



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Maestría y Doctorado en Ciencias Bioquímicas
INSTITUTO DE BIOTECNOLOGÍA

“ANÁLISIS DE LA CORRELACIÓN DE CARACTERÍSTICAS FISICOQUÍMICAS,
MECANISMOS DE EXPORTACIÓN EXTRACELULAR Y TIPO DE ENVOLTURA
CELULAR DE LAS ESPECIