-4 mutante por delección del gen <i>qrp</i>
8325-4 <i>qrp</i> ⁻ pCN51	<i>S. aureus</i> 8325-4 <i>qrp</i> ⁻ con el plásmido pCN51

8325-4 <i>qrp</i> ⁻ pCN5- <i>qrp</i>	<i>S. aureus</i> 8325-4 <i>qrp</i> ⁻ con el plásmido pCN51- <i>qrp</i> (complementada)
8325-4 pCN5- <i>qrp</i>	<i>S. aureus</i> 8325-4 con el plásmido pCN51- <i>qrp</i> (sobre expresada)
15981 <i>qrp</i> ⁻	<i>S. aureus</i> 15981 mutante por deleción del gen <i>qrp</i>
15981 <i>qrp</i> ⁻ pCN51	<i>S. aureus</i> 15981 <i>qrp</i> ⁻ con el plásmido pCN51
15981 <i>qrp</i> ⁻ pCN51- <i>qrp</i>	<i>S. aureus</i> 15981 <i>qrp</i> ⁻ con el plásmido pCN51- <i>qrp</i> (complementada)
15981 pCN51- <i>qrp</i>	<i>S. aureus</i> 15981 con el plásmido pCN51- <i>qrp</i> (sobreexpresada)
132 <i>qrp</i> ⁻	<i>S. aureus</i> 132 mutante por deleción del gen <i>qrp</i>
132 <i>qrp</i> ⁻ pCN51	<i>S. aureus</i> 132 <i>qrp</i> ⁻ con el plásmido pCN51
132 <i>qrp</i> ⁻ pCN51- <i>qrp</i>	<i>S. aureus</i> 132 <i>qrp</i> ⁻ con el plásmido pCN51- <i>qrp</i> (complementada)
132 pCN51- <i>qrp</i>	<i>S. aureus</i> 132 con el plásmido pCN51- <i>qrp</i> (sobreexpresada)
<i>E. coli</i> XL1 blue pMAD-AD	<i>E. coli</i> XL1 blue con el vector pMAD-AD.
<i>E. coli</i> XL1 blue pCN51- <i>qrp</i>	<i>E. coli</i> XL1 blue con el vector pCN51- <i>qrp</i> .
RN4221 pMAD-AD	<i>S. aureus</i> RN4220 con el vector pMAD-AD, usado para hacer deleción del gen <i>qrp</i> .
RN4220 pCN51- <i>qrp</i>	<i>S. aureus</i> RN4220 con el plásmido pCN51 con el gen <i>qrp</i> (sobreexpresada)
<i>Eco</i> TOP10- <i>qrp</i>	<i>E. coli</i> TOP10 con el plásmido pGEM-T1
<i>Eco</i> TOP10pET- <i>qrp</i>	<i>E. coli</i> TOP10 con el plásmido pET-CT1
<i>Eco</i> BL21- <i>qrp</i>	<i>E. coli</i> BL21 con el plásmido pET-CT1

*qrp*_A y *qrp*_B; y para el fragmento corriente abajo (fCD) que tiene un producto esperado de 655 pb se usaron los primers *qrp*_C y *qrp*_D. Estos fragmentos fueron obtenidos a partir de DNA genómico obtenido de la cepa NCTC8325-4. Una colonia de la cepa fue crecida en 3 mL medio LB durante 12 horas, la suspensión bacteriana fue centrifugada y el DNA extraído a partir del precipitado de bacterias utilizando el método fenol-cloroformo como ha sido estandarizado previamente en nuestro laboratorio. El DNA fue cuantificado por su absorbancia a 260 nm usado un nanodrop y entre 50 ng y 100 ng fueron usados para cada PCR.

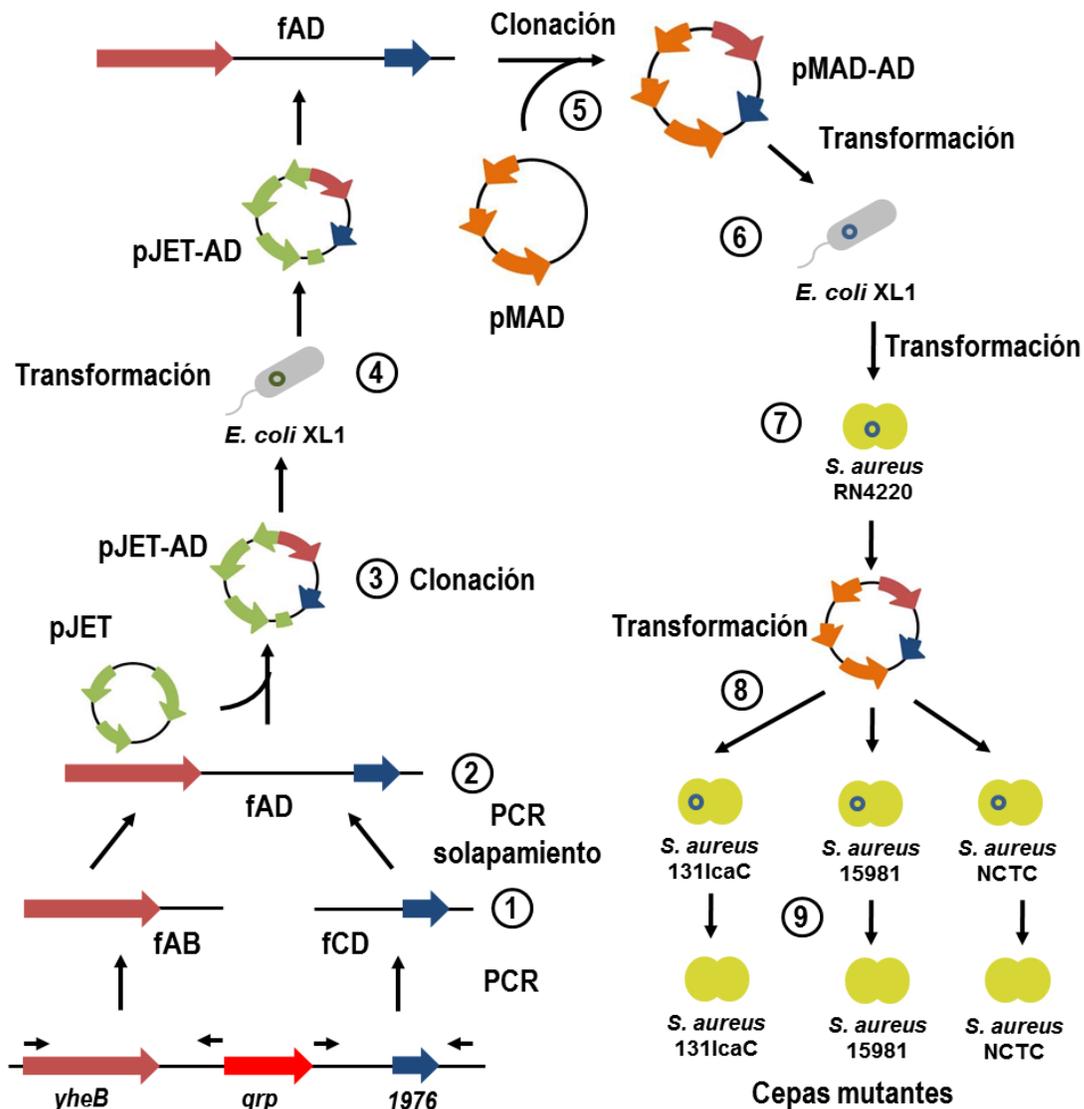


Figura 4-1: Proceso para la generación de las cepas mutantes con la delección del gen *qrp*. 1. Amplificación por PCR de los fragmentos *fAB* y *fCD* localizados corriente arriba y abajo del gen *qrp*, respectivamente, 2. Generación de un fragmento de DNA que contiene los dos fragmentos anteriores por medio de PCR de solapamiento (“*overlap extension PCR*” [98]), (*fAD*), 3. Clonación del fragmento de DNA (*fAD*) en el plásmido *pJET*, 4. Transformación de *pJET-AD* en *Escherichia coli* XL1 blue, 5. Liberación y clonación del fragmento de DNA (*fAD*) en el plásmido *pMAD*, 6. Transformación de *pMAD-AD* en *Escherichia coli* XL1 blue, 7. Purificación y transformación del plásmido *pMAD-AD* en la cepa RN4220 de *Staphylococcus aureus*, 8. Purificación y transformación del plásmido *pMAD-AD* en las cepas NCTC8325-4, 15981 y 132, 9. Delección del gen *qrp* en las tres cepas de interés y confirmación por PCR.

Tabla 4-2. Oligonucleótidos usados para la producción de las cepas mutantes, cepas complementadas y proteínas recombinantes.

Código LGMB	Secuencia	Uso
qrp_A	<u>GGATCC</u> GAACTTATACGTTTAACATCTCATCC	Amplificación fragmento DNA corriente arriba del gen <i>qrp</i> . BamHI
qrp_B	TCTATTTAAAGGATTAATAATATTATTTTAA TAGGTTATAGCACTCCTTTGTTG	
qrp_C	TATTAATAATATTTTTCAATCCTTTAAATAGA	Amplificación fragmento DNA corriente arriba del gen <i>qrp</i> . EcoRI
qrp_D	<u>GAATTC</u> CCAGTCGCCATTTTTATTTTCG	
qrp_E	CAAATGTTTATGACAAAAGAGAGTATTG	Confirmación de la delección del gen <i>qrp</i>
qrp_F	AGGTGTAGTGATTCATTAGCTGATTTA	
pJET-S	CGACTCACTATAGGGAGAGCGGC	Primer corriente arriba del sitio de clonación del plásmido pJET Primer corriente abajo del sitio de clonación del plásmido pJET
pJET-AS	AAGAACATCGATTTTCCATGGCAG	
qrp_S	<u>GAATTC</u> TTAGTCAGCGTAAATTTCTGCTAAT	Amplificación completa del gen <i>qrp</i> en <i>S. aureus</i> para clonar en pCN51.
qrp_AS	<u>GGATCC</u> TGGCAGTAAATTTCTATGA TTATGCAAAT	
GP545	CGGTTCTCTAG <u>AATGGCAGTAAATTTATATGA</u>	Amplificación completa del gen <i>qrp</i> en <i>S. aureus</i> para producir recombinantes
GP538	TCGTGCATGCATGTCAGCGTAAATTTTCG TCTAATG	
GP539	GAGTTCCGTGAAACTCAAAT	Amplificación de un fragmento del gen <i>qrp</i> en <i>S. aureus</i>
GP540	TGATTTCTTGAATACTTGACTC	
GP515	TGGGGGTGCAACATTTTGAATA	Amplificación de un fragmento del gen <i>sarX</i> en <i>S. aureus</i>
GP516	TCATCTTTGCAATGCTTCATCGT	
GP544	CGGTTCTCTAGAATGGCTGAAAAATTTAAG	Amplificación completa del gen <i>sarX</i> en <i>S. aureus</i>
GP510	TCGTGCATGCATAATTTTAAAAATTGTTTC	
GP438	TAATACGACTCACTATAGGG	Primer promotor T7

Los 2 fragmentos fAB y fCD fueron purificados a partir del gel de agarosa usando el kit QIAkit Gel Extraction (Qiagen®) y utilizados en una PCR de solapamiento (“*overlap extension PCR*” [98]) para obtener un único fragmento de DNA (fAD) con un tamaño de 1200 pb usando adicionalmente los primers qrp_A y qrp_D y la DNA polimerasa “Phusion”®. El fragmento fAD fue purificado a partir del gel y clonado en el plásmido pJET® usando DNA ligasa y posteriormente transformado en células de *E. coli* XL1 blue por electroporación usando las siguientes condiciones: 1250 v, 25uF, 200Ω y un tiempo de exposición entre 3.9 y 4 ms. La suspensión bacteriana fue crecida en 1 ml de medio

LB durante 1 hora a 37°C a 250 rpm y las bacterias transformantes fueron seleccionadas en placas de agar LB suplementadas con 100 µg/mL de ampicilina. El plásmido recombinante pJET-AD (Tabla 4-1) fue confirmado por medio de PCR usando los primers pJET-S, pJET-AS, qrp_B y qrp_C.

El fragmento fAD fue liberado del plásmido pJET-AD por medio de restricción con las enzimas *BamHI* y *EcoRI* (*FastDigest* Thermo Scientific®) de acuerdo a las condiciones sugeridas por el fabricante, usando los sitios de restricción que fueron incorporados previamente con los primers (Tabla 4-2), purificado y clonado en el plásmido pMAD, el cual fue previamente cortado con las enzimas *BamHI* y *EcoRI*. El plásmido recombinante pMAD-AD (Tabla 4-1) fue transformado en células de *E. coli* XL1 blue por electroporación usando las condiciones anteriormente descritas. Las bacterias transformantes fueron seleccionadas en placas de agar LB suplementadas con 100 µg/mL de ampicilina y X-Gal (80 µg/mL). El plásmido recombinante pMAD-AD (Tabla 4-1) fue confirmado por medio de PCR usando los primers qrp_A y qrp_B y secuenciación convencional (Sanger). Luego el plásmido recombinante pMAD-AD fue purificado desde las células de *E. coli* y transformado en la cepa RN4220 de *S. aureus* por electroporación usando las condiciones anteriormente descritas. Las células transformantes fueron crecidas en 1 mL de medio LB por 3 horas, incubadas a 28°C (temperatura a la cual el plásmido pMAD puede replicarse) a 200 rpm y luego seleccionadas en cajas de TSA (agar tripticasa-soya) suplementadas con eritromicina (1,5 µg/mL) y X-Gal (80 µg/mL) e incubadas a 28°C durante 2 o 3 días. El plásmido recombinante es primero transformado en la cepa RN4220, la cual es deficiente en el sistema de modificación-restricción y permite la adquisición de DNA foráneo, esto favorece la aceptación y permanencia del plásmido en otras cepas de *S. aureus*.

El plásmido pMAD-AD, previamente confirmado por PCR, fue purificado desde la cepa RN4220 y transformado en la cepa NCTC8325-4 por electroporación usando las condiciones anteriormente descritas. Las células transformantes fueron crecidas en 1 mL de medio SMMP por 3 horas, incubadas a 28°C a 200 rpm y luego seleccionadas en cajas de TSA (agar tripticasa-soya) suplementadas con eritromicina (1,5 µg/mL) y X-Gal (80 µg/mL) e incubadas a 28°C durante 2 o 3 días. Para la confirmación molecular de las bacterias transformadas, las colonias azules fueron sembradas independientemente en 5 mL de medio TSB suplementado con eritromicina (1,5 µg/mL) a 28°C a 250 rpm

durante 24 horas. Después de confirmar que la cepa NCTC8325-4 había adquirido y mantenido el plásmido pMAD-AD se procedió a inducir la delección del gen *qrp*, para esto 5 colonias fueron sembradas independientemente en 5 mL de medio TSB suplementado con eritromicina (1,5 µg/mL) a 44°C a 250 rpm durante 24 horas. Luego se realizó una dilución seriada 1:10 por tres veces (dilución final 1:1000) y 60 µL de cada suspensión bacteriana fueron crecidos en placas de agar TSB suplementado con eritromicina (1,5 µg/mL) y X-Gal (80 µg/mL) a 44°C durante 24 horas. Después de este tiempo, las colonias de color celeste (no las colonias azules) fueron resembradas en placas de agar-TSB suplementado con eritromicina (1,5 µg/mL) y X-Gal (80 µg/mL) a 44°C durante 24 horas. De cada caja se seleccionaron entre 2 y 5 colonias y fueron crecidas de forma independiente en 5 mL de caldo TSB sin antibiótico a 28°C a 200 rpm durante toda la noche (16 horas). A partir de este crecimiento bacteriano se realizó una dilución seriada 1:10 por 6 veces (dilución final 1:10⁻⁶) y 60 µL de cada suspensión bacteriana fueron crecidos en placas de agar TSB suplementado X-Gal (80 µg/mL) (sin antibiótico) a 37°C durante 48 horas o hasta que se distinguieron bien las colonias blancas. En estas colonias blancas existe un 50% de probabilidad que correspondan a cepas mutantes, por lo cual es necesario confirmar la mutación por PCR en el mayor número de colonias. Para esto cada colonia fue crecida por duplicado en placas de agar suplementadas con o sin eritromicina (1,5 µg/mL) y X-Gal (80 µg/mL) e incubadas a 28°C durante 24 horas. A las colonias blancas que crecieron en el agar TSB y que no crecieron en el agar suplementado con eritromicina se les extrajo el DNA y la delección del gen *qrp* fue confirmada por medio de PCR utilizando los primers *qrp_F* y *qrp_E*. Este mismo procedimiento fue realizado para las cepas 15981 y 132. Las cepas mutantes fueron almacenadas a -80°C en medio TSB con glicerol al 10% hasta su uso.

Obtención de cepas complementadas y sobre-expresadas. En la figura 4-2 se muestran los pasos realizados para la obtención de las cepas complementadas y sobre-expresadas, para lo cual se utilizó el plásmido pCN51 [97]. Se diseñaron primers para amplificar el marco abierto de lectura completo del gen *qrp* desde la cepa NCTC8325-4 (*qrp_S*, *qrp_AS*, Tabla 4-2). Los primers utilizados tenían incorporados sitios de reconocimiento para las enzimas *Bam*HI y *Eco*RV. El gen completo amplificado (357 pb) fue purificado a partir del gel de agarosa y clonado en el plásmido pJET® usando DNA ligasa y posteriormente transformado en células de *E. coli* XL1 blue por electroporación

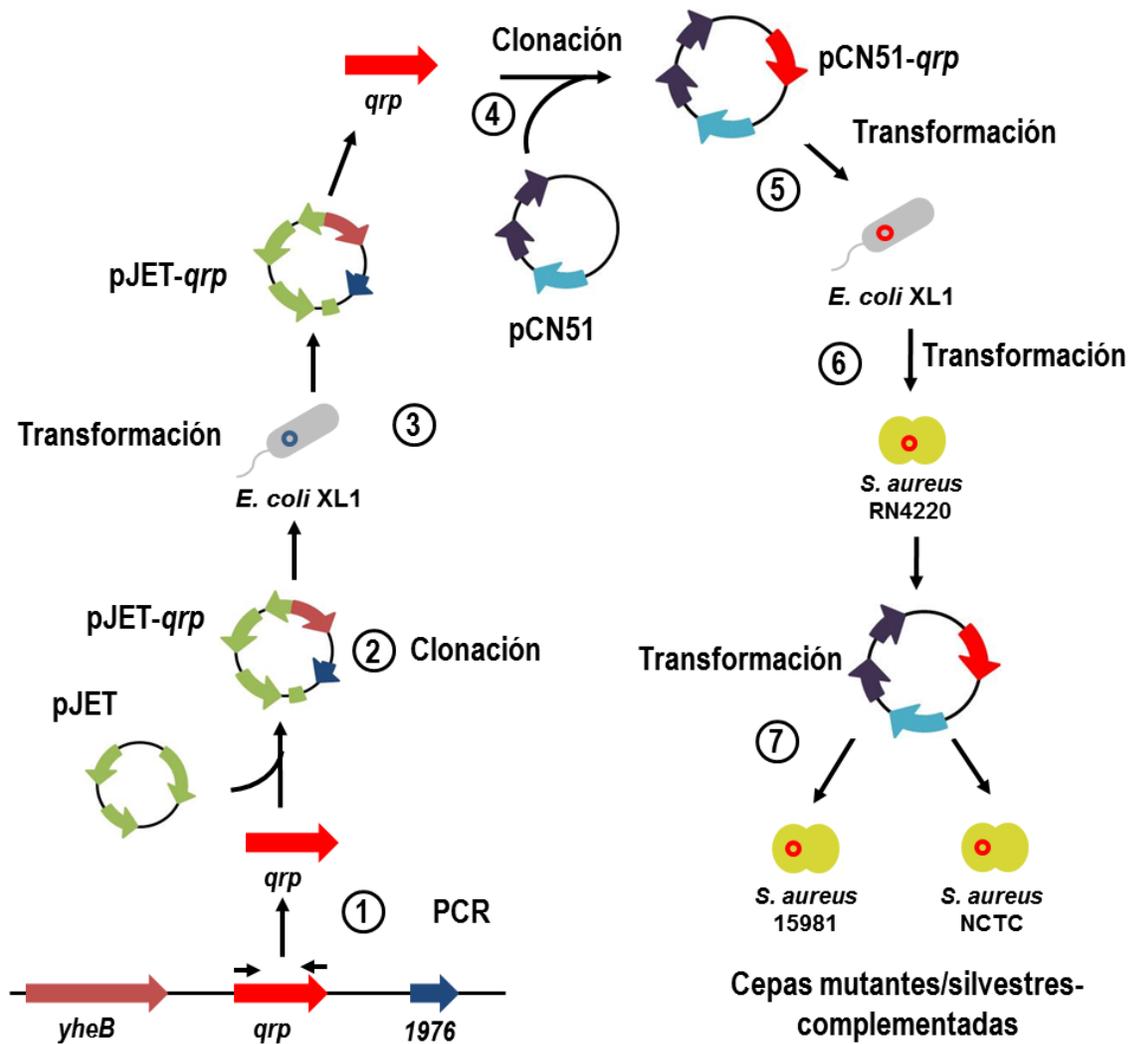


Figura 4-2: Proceso para la generación de las cepas complementadas y sobre-expresadas. 1. Amplificación por PCR del ORF completo del gen *qrp*, 2. Clonación del gen *qrp* en el plásmido pJET, 3. Transformación del plásmido recombinante pJET-*qrp* en *Escherichia coli* XL1 blue, 4. Liberación y clonación del gen *qrp* en el plásmido pCN51, 5. Transformación del plásmido recombinante pCN51-*qrp* en *Escherichia coli* XL1 blue, 6. Purificación y transformación del plásmido pCN51-*qrp* en la cepa RN4220 de *Staphylococcus aureus*, 7. Purificación y transformación del plásmido pCN51-*qrp* en las cepas 8325-4 *qrp*⁻ (mutante) y 8325-4 (silvestre) y confirmación por PCR.

usando las siguientes condiciones: 1250 v, 25uF, 200Ω y un tiempo de exposición de 4 ms. La suspensión bacteriana fue crecida en 1 mL de medio LB durante 1 hora a 37°C a

250 rpm y las bacterias transformantes fueron seleccionadas en placas de agar LB suplementadas con 100 µg/mL de ampicilina. El plásmido recombinante pJET-*qrp* (Tabla 4-1) fue confirmado por medio de PCR usando los primers pJET-S, pJET-AS, *qrp_S* y *qrp_AS*.

El gen *qrp* fue liberado del plásmido pJET-*qrp* por medio de restricción con las enzimas *Bam*HI y *Eco*RI (*FastDigest* Thermo Scientific®), purificado y clonado en el plásmido pCN51, el cual también fue previamente tratado con las enzimas *Bam*HI y *Eco*RI. El plásmido recombinante pCN51-*qrp* (Tabla 4-1) fue transformado en células de *E. coli* XL1 blue por electroporación usando las condiciones anteriormente descritas. Las bacterias transformantes fueron seleccionadas en placas de agar LB suplementadas con 100 µg/mL de ampicilina a 37°C. El plásmido recombinante pCN51-*qrp* fue confirmado por medio de PCR usando los primers *qrp_S* y *qrp_AS*, purificado y transformado en la cepa RN4220 de *S. aureus* por electroporación usando las condiciones anteriormente descritas. Las células transformantes fueron crecidas en 1 mL de medio SMMP a 37°C y 200 rpm por 3 horas; y luego seleccionadas en cajas de TSA (agar tripticasa-soya) suplementadas con eritromicina (10 µg/mL) e incubadas a 37°C durante 16 horas (O/N). A partir de las bacterias transformantes se purificó el plásmido pCN51-*qrp* y se transformó nuevamente en las cepas: 8325-4 *qrp*⁻ (mutante) y 8325-4 (silvestre) por electroporación con las condiciones anteriormente descritas. Las células transformantes fueron crecidas en 1 mL de medio LB a 37°C y 200 rpm por 3 horas; y luego seleccionadas en cajas de TSA (agar tripticasa-soya) suplementadas con eritromicina (10 µg/mL) e incubadas a 37°C durante 16 horas (O/N). Las cepas complementada (8325-4 *qrp*⁻ pCN5-*qrp*) y sobre-expresada (8325-4 pCN5-*qrp*) fueron confirmadas por PCR y almacenadas a -80°C en medio TSB con glicerol al 10% y eritromicina (1,5 µg/mL) hasta su uso.

4.1.3 Ensayos de formación de biofilm y hemólisis en *Staphylococcus aureus*

La evaluación del posible efecto de la delección del gen *qrp* en la formación de biofilm fue determinado como sigue: una colonia de cada cepa fue sembrada en 5 mL de caldo TSB suplementado con glucosa al 0,25% (p/v) y eritromicina (10 µg/mL) e incubadas a 37°C y 200 rpm durante 16 horas (O/N). Al siguiente día, la densidad óptica (DO) a 600 nm fue

determinada, y a partir de este valor un 1 mL de una nueva suspensión bacteriana fue preparado a una DO de 0.1 unidades de absorbancia. Luego, 5 μ L de esta suspensión bacteriana más 195 μ L de caldo TSB suplementado con glucosa al 0,25% (p/v) fueron colocados en una microplaca de 96 pozos de poliestireno de fondo plano e incubada a 37°C en quietud por 24 horas. Después, el sobrenadante fue retirado y la microplaca lavada tres veces por inmersión en agua con cloro, fue secada por inversión y 200 μ L de solución de cristal violeta fueron adicionados a cada pozo. Al cabo de 2 minutos, se descartó la solución de cristal violeta y la microplaca fue lavada nuevamente por inmersión en agua destilada tres veces, 200 μ L de una solución de etanol:acetona (4:1) fueron adicionados y la placa fue dejada 30 minutos a temperatura ambiente. Finalmente, la solución coloreada (200 μ L) fue transferida a otra microplaca limpia y su absorbancia fue medida a 595 nm en un lector de Elisa.

Adicionalmente se evaluó el efecto de la delección del gen *qrp* en uno de principales y más conocidos mecanismos de virulencia que tiene *S. aureus*, la hemólisis de eritrocitos. Este ensayo de hemólisis fue realizado en agar sangre y en microplaca. Para el ensayo en agar, una colonia de cada cepa fue crecida en 3 mL de medio BHI a 37°C durante 16 horas, luego se centrifugó a 900 g por 10 minutos, el sobrenadante fue retirado y las bacterias lavadas con solución salina al 0,89% estéril y nuevamente centrifugadas. Luego 2 μ L del aglomerado bacteriano fueron colocados directamente en las cajas de agar sangre al 5%, las cuales fueron incubadas a 37°C durante 24 o 48 horas. Para el ensayo de microplaca, se partió del mismo aglomerado bacteriano para preparar una suspensión bacteriana equivalente a un MacFarlan de 2 (0,4 UA), de la cual se tomaron 30 μ L y fueron colocados en la microplaca más 270 μ L de una suspensión de eritrocitos de cordero al 3%, previamente lavados con solución salina hasta que el sobrenadante es traslucido (aproximadamente 8 lavados). Como control de hidrólisis total, 270 μ L de la suspensión de eritrocitos fueron tratados con 20 μ L de SDS al 1% (100% de hidrólisis). La microplaca fue incubada en quietud a 37°C durante 24 horas. Después la microplaca fue centrifugada a 500 g y 100 μ L del sobrenadante fueron transferidos a una nueva placa limpia de fondo redondo, la absorbancia fue evaluada a 540 nm en un lector de Elisa. El porcentaje de hemólisis fue calculado teniendo como referencia el valor de los eritrocitos tratados con SDS (100% de hidrólisis). La significancia estadística entre las diferentes cepas (silvestre vs mutantes) en los dos ensayos (formación de biofilm y hemólisis) fue calculada por medio de una prueba *t* a dos colas, con mínimo tres replicas.

4.1.4 Producción de la proteína recombinante Qrp

También se realizaron experimentos para evaluar la interacción sonda-proteína usando la proteína recombinante Qrp (rQrp).

Amplificación y clonación del gen *qrp*. Se diseñaron primers específicos para la amplificación del gen *qrp* completo por PCR, no se incluyó el codón de parada para permitir la adición de la cola de 6 histidinas en el extremo C-terminal usada para facilitar su purificación. También se diseñó una pareja de primers internos para el gen para confirmar la generación de los diferentes plásmidos recombinantes por PCR (Tabla 4-2). Todos los primers fueron diseñados por medio de los programas Primer3 y BLAST-primer disponibles en <http://bioinfo.ut.ee/primer3-0.4.0/> y <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>. Para facilitar la clonación del gen *qrp*, una secuencia de reconocimiento para las enzimas de restricción *Xba*I y *Nsi*I fueron adicionadas en el extremo 5' de los primers. La amplificación del gen *qrp* se realizó por PCR a partir de 100 ng de DNA fresco de la cepa USA300.

El producto amplificado (gen completo) fue purificado a partir de un gel de agarosa y clonados en el vector pGEM-T easy (Promega®) siguiendo las recomendaciones del fabricante. El plásmido recombinante fue posteriormente transformado en la cepa de *Escherichia coli* TOP10 por choque térmico de acuerdo a las condiciones previamente estandarizadas en nuestro Laboratorio, la suspensión bacteriana fue crecida en placas de agar con medio LB suplementado con ampicilina (100 µg/mL), IPTG (2mM), X-Gal (80 µg/mL) a 37 °C por toda la noche. Las bacterias transformadas con el plásmido (colonias blancas) fueron confirmadas por medio de PCR usando combinaciones de primers internos del gen y específicos del vector (Tabla 4-2). El plásmido recombinante pGEM-T1 fue purificado, y junto con el vector Champion™ pET303/CT-His (Invitrogen®), fueron sometidos a doble digestión, primero con la enzima de restricción *Xba*I (10U) y luego *Nsi*I (10U) (Thermo Scientific®). El fragmento de DNA liberado del plásmido fue purificado a partir del gel y ligado en el vector pET303/CT previamente digerido, utilizando la enzima T4 DNA ligasa (Invitrogen®) por 1 hora a temperatura ambiente, luego fue transformado en la cepa de *Escherichia coli* TOP10 por choque térmico, la cual fue crecida en placas de agar con medio LB suplementado con ampicilina (100 µg/mL) a 37 °C por toda la

noche. Las bacterias transformadas con el plásmido fueron confirmadas por medio de PCR. El plásmido pET-CT1 (Tabla 4-1) fue purificado y transformado en la cepa *Escherichia coli* BL21, la cual fue crecida en medio LB suplementado con ampicilina (100 µg/mL) a 37 °C por toda la noche. Las bacterias transformadas con el plásmido fueron confirmadas por medio de PCR y secuenciación.

Producción y purificación de la proteína rQrp. Después del proceso de optimización, se encontró que la proteína rQrp se produjo de forma soluble y con una elevada expresión a las 3 horas de inducción con IPTG. A partir de esto se inició su producción a mayor escala y purificación. Una colonia de la cepa *Eco* BL21-*grp* (Tabla 4-1) fue crecida en 50 mL de caldo LB suplementado con ampicilina (100 µg/mL) e IPTG (1mM) a 37 °C durante 3 horas a 150 rpm, luego la suspensión bacteriana se colocó en hielo y se centrifugó a 1700 g durante 5 minutos, el sobrenadante fue descartado y las bacterias suspendidas en 2 mL de buffer de unión (NaH₂PO₄ 50 mM, NaCl 0,5 M pH: 8,0; imidazol 25 mM) y sonicadas como fue descrito anteriormente. Esta suspensión se centrifugó a 7.000 g durante 10 minutos a 4°C, el sobrenadante (extracto de proteínas solubles) fue transferido a nuevo tubo y usado inmediatamente para la purificación de la proteína recombinante utilizando una resina de níquel (Ni-NTA Purification system, Invitrogen®) acoplada al sistema FPLC. Se optimizaron las condiciones de corrido así: La columna Ni-NTA fue equilibrada con 10 mL de buffer de unión (NaH₂PO₄ 50 mM, NaCl 0,5 M pH: 8,0;imidazol 25 mM), luego 10 mL de buffer de elución (NaH₂PO₄ 50 mM, NaCl 0,5 M pH: 8,0, imidazol 250 mM) y 10 mL de buffer de unión; se inyectó un 1 mL del sobrenadante del extracto de proteínas y se mantuvo en interacción con la resina durante 30 minutos, finalmente la proteína fue eluida con buffer de elución y se recolectaron fracciones de 500 µL. La purificación de la proteína recombinante fue confirmada por SDS-PAGE.

La unión de la proteína recombinante a las sonda de DNA fue determinada por EMSA utilizando las condiciones descritas para los extractos totales de la cepa USA300, se emplearon 1500 fmoles de sonda y diferentes cantidades de proteína recombinante.

4.2 Resultados

4.2.1 Evaluación del efecto de la delección del gen *qrp* en la formación de biofilm y hemólisis de eritrocitos en *Staphylococcus aureus*.

Con la identificación de la proteína Qrp en los ensayos de interacción con la sonda y el análisis de su estructura, el siguiente objetivo fue evaluar su participación en la formación del biofilm en *S. aureus*. La estrategia usada fue determinar el cambio en la formación de biofilm cuando se hace la delección completa del gen *qrp* dentro del genoma de la bacteria. La delección del gen *qrp* se realizó empleando la metodología de recombinación homóloga con el sistema pMAD [96], una metodología basada en el reemplazo alélico por recombinación homóloga en dos pasos, el primero con una integración del plásmido recombinante pMAD en el genoma y el segundo con una resolución del plásmido, el cual es curado cuando la cepa es crecida a una temperatura $>37^{\circ}\text{C}$. Adicionalmente, el proceso es seguido visualmente por la acción de la enzima β -galactosidasa sobre X-gal. La metodología estuvo compuesta de una serie de pasos que incluyó: construcción del plásmido recombinante pMAD con las secuencias que rodean el gen *qrp*, para dirigir y hacer específica la recombinación alélica. Este plásmido fue primero transformado en *E. coli*, para propagarlo, luego transformado en la cepa RN4220 de *S. aureus*, para que sufra las modificaciones adecuadas (metilaciones especie específicas) y pueda ser adquirido y mantenido eficientemente por las cepas de *S. aureus* donde se realizará la delección.

La delección del gen *qrp* en las tres cepas diferentes de *S. aureus* (NCTC8325, 15981 y 132) fue confirmada por medio de dos PCR, usando dos combinaciones diferentes de primers (Figura 4-3). Con las mutantes confirmadas, se procedió a determinar el efecto de la delección en la capacidad de formar biofilm. Como se muestra en la figura 4-4, cuando el gen *qrp* es eliminado, hay una disminución entre el 20 y 25% en la formación de biofilm de las cepas mutantes comparado con la cepa silvestre. Esta disminución fue encontrada tanto en la cepa NCTC8325 como en la 15981 y además en dos colonias de la misma cepa que realizaron el proceso de mutagénesis de manera independiente. En la colonia 1 de la cepa 15981 la disminución fue mayor (43%). Estas diferencias fueron estadísticamente significativas ($n=6$). Cuando la cepa mutada fue complementada con el gen, la formación de biofilm fue recuperada, no hasta el 100%, pero si hasta el 90%.

Estos resultados sugieren que el gen *qrp* podría estar relacionado con la formación de biofilm, aunque su ausencia no suprime totalmente esta capacidad, si la afecta ligeramente.

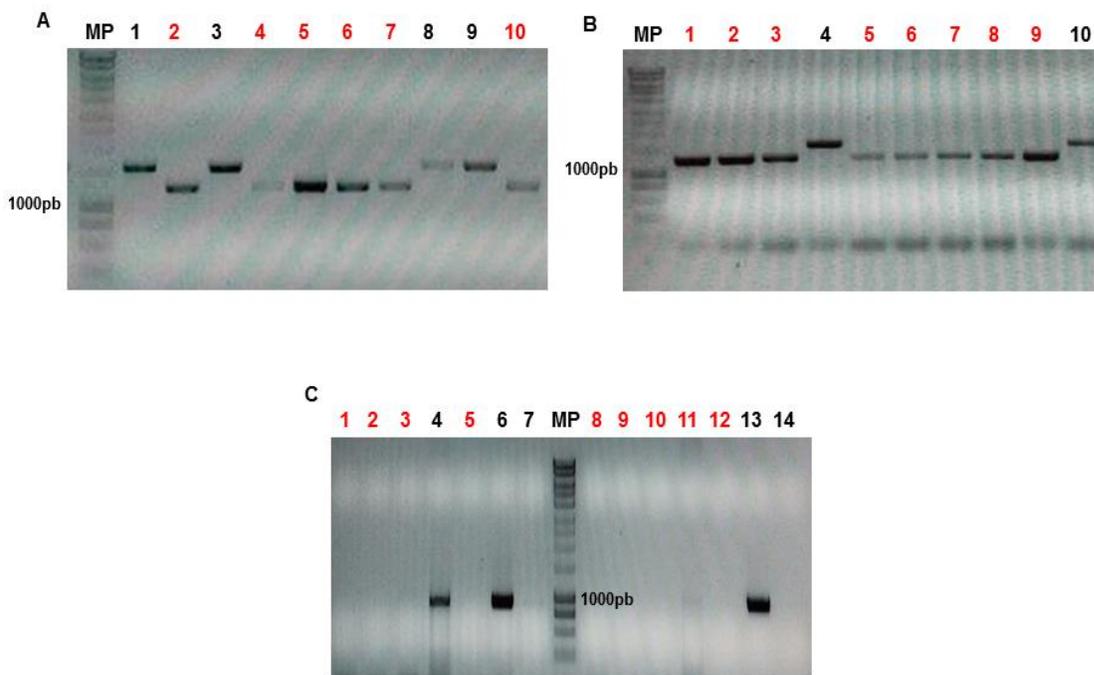


Figura 4-3: Confirmación molecular de la delección del gen *qrp* en las cepas de *Staphylococcus aureus*. **A.** colonias de la cepa NCTC8325 usando los primers *qrp_E* y *qrp_F* (Tabla 4-2), el tamaño esperado para las colonias mutantes (números en rojo) es de 1315 pb y para las colonias silvestres 1611 pb; **B.** colonias de la cepa 15981 usando los primers *qrp_E* y *qrp_F* y **C.** confirmación de las colonias mutantes de las tres cepas usando los primers *qrp_E* y *qrp_AS*, no se espera amplificación en las colonias mutantes (números en rojo) y el tamaño esperado para las colonias silvestres es de 935 pb.

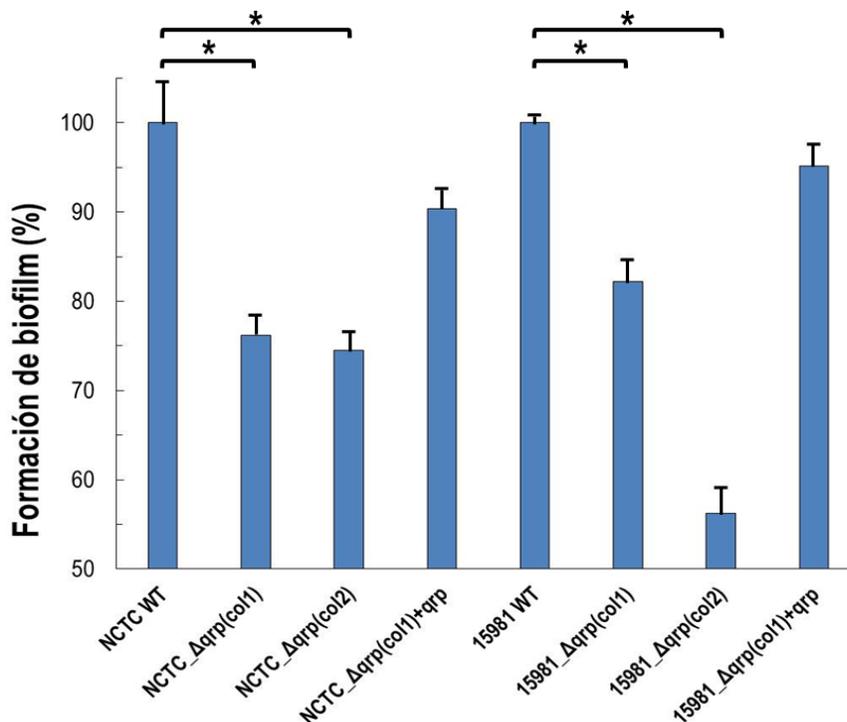


Figura 4-4: Efecto de la delección del gen *qrp* en la formación de biofilm en *Staphylococcus aureus*. Evaluación de la capacidad de formar biofilm de las cepas silvestres NCTC8325 y 15981, y sus respectivas cepas mutantes y complementadas. Adicionalmente se muestra el resultado obtenido en dos colonias donde el proceso de mutagénesis se realizó de manera independiente para cada cepa. El asterisco denota que hay una significancia estadística entre los valores de las medias de los dos grupos de datos usando la prueba t-Student, a un intervalo de confianza del 95% y dos colas (n=6).

Adicionalmente a la formación del biofilm, se determinó el efecto de la delección del gen *qrp* en la hemólisis de eritrocitos, uno de los mecanismos de virulencia más conocidos de *S. aureus*. Contrario a lo que se observó en biofilm, la delección del gen ocasionó un ligero aumento de la capacidad de hemólisis de la cepa NCTC8325-4, un mayor halo de hemólisis en la placa de agar y un aumento en el ensayo de microplaca (aproximadamente 8%), respecto a la cepa silvestre (Figura 4-5). Cuando la cepa mutante es complementada o la proteína es sobre-expresada en la silvestre, se observa una disminución en la hemólisis, cercana al 14%. Sin embargo, estas diferencias no fueron estadísticamente significativas.

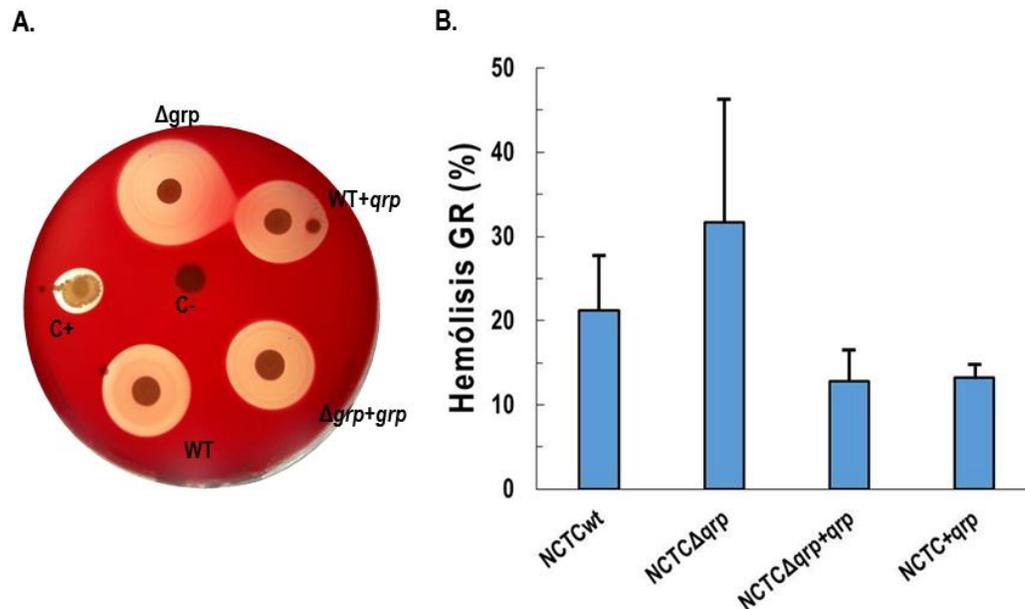


Figura 4-5: Efecto de la delección del gen *qrp* en la hemólisis de eritrocitos por *Staphylococcus aureus*. Evaluación de la capacidad de hemólisis de la cepa silvestre NCTC8325, y su respectiva cepa mutante, complementada y sobre-expresada. A. Ensayo en placa de agar de la cepa silvestre (WT), mutada (Δgrp), sobre-expresada (WT+*qrp*) y complementada ($\Delta grp+qrp$). C+: Control positivo (cepa de *Staphylococcus aureus* ATCC29213), C-: Control negativo (cepa de *Klebsiella pneumoniae*). B. Ensayo de microplaca. No se encontró una significancia estadística entre los valores de las medias de los dos grupos de datos usando la prueba t-Student, a un intervalo de confianza del 95%.

4.2.2 Interacción Sonda-Proteína recombinante

Producción de la proteína recombinante rQrp. La amplificación del gen *qrp* completo (sin codón de parada) generó un producto de 366 pb, el cual después de ser purificado fue ligado en el vector pGEM-T para facilitar el corte con las enzimas *Xba*I y *Nsi*I. Luego de la digestión con las enzimas fue clonado en el vector pET-CT. Los plásmidos recombinantes fueron confirmados molecularmente por medio de PCR utilizando primers específicos del gen y del vector (Tabla 4-2) y por secuenciación (Figura 4-6).

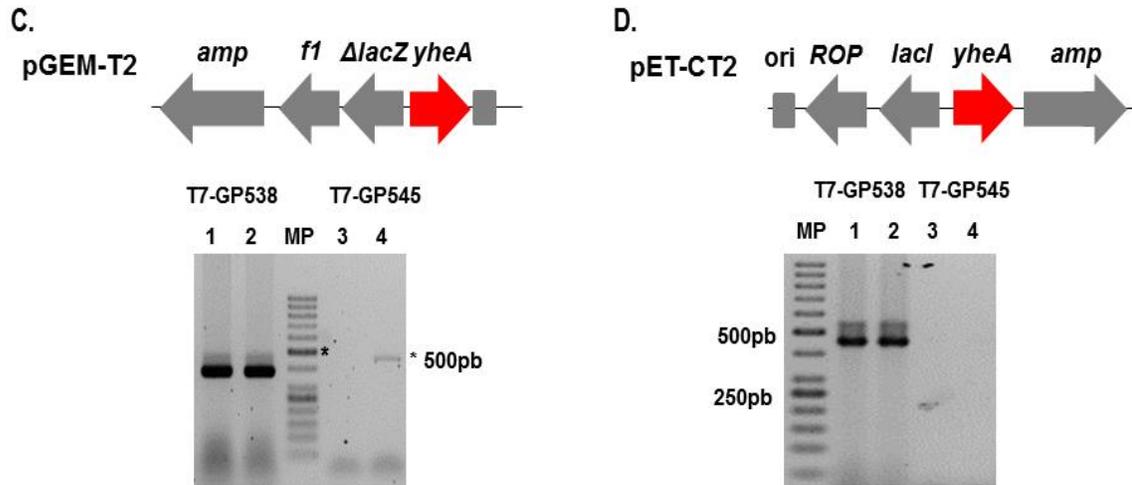


Figura 4-6: Confirmación molecular de los plásmidos recombinantes de los genes *qrp* y *sarX*. Amplificación por PCR de genes específicos de los diferentes plásmidos recombinantes generados mostrando la dirección del gen. **A y B.** Plásmidos pGEM-T1 y pET-CT1 (gen *qrp* clonado dentro del vector pGEM-T y pET-CT, respectivamente), se confirmaron dos plásmidos purificados desde dos colonias transformadas diferentes. Carril 1 y 2: PCR con los primers T7-GP538 (antisentido del gen), y Carril 3 y 4: PCR con los primers T7-GP545 (sentido del gen). **C y D.** Plásmidos pGEM-T2 y pET-CT2 (gen *sarX* clonado dentro del vector pGEM-T y pET-CT, respectivamente), se confirmaron dos plásmidos purificados desde dos colonias transformadas diferentes. Carril 1 y 2: PCR con los primers T7-GP510 (antisentido del gen), y Carril 3 y 4: PCR con los primers T7-GP509 (sentido del gen). MP: Marcador de masa molecular.

Adicionalmente, la PCR y la secuenciación indicaron que el gen *qrp* estaba en la dirección correcta para su expresión e inserción de las 6 histidinas. El plásmido pET-CT1 fue transformado en la cepa de *E. coli* BL21 para la expresión de la proteína recombinante. La proteína rQrp fue purificada a partir de 50 mL de cultivo de la cepa transformada inducida con IPTG (ver materiales y métodos). Una separación de un extracto total de proteínas solubles de la cepa transformada e inducida con IPTG por SDS-PAGE muestra la expresión de una proteína de aproximadamente 15 kDa, que no se observa en la cepa sin inducir (Figura 4-7B), y que corresponde a la proteína rQrp; con este resultado se procedió a su purificación por FPLC.

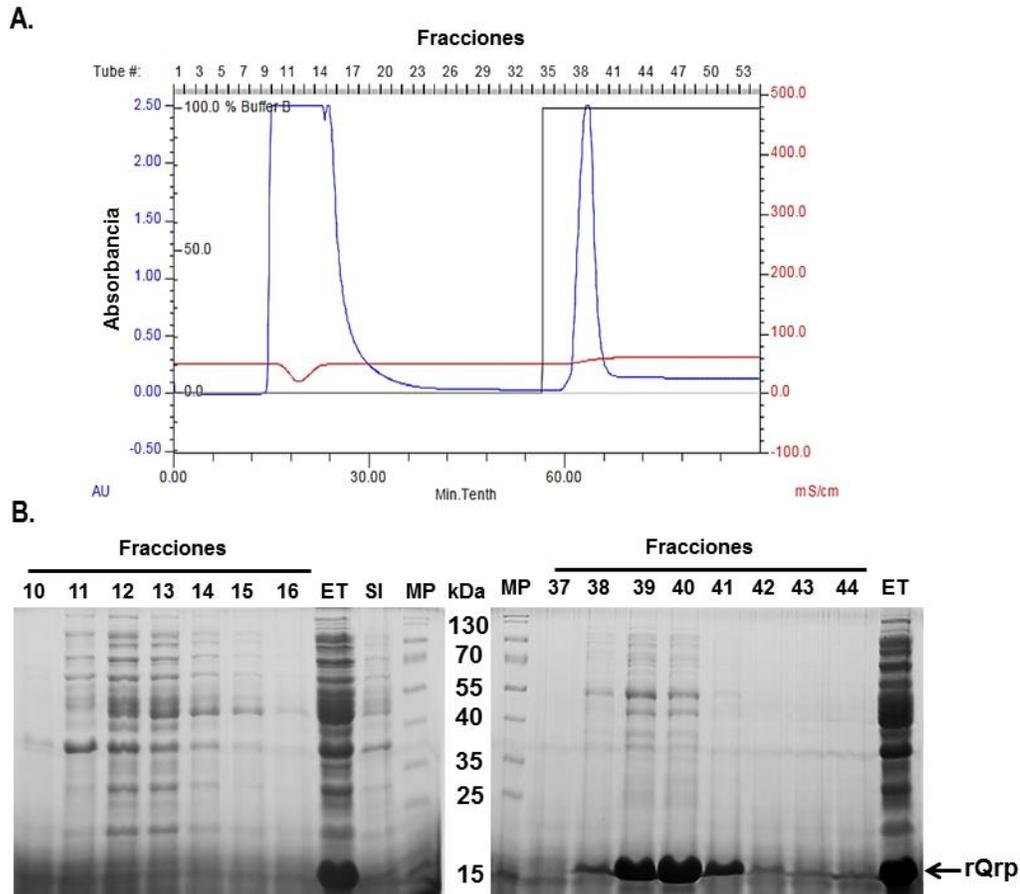


Figura 4-7: Seguimiento de la purificación de la proteína recombinante rQrp. A. Cromatograma del proceso de purificación de la proteína rQrp por medio de cromatografía de afinidad con la resina Ni-NTA. La línea negra indica el inicio del buffer de elución suplementado con imidazol. **B.** SDS-PAGE de las fracciones obtenidas de la cromatografía de afinidad. ET: Extracto total de proteínas con inducción con IPTG, SI: extracto total de proteínas sin inducción. MP: marcador de masa molecular.

Como se muestra en la figura 4-7, la cromatografía de afinidad con la resina de níquel del extracto total produjo dos picos principales, el primero correspondiente a las proteínas que no se unen a la fase estacionaria (que son la mayor cantidad de proteínas del extracto) y el segundo que se produce exactamente después de la aplicación del buffer de elución con imidazol, en las fracciones 38 a 40, el cual corresponde a la proteína rQrp, como es confirmado por SDS-PAGE (Figura 4-7B). Aunque se observa la presencia de otras proteínas contaminantes, la proteína de interés estaba en una mayor cantidad, por

lo cual estas dos fracciones fueron utilizadas para determinar la posible unión entre la proteína rQrp y la sonda de DNA que simula la secuencia palindrómica del gen *icaA*.

Interacción sonda-proteína: De la cromatografía de afinidad se obtuvieron dos fracciones (500 μ L cada una) con proteína recombinante, a partir de las cuales se determinó su capacidad de unión a la sonda de DNA, para esto, diferentes cantidades de la proteína (entre 5 μ g a 80 μ g) fueron mezcladas con 1500 fmoles de sonda (la misma cantidad usada con los extractos totales) y evaluados por EMSA de acuerdo a las condiciones empleadas previamente. Sin embargo, ninguna señal de retardamiento fue obtenida (datos no mostrados), aun colocando mayor cantidad de proteína. Este resultado puede deberse a varios factores como: poca cantidad de proteína, mal plegamiento de ésta, interferencia de algún componente del proceso de expresión y purificación o necesidad de otras proteínas para su unión a DNA. Para intentar determinar la necesidad de otras proteínas en la interacción con el DNA, el EMSA se realizó adicionando pequeñas cantidades de extracto total de la cepa USA300 (50 μ g a 3 μ g) a rQrp, es importante mencionar que con 50 μ g de extracto total se logra obtener una señal débil de retardamiento (Figura 2-2); aunque se esperaba intensificar la señal con la mezcla del extracto total y la proteína recombinante. No se obtuvo ninguna señal de retardamiento de la sonda (Figura 4-8), es decir, la solución de proteína recombinante eliminaba completamente la señal, lo cual sugiere que probablemente el buffer de elución interfiere con la unión DNA-proteína.

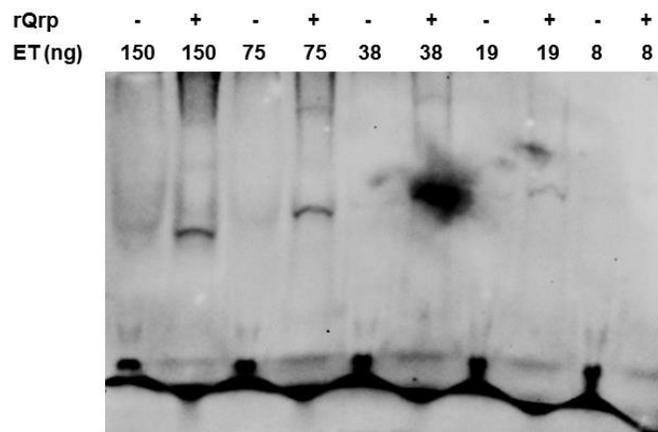


Figura 4-8: Ensayos de retardamiento en gel de la proteína rQrp y la sonda de DNA de la secuencia palindrómica. Evaluación de la interacción de la proteína rQrp (80 ng) con la sonda de DNA (1500 fmoles) en presencia de diferentes cantidades de extracto total de proteínas (Fracción 40, Figura 4-7).

Con el propósito de eliminar el buffer de elución y purificar aún más la proteína rQrp, se realizó una CET de las fracciones 39 y 40 (Figura 4-9). En la CET se obtuvo un pico mayoritario en las fracciones 29, 30 y 31, que, de acuerdo a los resultados del SDS-PAGE, corresponde a la proteína rQrp (Figura 4-9), la cual aparece más pura y no se detectan proteínas contaminantes.

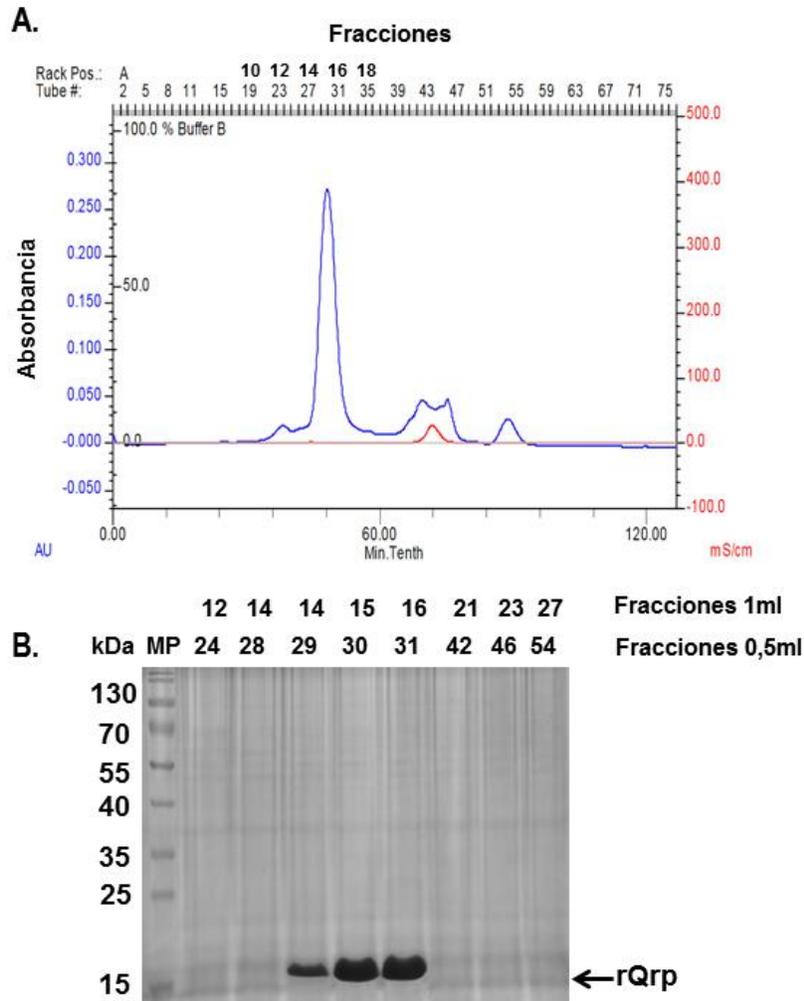


Figura 4-9: Comportamiento de la proteína recombinante rQrp en la CET. A. Cromatograma de la proteína rQrp en la CET. Los primeros números indican las fracciones recolectadas (0,5 mL) y su relación respecto a los valores que tendrían las fracciones de 1 mL (valores en negrilla). **B.** SDS-PAGE de las fracciones obtenidas de la CET de la proteína rQrp. MP: marcador de masa molecular.

Un resultado que llama la atención es la relativamente rápida elución de la proteína rQrp, en fracciones donde se esperan proteínas de mayor tamaño (>50 kDa), al comparar con la proteína nativa, un comportamiento similar fue encontrado (Figura 2-3) y confirmado por el SDS-PAGE, donde la única proteína de tamaño pequeño fue Qrp, las restantes fueron superiores a 40 kDa (Figura 2-6C). Este resultado puede sugerir que la proteína Qrp podría estar formando complejos de mayor tamaño, aunque es necesario de más evidencia experimental. Ya con la proteína rQrp con mayor grado de pureza y sin el buffer de elución, se realizó nuevamente el EMSA con fracción 30 y la sonda de DNA. Como se muestra en la figura 4-10, la adición de la proteína rQrp parece aumentar las señales de la sonda retardada (en 4 y 2 ul) respecto al extracto total solo. Este resultado sugiere que la proteína Qrp si podría tener el potencial de unirse a la secuencia palindrómica del gen *icaA*, pero tal vez en forma de complejo con otras proteínas de la bacteria.

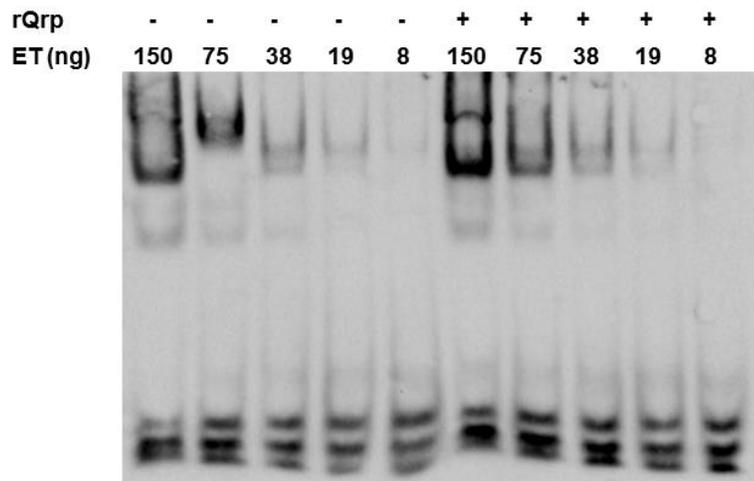


Figura 4-10: Ensayos de retardamiento en gel de la proteína rQrp y la sonda de DNA de la secuencia palindrómica. Evaluación de la interacción de la proteína rQrp (80 ng) con la sonda de DNA en presencia de diferentes cantidades de extracto total de proteínas después de la CET (Fracción 30, Figura 4-4).

4.3 Discusión

Las principales contribuciones al conocimiento de este estudio son: Primero, la identificación de una nueva familia de proteínas que contienen el dominio *com_ylbF* en *Staphylococcus aureus*; segundo, la participación de una de ellas (YheA/Qrp) en la formación de biofilm en esta bacteria, así como se ha visto en *Bacillus subtilis* para YlbF y YmcA; tercero, las tres proteínas con el dominio *com_ylbF* (YlbF, YmcA y Qrp/YheA) tienen una estructura tridimensional altamente conservada a pesar de su bajo porcentaje de identidad (<16%), cuarto, la identificación de un posible motivo funcional dentro de cada una de las tres proteínas y quinto, para Qrp (YheA), la probable importancia del aminoácido glutamina en su función.

El objetivo del presente estudio fue identificar una proteína de unión al gen *icaA* y determinar su participación en la formación del biofilm en *S. aureus*. Desde un enfoque bioquímico, se utilizó un proceso de purificación que incluyó varios pasos, no para tratar de tener la proteína 100% pura, sino para obtener un extracto enriquecido de la proteína de interés parcialmente purificada que facilitara su identificación por secuenciación. Esta proteína mantuvo su capacidad de unión a la sonda a largo del proceso de purificación, el cual incluyó diversos pasos basados en propiedades fisicoquímicas diferenciales como capacidad de agregación, separación por tamaño y por carga superficial. Estos resultados sugieren que probablemente la interacción sonda-proteína no es producto de un artefacto técnico, si no que puede existir una interacción real. Después de todo el proceso, y de secuenciar más de una docena de proteínas en las fracciones donde se detectaba la interacción, tanto en crecimiento platótico como formando biofilm, el análisis permitió identificar dos proteínas, SarX y una proteína hipotética (ortólogo de la proteína YheA de *B. subtilis*), la cual hemos denominado como proteína Qrp. Un estudio realizado por Cue y colaboradores ya había demostrado que SarX tenía la capacidad de unirse a la región 5' del gen *icaA* [72]. Respecto a la proteína Qrp, la simulación de su estructura tridimensional mostró que está compuesta exclusivamente de α -hélices y que puede actuar como un homodímero, adoptando una conformación espacial con un alto potencial de comportarse como una proteína de unión a DNA. Otra característica interesante encontrada en esta proteína Qrp fue su alto contenido de glutaminas (13,2%), algunas de ellas muy posiblemente implicadas en su función. En células eucariotas se ha descrito la importancia de la glutamina en algunos factores de transcripción [99].

La proteína YlbF fue identificada por primera vez en el año 2000 en cepas de *Bacillus subtilis* incapaces de producir esporas, volverse competentes y formar biofilm, producto de la truncación aleatoria del gen *ylbF* en experimentos de mutagénesis usando el transposón Tn10 [78]. El análisis de la proteína YlbF permitió encontrar un nuevo dominio proteico denominado *com_ylbF*, el cual hasta ese momento era específico de esta proteína. Este dominio tiene un tamaño de 120 aminoácidos y abarca casi toda la proteína (la proteína completa es de 141 aminoácidos). Sin embargo, dos proteínas adicionales con este dominio *com_ylbF* ya han sido identificadas, YmcA y YheA. La primera también identificada en experimentos de mutagénesis aleatoria en cepas de *B. subtilis* deficientes en la formación de biofilm [80], pero también su delección inhabilita a la bacteria para formar endosporas y adquirir DNA exógeno [79]. De la proteína YheA no se sabe absolutamente nada. Los análisis realizados muestran que los ortólogos de YheA están presentes en más bacterias del *phylum* firmicutes que los correspondientes a YlbF y YmcA. Es decir, la proteína YheA está más ampliamente distribuida que sus contrapartes, incluso dentro de bacterias que no tienen la habilidad de producir esporas como *S. aureus*, *Streptococcus* spp., *Enterococcus faecalis*, *Lactobacillus acidophilus*, entre otros (Figura 3-7). Este resultado sugiere que la proteína YheA podría estar involucrada en algunos procesos biológicos básicos conservados dentro del *phylum* y que van más allá de su participación exclusiva en la producción de esporas o formar biofilm.

Toda la información de la posible función de las proteínas YlbF y YmcA provienen de los estudios realizados en *B. subtilis*, nada es conocido sobre estas proteínas en otras bacterias Gram positivas. En este estudio se encontró que *S. aureus*, además de poseer el gen *yheA*, también posee los genes *ylbF* y *ymcA* (así como el gen *yaat*), inmersos en un entorno genético moderadamente diferente en las dos bacterias (Figura 3-3), producto de la inserción diferencial de genes. Curiosamente, el gen *yheB*, localizado corriente arriba del gen *yheA*, se ha mantenido conservado en las dos especies, sugiriendo que podría estar facilitando la función de YheA y que pueden actuar como un operón. Sin embargo, experimentos de transcriptómica usando RNA-seq indican que ellos son transcritos independientemente (datos no publicados). YheB es una proteína de membrana y tal vez pueda estar favoreciendo el reclutamiento y actividad de YheA hacia la membrana de la bacteria, donde se encuentra RNasa Y.

Aunque los modelos “*in silico*” de proteínas pueden contener errores, ellos pueden ser útiles en la predicción de posibles funciones desde el análisis de sus rasgos estructurales y lo más importante pueden eficientemente direccionar los experimentos biológicos [100]. Actualmente, dos roles, aparentemente contradictorios, han sido atribuidos a YlbF y YmcA (en interacción con YaaT); el primero, acelerando la fosforilación del regulador transcripcional Spo0A a través de la formación de dos complejos $(4\text{Fe-4S})^{+2}$ y la interacción con las proteínas Spo0B y Spo0F [79, 82, 83]; y el segundo, regulando la degradación del mRNA del represor SinR a través de una posible interacción con la endoribonucleasa RNasa Y [81, 92]. En el primer caso, la actividad de YlbF-YmcA-YaaT, radica en la formación de los complejos $(4\text{Fe-4S})^{+2}$ por YlbF y YaaT. Solo se ha relacionado el motivo metalotioneína de YlbF, el cual está localizado en el C-terminal de la proteína, por fuera del dominio *com_ylbF*, en la formación de estos complejos [83]. En el segundo caso la función de las proteínas no ha sido asociada a su estructura [81]. Sorprendentemente, en ninguna de las dos hipótesis el dominio *com_ylbF*, que constituye casi toda la proteína, se ha relacionado con su función. En este aspecto, los resultados del presente estudio aportan al nuevo conocimiento de esta familia de proteínas; el análisis “*in silico*” muestra que el dominio *com_ylbF* de estas tres proteínas, YlbF, YmcA y YheA tiene una estructura tridimensional altamente conservada, tanto en *B. subtilis* como en *S. aureus*, conformada únicamente por α -hélices (Figura 3-5, 3-6). Adicionalmente, un motivo putativo conformado por 4 a 7 aminoácidos conservados fue identificado en cada una de las tres proteínas, aunque su secuencia es diferente y específica de cada proteína, si mantienen una posición conservada dentro de la estructura tridimensional de la proteína en la tercera α -hélice (Figura 3-9). Ahora, si las tres proteínas actúan como un homodímero, tal como ha sido reportado experimentalmente para la proteína YmcA [79] (<http://www.rcsb.org/structure/2PIH>), los motivos putativos de cada proteína estarían orientados hacia el interior del dímero, donde ellos podrían interactuar más adecuadamente con sus sustratos simulando una “tenaza”, con la habilidad para abrir de manera diferencial sus “mandíbulas” (Figura 3-9).

El hallazgo de los motivos putativos inmediatamente promueve la necesidad de realizar experimentos de mutagénesis sitio-dirigida, alterando estos aminoácidos conservados (o los aminoácidos conservados de los extremos N y C terminal responsables tal vez de la dimerización de la proteína), no solo en Qrp (YheA), sino también en YlbF y YmcA, tanto en *B. subtilis* como en *S. aureus*, para confirmar la importancia de estos motivos en la

función de las proteínas y su relación con la formación de biofilm, esporas y competencia (estas dos últimas en *B. subtilis*).

El estado actual del conocimiento de las proteínas YheA, YlbF y YmcA aumenta la importancia de estudiar su función en otras bacterias como *S. aureus*, para proporcionar más evidencia experimental que soporten su participación ya sea en la fosforilación de Spo0A o en la maduración y decaimiento del RNA mensajero (degradación de SinR). En este sentido, este estudio aporta evidencia experimental sobre la participación de la proteína Qrp en la formación de biofilm en *S. aureus*, ya que la delección del gen *qrp* afecta moderadamente la formación de biofilm, así como se ha establecido que la delección de los genes *ylbF* y *ymcA* afectan la formación de biofilm en *B. subtilis*. Esta participación podría estar mediada por la interacción de la proteína Qrp con el gen *icaA* a través del reconocimiento de la secuencia palindrómica. Aunque es muy importante obtener más evidencia experimental que confirmen de manera contundente la interacción de la proteína con esta secuencia, ya que los ensayos con la proteína recombinante no fueron contundentes, tal vez debido a dos causas principales: mal plegamiento de la proteína recombinante o la necesidad de la interacción de Qrp con otras proteínas, como ha sido reportado para YlbF y YmcA, las cuales forman un complejo ternario “*in vivo*” con la proteína YaaT para ejercer su función [79]. Adicionalmente, se intentó profundizar en el efecto de la delección del gen *qrp* en otros procesos básicos de *S. aureus*, como la virulencia, ya que la delección de *ylbF* y *ymcA* afectan otros procesos a parte de la formación del biofilm en *B. subtilis* (como la competencia y formación de esporas), y un resultado inesperado fue encontrado, la supresión del gen produjo un aumento de la hemólisis de eritrocitos (Figura 4-7). Este resultado sugiere que Qrp podría estar afectando otros procesos de regulación de la transcripción más globales en la bacteria.

Curiosamente, *S. aureus* no posee proteínas ortólogas relacionadas con Spo0A ni tampoco con SinR, pero si se ha identificado un ortólogo de la proteína RNasa Y, lo cual apoyaría en mayor medida la hipótesis de la participación de las proteínas *com_ylbF* con la función de esta ribonucleasa. En *S. aureus* es muy limitada la información sobre las características y función de esta enzima, algunos estudios han revelado que RNasa Y no es esencial para la supervivencia de la bacteria, ya que la delección del gen que la codifica, *rny* (el cual se encuentra localizado muy cerca del gen *ymcA*, Figura 3-3) no produce cambios fenotípicos significativos en el crecimiento de la bacteria (contrario a lo que sucede en *B. subtilis*); sin embargo si tiene un impacto muy significativo en la

disminución de la virulencia en modelos “*in vivo*” [101-105]. Se ha podido establecer que RNasa Y participa en el procesamiento y degradación de algunos mRNA [105, 106], pero también participa en la estabilización de algunos transcritos de mRNA, como por ejemplo el operón *SaeRS* [103, 105, 107], el cual es un sistema de dos componentes que actúa como regulador global de algunos factores de virulencia [107]. Interesantemente, en *S. aureus* el regulador transcripcional global *agr* produce los mismos efectos fenotípicos a los encontrados con la delección del gen *qrp*, disminuye la formación del biofilm y aumenta la virulencia (entre ellos la hemólisis) [54]. *Agr* es un operón bifuncional que produce la transcripción de dos moléculas de RNA (RNAII y RNAIII), el primero codifica los genes *agrA*, *agrB*, *agrC* y *agrD*, relacionados con el sistema de quorum sensing (síntesis y respuesta del péptido autoinductor); y el segundo cumple dos funciones, alberga el gen para la δ -hemolisina, relacionada con la hemólisis de eritrocitos, y además funciona como RNA antisentido involucrado en la inhibición de varios genes que codifican proteínas de superficie en la bacteria, relacionadas con la formación de biofilm [54]. Entonces, cuando el operón *agr* es activado, se aumenta la virulencia (hemólisis), pero se disminuye la formación de biofilm al reprimir la producción de proteínas de adhesión. Con estos antecedentes es plausible sugerir que si las proteínas *com_ylbF* interactúan con RNasa Y estabilizando la degradación del mRNA de SinR en *B. subtilis*, en *S. aureus* la proteína Qrp podría estar estabilizando la degradación de cualquiera de los dos RNA producidos por *agr* (RNAII o RNAIII) a través de su interacción con RNasa Y y de esta forma estar favoreciendo la formación del biofilm. Es decir, cuando la proteína Qrp no se produce (mutante) los RNA de *agr* permanecen más tiempo dentro de la bacteria, produciendo una disminución de la formación de biofilm y aumentando la hemólisis. A partir de este análisis es muy importante determinar si la delección del gen *qrp* afecta la actividad enzimática de la RNasa Y y si afecta los niveles de los RNA de *agr* en *S. aureus*.

Finalmente, el nombre de las proteínas YheA, YlbF y YmcA provienen de una nomenclatura general utilizada en la anotación automática de los genes producto de la secuenciación de la cepa 168 de *B. subtilis*. Para YlbF y YmcA se han propuesto dos nombres diferentes RcsA y RcsB [81]; o RicF y RicA [83], respectivamente. De acuerdo a los resultados encontrados en este estudio, relacionados con el alto contenido de glutamina, tanto en la proteína completa como en el motivo putativo **QQKQMXG**, como su probabilidad de participación en la función, se sugiere renombrar a la proteína YheA

como proteína rica en glutamina o Qrp (por su nombre en inglés Glutamine-rich protein) en *S. aureus*.

5. Conclusiones y recomendaciones

5.1 Conclusiones

Staphylococcus aureus es una de las bacterias que mayor impacto clínico genera debido a su capacidad de producir infecciones sistémicas. Aunque es una bacteria muy estudiada, aún no son claramente entendidos los mecanismos moleculares involucrados en su patogénesis. La capacidad de formar biofilm que posee esta bacteria contribuye a su supervivencia, persistencia y diseminación. Es importante comprender los mecanismos moleculares que intervienen en la formación del biofilm porque así se podrán diseñar nuevas estrategias para su inhibición.

Con la realización de este estudio se logró identificar una nueva familia de proteínas que contienen el dominio *com_ylbF* en *Staphylococcus aureus*, de las cuales una de ellas, Qrp/YheA, está participando en dos procesos fundamentales de la bacteria la formación de biofilm y la hemólisis de eritrocitos. En el primer caso, a través de una interacción, directa o indirecta, con una secuencia palindrómica inmersa dentro del gen *icaA*, y en el segundo caso, posiblemente por su participación en la degradación/estabilización de mRNA a través de la modulación de la actividad de la endoribonucleasa RNasa Y.

Las proteínas YlbF, YmcA y Qrp/YheA, que contienen el dominio *com_ylbF*, tienen una estructura tridimensional compuesta exclusivamente de α -hélices y en forma de “pinza” con la cual podrían interactuar más específicamente con sus sustratos (ácidos nucleicos). Notablemente, esta estructura es altamente conservada entre las tres proteínas, a pesar de su bajo porcentaje de identidad, y es muy probable que el dominio *com_ylbF* esté implicado en la función de estas proteínas.

Se identificó un posible motivo funcional dentro del dominio *com_ylbF*, el cual está conformado por 6 o 7 aminoácidos, tiene una secuencia específica en cada una de las tres proteínas (YlbF, YmcA y Qrp/YheA) y podría estar relacionado con la especificidad de la función que cada una de ellas puede estar realizando en las bacterias Gram positivas.

De acuerdo a la importancia que puede tener la glutamina en la función de la proteína, se sugiere renombrar a YheA como proteína rica en glutamina o Qrp (por su nombre en inglés Glutamine-rich protein) en *Staphylococcus aureus*.

5.2 Recomendaciones

La identificación de esta una nueva familia de proteínas en *Staphylococcus aureus* incentiva a continuar con su caracterización, no solo de la proteína Qrp, sino también de YlbF y YmcA (incluso de YaaT), para establecer sus funciones dentro de la bacteria. La evidencia obtenida hasta el momento en *Bacillus subtilis* y *Staphylococcus aureus* sugiere que estas proteínas *com_ylbF* están participando en procesos biológicos básicos de bacterias Gram positivas, así que su estudio nos permitirá avanzar en el conocimiento de los mecanismos moleculares involucrados en la formación de biofilm y virulencia. Es importante establecer experimentalmente su estructura tridimensional y su interacción con ácidos nucleicos, así como su participación en la formación del biofilm. Todo lo que se ha encontrado en estas proteínas será nuevo conocimiento y por lo cual altamente publicable en revistas especializadas de alto impacto.

Los resultados encontrados en este estudio fueron divulgados en el III congreso colombiano de bioquímica y biología molecular (C2B2) y en el artículo de investigación titulado “*Identification and “in silico” Structural Analysis of the Glutamine-rich Protein Qrp (YheA) in Staphylococcus aureus*” publicado en la revista The Open Bioinformatics Journal.

Anexo A: Protocolos de las cromatografías utilizadas en la purificación de las proteínas SarX y Qrp.

Protocolo utilizado en la Cromatografía de filtración en gel (CFG).

Buffer: Tris-HCl 20 mM pH 7.5

Paso	Volumen	Descripción			
1	0.00	Colectar fracciones de 0.50 mL durante todo el corrido			
2	0.00	Lámpara (detector UV)	Encendido		
3	0.00	Flujo isocrático	A:A1: ácido	100%	Vol: 2.00 mL
			B:B1: agua	0%	Flujo: 0.50%
4	2.00	Carga/inyección muestra	Muestra Loop estático	Auto- inyección	Vol: 2.00 mL Flujo: 0.50%
5	4.00	Cambio de válvula	Válvula AV 7-3		Carga de
6	4.00	Flujo isocrático	A:A1: ácido	100%	Vol: 36 mL
			B:B1: agua	0%	Flujo: 0.50 mL
	40.00	Fin del protocolo			

Protocolo utilizado en la cromatografía de intercambio aniónico (AEC) para las fracciones de GFC.

Buffer A: Tris-HCl 20mM pH 7.5

Buffer B: Tris-HCl 20mM pH 7.5 + NaCl 1M.

Paso	Volumen	Descripción			
1	0.00	Colectar fracciones de 0.50 mL durante todo el corrido			
2	0.00	Flujo isocrático	A:A1: ácido	100%	Vol: 3.00 mL
			B:B1: sal	0%	Flujo: 1.00 mL/min
3	3.00	Carga/inyección muestra	Muestra Loop estático	Auto-inyección	Vol: 2.00 mL
4	5.00	Cambio de válvula	Válvula AV 7-3		Flujo: 1.00 mL/min
5	5.00	Flujo isocrático	A:A1: ácido	100%	Vol: 2.00 mL
			B:B1: sal	0%	Flujo: 1.00 mL/min
6	7.00	Carga/inyección muestra	Muestra Loop estático	Auto-inyección	Vol: 2.00 mL
7	9.00	Cambio de válvula	Válvula AV 7-3		Flujo: 1.00 mL/min
8	9.00	Flujo isocrático	A:A1: ácido	100%	Vol: 1.00 mL
			B:B1: sal	0%	Flujo: 1.00 mL/min
9	10.00	Gradiente lineal	A:A1: ácido	100%----	Vol: 20 mL
			B:B1: sal	-0% 0%----- 100%	Flujo: 1 mL/min
10	30.00	Flujo isocrático	A:A1: ácido	0%	
			B:B1: sal	100%	
11	34.00	Fin del protocolo			

Anexo B: Péptidos utilizados para la identificación de la proteína Qrp.

La proteína Qrp en *Staphylococcus aureus* fue identificada en los dos estados de crecimiento (planctónico y biofilm) por medio de los siguientes péptidos:

Observed	Mr(expt)	Mr(calc)	Delta	Start	End	Miss	Ions	Peptide
778.4060	777.3987	777.4021	-0.0034	27	-	33	0 ---	K.EAFANVK.A
826.4023	825.3950	825.4021	-0.0071	41	-	46	0 ---	K.LFDEFR.E
954.4941	953.4868	953.4971	-0.0102	40	-	46	1 35	K.KLDEFR.E
1135.5625	1134.5552	1134.5669	-0.0117	47	-	55	0 ---	R.ETQINFQK.Q
1436.7258	1435.7185	1435.6943	0.0242	27	-	39	1 ---	K.EAFANVKANEESK.K
1647.7645	1646.7572	1646.7457	0.0115	56	-	69	0 ---	K.QMQGEEIAEEDLQK.A
1663.7371	1662.7298	1662.7406	-0.0108	56	-	69	0 ---	K.QMQGEEIAEEDLQK.A + Oxidation (M)
1880.9403	1879.9330	1879.9428	-0.0098	2	-	17	0 102	M.AVNLYDYANQLEQALR.E
2559.2568	2558.2495	2558.2282	0.0214	70	-	92	1 ---	K.AQEQAQAIKDNISALMNAEQK.M
2575.2351	2574.2278	2574.2231	0.0047	70	-	92	1 ---	K.AQEQAQAIKDNISALMNAEQK.M + Oxidation (M)

Anexo C: Secuencia de los genes *qrp* y *yheB* y sus proteínas codificantes.

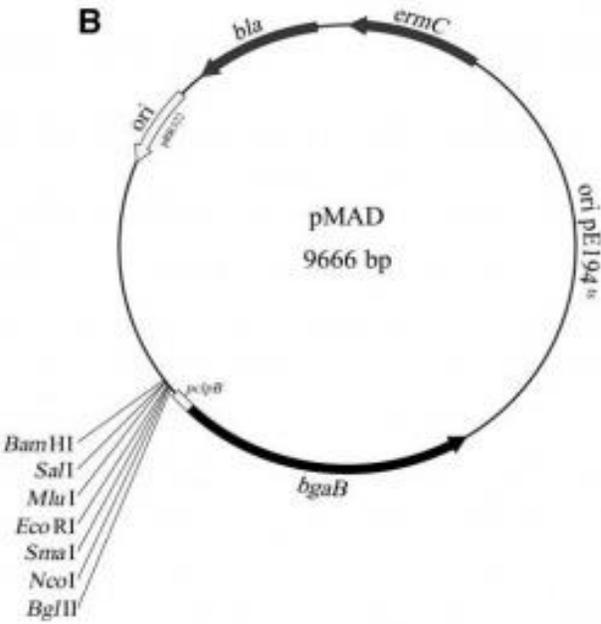
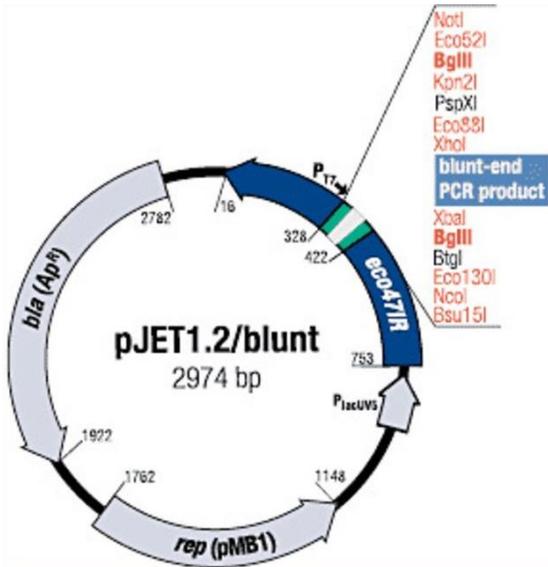
Secuencia de los genes *qrp* y *yheB* y las proteínas que codifican. En azul las posibles cajas Pribnow, en rojo las posibles secuencias de unión a ribosoma (Shine-Dalgarno) y en verde las posibles secuencias de terminación independientes de Rho identificadas por medio del programa Transterm se simbolizan por los loops en negro.

```

gcagaaagtgaatgattaatgaattaaataaaatagtggtataatgaataa+1ctatgtgtgcacaaggagtccaaaattatgaatgcactatttatt
M N A L F I
atcatatttatgattgtagtagggcgataaattggtggcattactaactgtaattgcaattagaatgctgtttccaccctttaaacatactatataatttaa
I I F M I V V G A I I G G I T N V I A I R M L F H P F K P Y Y I F K
atttagagtgccatttacaccagggttaataaccgaaaagacgcgaagaaattgcaactaaaattttaaggccaagtgatgaagagcatttgcttacagaa
F R V P P T P G L I P K R R E E I A T K I G Q V I E E H L L T E T L
actttaatgaaaaattgaaaagcgcgacaaatcacagcaagcaatagaatctatgattcaacagcagttacaaaagttgacgaaagatcaattgtcaataaaa
I N E K L K S E Q S Q Q A I E S M I Q Q Q L Q K L T K D Q L S I K Q
caaatctctccaaatcgatattgatttagaacaagtattacaactaatggaatcaatataatgaaatcaaatgaaataattattatacaaaagcatca
I T S Q I D I D L E Q V L Q T N G N Q Y I E S Q L N N Y Y T K H Q N
aaaccaacaatagcatctttattgcaaaatcaactgtaacatttttaatcagcatgtagataacgcaacagacttattatgtgatcgtgcaaggaatt
Q T I A S L L P N Q L V T F L N Q H V D N A T D L L C D R A R N Y L
atttactcatctgcaaaagcacaagatataatgatatggtggatctttttcaatgagaaaggttaagttaattggatggtgcaaatggttatgaca
S S A K G T Q D I N D M L D T F F N E K G K L I G M L Q M F M T K E
aaagagagtattgcagatcgcattcaacaagaacttatacgtttaacatctcatcctaaagcaagaacaattgtgacatcgttaattccaatgaatatca
S I A D R I Q Q E L I R L T S H P K A R T I V T S L I T N E Y Q T F
aacttttaagataaaaccattgaaatgaaactggttagatgcacatgcagtttaatgaaatcctatcagtgatgtaacaacatgatgcaagcaaac
K D K P L N E L L D A S Q F N E I A E N L S V Y V T T Y A S K Q A N
aagcgaataagccagtggttacgtttaatgcctaattggtgattattagaaggtcaactctctagcaaatgcaaatttaatcattgaaaaattatcg
K P V V T L M P Q F V D Y L E G Q L S S K L A N L I I E K L S I H L
atctacttatctacaattatgaaaaagtggtttacagaggatttaattgaaagacaaaattaatcogttagacttagattatgaaaaattaattataga
S T I M K K V D L R G L I E E Q I N T F D L D Y I E K L I I E I A N
aattgctaacaagaactaaagtgtgatgtcattaggtttcatttaggtggtattattggatttttccaaggttgggtgcaatcttgtctaaaaaa
K E L K L I M S L G F I L G G I I G F F Q G L V A I F V -
gcgagatggtgtaaagttaatacgcagcttttaataaacatcaacaaaaggagtgcataaacatggcagtaaattatgatgatgcaaatcaattag
M A V N L Y D Y A N Q L
aacaagctttaagagaaagcgaagaatacaaaagcaatcaaaagaagcattogctaaatgtaaaagctaaagaaatctaaaaagttattcgaagagttccogt
E Q A L R E S E E Y K A I K E A F A N V K A N E E S K K L F D E F
gaaactcaaattaacttccaacaaaaacaaatgcaaggtgaagaaattgctgaagaagatttacaaaaagcgcgaagaacaagcgaagcaattgaaaaaga
R E T Q I N F Q Q K Q M Q G E E I A E E D L Q K A Q E Q A Q A I E K
tgaaaacatctctgcattaatgaaatgctgaacaaaaatgagtcagatttccaagaaatcaaccaattactgtaaaccattagacgaaatttacgctg
D E N I S A L M N A E Q K M S Q V F Q E I N Q I I V K P L D E I Y A
actaattaaaaaataatttttcaatcctttaaatagaggaaatagactgcccagagagctcogagactttgtgggtagtttttttagtttgataaacgaa
D -

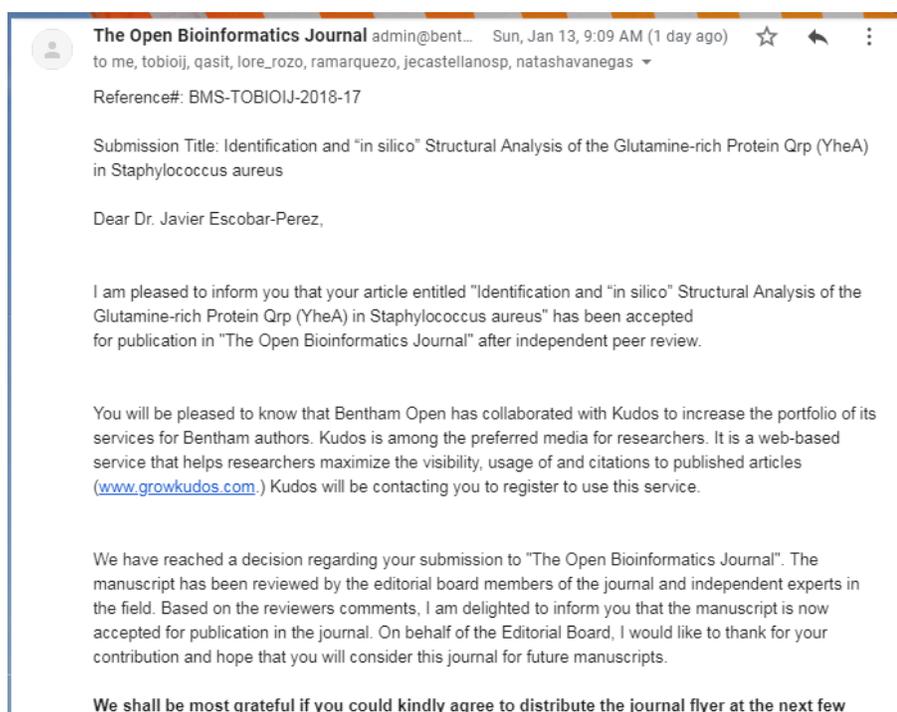
```

Anexo D: Diagrama de los plásmidos pJET y pMAD



Anexo E: soporte de la divulgación de los resultados

Aceptación del artículo de investigación titulado "*Identification and "in silico" Structural Analysis of the Glutamine-rich Protein Qrp (YheA) in Staphylococcus aureus*" publicado en la revista The Open Bioinformatics Journal.



Bibliografía

1. Balasubramanian D, Harper L, Shopsin B, Torres VJ. *Staphylococcus aureus* pathogenesis in diverse host environments. *Pathog Dis* 2017; 75.
2. Chambers HF. The changing epidemiology of *Staphylococcus aureus*? *Emerg Infect Dis* 2001; 7: 178-82.
3. Diekema D. J PMA, Schmitz F. J, Smayevsky J, Bell J, Jones R. N, and Beach M,. Survey of Infections Due to *Staphylococcus* Species: Frequency of Occurrence and Antimicrobial Susceptibility of Isolates Collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY Antimicrobial Surveillance Program, 1997–1999. *Clinical Infectious Diseases* 2001; 32: 114–32.
4. Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG, Jr. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* 2015; 28: 603-61.
5. Kluytmans J, van Belkum A, Verbrugh H. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev* 1997; 10: 505-20.
6. van Belkum A, Melles DC, Nouwen J *et al.* Co-evolutionary aspects of human colonisation and infection by *Staphylococcus aureus*. *Infect Genet Evol* 2009; 9: 32-47.
7. Chugh. *Staphylococcus aureus*: the modern Ghengis Khan. . *Journal of Infection in Developing Countries* 2007; 1 125-8.
8. Iwase T, Uehara Y, Shinji H *et al.* *Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* 2010; 465: 346-9.
9. Bannerman. *Staphylococcus*, *Micrococcus* and other Catalase – Positive Cocci That Grow Aerobically, . *Manual of Clinical Microbiology* 2003; Chap. 28. : 384-5. .
10. Giacometti A, Cirioni O, Schimizzi AM *et al.* Epidemiology and microbiology of surgical wound infections. *J Clin Microbiol* 2000; 38: 918-22.
11. Grupo para el control de la resistencia en Bogotá GREBO. Boletín Informativo N°4. 2012.
12. Escobar NM, Marquez IA, Quiroga JA *et al.* Using Positive Deviance in the prevention and control of MRSA infections in a Colombian hospital: a time-series analysis. *Epidemiol Infect* 2017; 145: 981-9.
13. Escobar JA, Marquez-Ortiz RA, Alvarez-Olmos MI *et al.* Detection of a new community genotype methicillin-resistant *Staphylococcus aureus* clone that is unrelated to the USA300 clone and that causes pediatric infections in Colombia. *J Clin Microbiol* 2013; 51: 661-4.

14. Portillo BC, Moreno JE, Yomayusa N *et al.* Molecular epidemiology and characterization of virulence genes of community-acquired and hospital-acquired methicillin-resistant *Staphylococcus aureus* isolates in Colombia. *Int J Infect Dis* 2013; 17: e744-9.
15. Escobar-Perez JA, Castro BE, Marquez-Ortiz RA *et al.* [Methicillin-sensitive *Staphylococcus aureus* isolates related to USA300 clone: Origin of community-genotype MRSA in Colombia?]. *Biomedica* 2014; 34 Suppl 1: 124-36.
16. Marquez-Ortiz RA, Alvarez-Olmos MI, Escobar Perez JA *et al.* USA300-related methicillin-resistant *Staphylococcus aureus* clone is the predominant cause of community and hospital MRSA infections in Colombian children. *Int J Infect Dis* 2014; 25: 88-93.
17. Escobar-Perez J, Reyes N, Marquez-Ortiz RA *et al.* Emergence and spread of a new community-genotype methicillin-resistant *Staphylococcus aureus* clone in Colombia. *BMC Infect Dis* 2017; 17: 108.
18. Archer NK, Mazaitis MJ, Costerton JW, Leid JG, Powers ME, Shirtliff ME. *Staphylococcus aureus* biofilms: properties, regulation, and roles in human disease. *Virulence* 2011; 2: 445-59.
19. Corne P, Marchandin H, Jonquet O, Campos J, Banuls AL. Molecular evidence that nasal carriage of *Staphylococcus aureus* plays a role in respiratory tract infections of critically ill patients. *J Clin Microbiol* 2005; 43: 3491-3.
20. Wertheim HF, Melles DC, Vos MC *et al.* The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect Dis* 2005; 5: 751-62.
21. Nuryastuti T, Krom BP, Aman AT, Busscher HJ, van der Mei HC. Ica-expression and gentamicin susceptibility of *Staphylococcus epidermidis* biofilm on orthopedic implant biomaterials. *J Biomed Mater Res A* 2011; 96: 365-71.
22. Singh R, Ray P, Das A, Sharma M. Penetration of antibiotics through *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms. *J Antimicrob Chemother* 2010; 65: 1955-8.
23. Arciola CR, Campoccia D, Speziale P, Montanaro L, Costerton JW. Biofilm formation in *Staphylococcus* implant infections. A review of molecular mechanisms and implications for biofilm-resistant materials. *Biomaterials* 2012; 33: 5967-82.
24. Houston P, Rowe SE, Pozzi C, Waters EM, O'Gara JP. Essential role for the major autolysin in the fibronectin-binding protein-mediated *Staphylococcus aureus* biofilm phenotype. *Infect Immun* 2011; 79: 1153-65.
25. Abee T, Kovacs AT, Kuipers OP, van der Veen S. Biofilm formation and dispersal in Gram-positive bacteria. *Curr Opin Biotechnol* 2011; 22: 172-9.
26. Waters EM, Rowe SE, O'Gara JP, Conlon BP. Convergence of *Staphylococcus aureus* Persister and Biofilm Research: Can Biofilms Be Defined as Communities of Adherent Persister Cells? *PLoS Pathog* 2016; 12: e1006012.

27. Cramton SE, Gerke C, Schnell NF, Nichols WW, Gotz F. The intercellular adhesion (*ica*) locus is present in *Staphylococcus aureus* and is required for biofilm formation. *Infect Immun* 1999; 67: 5427-33.
28. O'Neill E, Pozzi C, Houston P *et al.* A novel *Staphylococcus aureus* biofilm phenotype mediated by the fibronectin-binding proteins, FnBPA and FnBPB. *J Bacteriol* 2008; 190: 3835-50.
29. Heilmann C, Schweitzer O, Gerke C, Vanittanakom N, Mack D, Gotz F. Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. *Mol Microbiol* 1996; 20: 1083-91.
30. Gerke C, Kraft A, Sussmuth R, Schweitzer O, Gotz F. Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the *Staphylococcus epidermidis* polysaccharide intercellular adhesin. *J Biol Chem* 1998; 273: 18586-93.
31. Blevins JS, Beenken KE, Elasri MO, Hurlburt BK, Smeltzer MS. Strain-dependent differences in the regulatory roles of *sarA* and *agr* in *Staphylococcus aureus*. *Infect Immun* 2002; 70: 470-80.
32. Diep BA, Gill SR, Chang RF *et al.* Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* 2006; 367: 731-9.
33. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of persistent infections. *Science* 1999; 284: 1318-22.
34. Vergara-Irigaray M, Valle J, Merino N *et al.* Relevant role of fibronectin-binding proteins in *Staphylococcus aureus* biofilm-associated foreign-body infections. *Infect Immun* 2009; 77: 3978-91.
35. Lemon KP, Earl AM, Vlamakis HC, Aguilar C, Kolter R. Biofilm development with an emphasis on *Bacillus subtilis*. *Curr Top Microbiol Immunol* 2008; 322: 1-16.
36. O'Toole G, Kaplan HB, Kolter R. Biofilm formation as microbial development. *Annu Rev Microbiol* 2000; 54: 49-79.
37. Buttner H, Mack D, Rohde H. Structural basis of *Staphylococcus epidermidis* biofilm formation: mechanisms and molecular interactions. *Front Cell Infect Microbiol* 2015; 5: 14.
38. Mielich-Suss B, Lopez D. Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environ Microbiol* 2015; 17: 555-65.
39. Cairns LS, Hobley L, Stanley-Wall NR. Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. *Mol Microbiol* 2014; 93: 587-98.
40. Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat Rev Microbiol* 2013; 11: 157-68.
41. Otto M. *Staphylococcus epidermidis*--the 'accidental' pathogen. *Nat Rev Microbiol* 2009; 7: 555-67.

42. Foster TJ, Hook M. Surface protein adhesins of *Staphylococcus aureus*. Trends Microbiol 1998; 6: 484-8.
43. O'Gara JP. *ica* and beyond: biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*. FEMS Microbiol Lett 2007; 270: 179-88.
44. P MMTMJMB. Biología de los Microorganismos. Prentice-Hall. 2003. 10ª edición. ; Madrid. : 1096.
45. Rowe SE, Mahon V, Smith SG, O'Gara JP. A novel role for SarX in *Staphylococcus epidermidis* biofilm regulation. Microbiology 2011; 157: 1042-9.
46. Marti M, Trottonda MP, Tormo-Mas MA *et al*. Extracellular proteases inhibit protein-dependent biofilm formation in *Staphylococcus aureus*. Microbes Infect 2010; 12: 55-64.
47. Mack D, Fischer W, Krokotsch A *et al*. The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis. J Bacteriol 1996; 178: 175-83.
48. Jefferson KK, Pier DB, Goldmann DA, Pier GB. The teicoplanin-associated locus regulator (TcaR) and the intercellular adhesin locus regulator (IcaR) are transcriptional inhibitors of the *ica* locus in *Staphylococcus aureus*. J Bacteriol 2004; 186: 2449-56.
49. O'Gara JP, Humphreys H. *Staphylococcus epidermidis* biofilms: importance and implications. J Med Microbiol 2001; 50: 582-7.
50. Conlon KM, Humphreys H, O'Gara JP. *icaR* encodes a transcriptional repressor involved in environmental regulation of *ica* operon expression and biofilm formation in *Staphylococcus epidermidis*. J Bacteriol 2002; 184: 4400-8.
51. Ruiz de los Mozos I, Vergara-Irigaray M, Segura V *et al*. Base pairing interaction between 5'- and 3'-UTRs controls *icaR* mRNA translation in *Staphylococcus aureus*. PLoS Genet 2013; 9: e1004001.
52. Yu L, Hisatsune J, Hayashi I *et al*. A Novel Repressor of the *ica* Locus Discovered in Clinically Isolated Super-Biofilm-Elaborating *Staphylococcus aureus*. MBio 2017; 8.
53. Valle J, Toledo-Arana A, Berasain C *et al*. SarA and not sigmaB is essential for biofilm development by *Staphylococcus aureus*. Mol Microbiol 2003; 48: 1075-87.
54. George EA, Muir TW. Molecular mechanisms of agr quorum sensing in virulent staphylococci. Chembiochem 2007; 8: 847-55.
55. Ma R, Qiu S, Jiang Q *et al*. AI-2 quorum sensing negatively regulates *rbf* expression and biofilm formation in *Staphylococcus aureus*. Int J Med Microbiol 2017; 307: 257-67.
56. Feuillie C, Formosa-Dague C, Hays LM *et al*. Molecular interactions and inhibition of the staphylococcal biofilm-forming protein SdrC. Proc Natl Acad Sci U S A 2017; 114: 3738-43.

57. Cucarella C, Tormo MA, Knecht E *et al.* Expression of the biofilm-associated protein interferes with host protein receptors of *Staphylococcus aureus* and alters the infective process. *Infect Immun* 2002; 70: 3180-6.
58. Dziwanowska K, Patti JM, Deobald CF, Bayles KW, Trumble WR, Bohach GA. Fibronectin binding protein and host cell tyrosine kinase are required for internalization of *Staphylococcus aureus* by epithelial cells. *Infect Immun* 1999; 67: 4673-8.
59. Lasa I, Penades JR. Bap: a family of surface proteins involved in biofilm formation. *Res Microbiol* 2006; 157: 99-107.
60. Taglialegna A, Navarro S, Ventura S *et al.* Staphylococcal Bap Proteins Build Amyloid Scaffold Biofilm Matrices in Response to Environmental Signals. *PLoS Pathog* 2016; 12: e1005711.
61. Corrigan RM, Rigby D, Handley P, Foster TJ. The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. *Microbiology* 2007; 153: 2435-46.
62. Boles BR, Horswill AR. Agr-mediated dispersal of *Staphylococcus aureus* biofilms. *PLoS Pathog* 2008; 4: e1000052.
63. Kong KF, Vuong C, Otto M. Staphylococcus quorum sensing in biofilm formation and infection. *Int J Med Microbiol* 2006; 296: 133-9.
64. Klein PLMHJP. Microbiología. McGraw-Hill Interamericana 2004.; 5ª edición. : 1272.
65. Liu Q, Wang X, Qin J *et al.* The ATP-Dependent Protease ClpP Inhibits Biofilm Formation by Regulating Agr and Cell Wall Hydrolase Sle1 in *Staphylococcus aureus*. *Front Cell Infect Microbiol* 2017; 7: 181.
66. Spronk CA, Bonvin AM, Radha PK, Melacini G, Boelens R, Kaptein R. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure* 1999; 7: 1483-92.
67. Dattananda CS, Rajkumari K, Gowrishankar J. Multiple mechanisms contribute to osmotic inducibility of proU operon expression in *Escherichia coli*: demonstration of two osmoresponsive promoters and of a negative regulatory element within the first structural gene. *J Bacteriol* 1991; 173: 7481-90.
68. Mullin DA, Newton A. A sigma 54 promoter and downstream sequence elements ftr2 and ftr3 are required for regulated expression of divergent transcription units flaN and flbG in *Caulobacter crescentus*. *J Bacteriol* 1993; 175: 2067-76.
69. Mitra R, Das HK, Dixit A. Identification of a positive transcription regulatory element within the coding region of the nifLA operon in *Azotobacter vinelandii*. *Appl Environ Microbiol* 2005; 71: 3716-24.
70. Resch A, Leicht S, Saric M *et al.* Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics* 2006; 6: 1867-77.
71. Edelman. B. Protein Methods. Wiley-Liss. New York. 1993: 96-127.

72. Cue D, Lei MG, Lee CY. Activation of *sarX* by *Rbf* is required for biofilm formation and *icaADBC* expression in *Staphylococcus aureus*. *J Bacteriol* 2013; 195: 1515-24.
73. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 2007; 8: R22.
74. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics* 2014; 48: 3 13 1-6.
75. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008; 9: 40.
76. Pettersen EF, Goddard TD, Huang CC *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004; 25: 1605-12.
77. Cheung AL, Nishina KA, Trotonda MP, Tamber S. The *SarA* protein family of *Staphylococcus aureus*. *Int J Biochem Cell Biol* 2008; 40: 355-61.
78. Tortosa P, Albano M, Dubnau D. Characterization of *ylbF*, a new gene involved in competence development and sporulation in *Bacillus subtilis*. *Mol Microbiol* 2000; 35: 1110-9.
79. Carabetta VJ, Tanner AW, Greco TM, Defrancesco M, Cristea IM, Dubnau D. A complex of *YlbF*, *YmcA* and *YaaT* regulates sporulation, competence and biofilm formation by accelerating the phosphorylation of *Spo0A*. *Mol Microbiol* 2013; 88: 283-300.
80. Branda SS, Gonzalez-Pastor JE, Dervyn E, Ehrlich SD, Losick R, Kolter R. Genes involved in formation of structured multicellular communities by *Bacillus subtilis*. *J Bacteriol* 2004; 186: 3970-9.
81. DeLoughery A, Dengler V, Chai Y, Losick R. Biofilm formation by *Bacillus subtilis* requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor *SinR*. *Mol Microbiol* 2016; 99: 425-37.
82. Dubnau EJ, Carabetta VJ, Tanner AW, Miras M, Diethmaier C, Dubnau D. A protein complex supports the production of *Spo0A-P* and plays additional roles for biofilms and the K-state in *Bacillus subtilis*. *Mol Microbiol* 2016; 101: 606-24.
83. Tanner AW, Carabetta VJ, Martinie RJ *et al*. The *RicAFT* (*YmcA-YlbF-YaaT*) complex carries two [4Fe-4S]²⁺ clusters and may respond to redox changes. *Mol Microbiol* 2017; 104: 837-50.
84. Grossman AD. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu Rev Genet* 1995; 29: 477-508.
85. Tucker AT, Bobay BG, Banse AV *et al*. A DNA mimic: the structure and mechanism of action for the anti-repressor protein *AbbA*. *J Mol Biol* 2014; 426: 1911-24.
86. Bai U, Mandic-Mulec I, Smith I. *SinI* modulates the activity of *SinR*, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction. *Genes Dev* 1993; 7: 139-48.

87. Kobayashi K. SlrR/SlrA controls the initiation of biofilm formation in *Bacillus subtilis*. *Mol Microbiol* 2008; 69: 1399-410.
88. Berka RM, Hahn J, Albano M *et al.* Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol Microbiol* 2002; 43: 1331-45.
89. Hamoen LW, Smits WK, de Jong A, Holsappel S, Kuipers OP. Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach. *Nucleic Acids Res* 2002; 30: 5517-28.
90. Ogura M, Yamaguchi H, Kobayashi K, Ogasawara N, Fujita Y, Tanaka T. Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J Bacteriol* 2002; 184: 2344-51.
91. Turgay K, Hamoen LW, Venema G, Dubnau D. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev* 1997; 11: 119-28.
92. DeLoughery A, Lalanne JB, Losick R, Li GW. Maturation of polycistronic mRNAs by the endoribonuclease RNase Y and its associated Y-complex in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 2018; 115: E5585-E94.
93. Cho KH. The Structure and Function of the Gram-Positive Bacterial RNA Degradosome. *Front Microbiol* 2017; 8: 154.
94. Subramaniam AR, Deloughery A, Bradshaw N *et al.* A serine sensor for multicellularity in a bacterium. *Elife* 2013; 2: e01501.
95. Chai Y, Norman T, Kolter R, Losick R. Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in *Bacillus subtilis*. *EMBO J* 2011; 30: 1402-13.
96. Arnaud M, Chastanet A, Debarbouille M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microbiol* 2004; 70: 6887-91.
97. Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Novick RP. Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl Environ Microbiol* 2004; 70: 6076-85.
98. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* 2010; 48: 463-5.
99. Gemayel R, Chavali S, Pougach K *et al.* Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol Cell* 2015; 59: 615-27.
100. Mills CL, Beuning PJ, Ondrechen MJ. Biochemical functional predictions for protein structures of unknown or uncertain function. *Comput Struct Biotechnol J* 2015; 13: 182-91.
101. Kaito C, Kurokawa K, Matsumoto Y *et al.* Silkworm pathogenic bacteria infection model for identification of novel virulence genes. *Mol Microbiol* 2005; 56: 934-44.

-
102. Nagata M, Kaito C, Sekimizu K. Phosphodiesterase activity of CvfA is required for virulence in *Staphylococcus aureus*. J Biol Chem 2008; 283: 2176-84.
 103. Marincola G, Schafer T, Behler J *et al.* RNase Y of *Staphylococcus aureus* and its role in the activation of virulence genes. Mol Microbiol 2012; 85: 817-32.
 104. Numata S, Nagata M, Mao H, Sekimizu K, Kaito C. CvfA protein and polynucleotide phosphorylase act in an opposing manner to regulate *Staphylococcus aureus* virulence. J Biol Chem 2014; 289: 8420-31.
 105. Bonnin RA, Bouloc P. RNA Degradation in *Staphylococcus aureus*: Diversity of Ribonucleases and Their Impact. Int J Genomics 2015; 2015: 395753.
 106. Khemici V, Prados J, Linder P, Redder P. Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. PLoS Genet 2015; 11: e1005577.
 107. Marincola G, Wolz C. Downstream element determines RNase Y cleavage of the saePQRS operon in *Staphylococcus aureus*. Nucleic Acids Res 2017; 45: 5980-94.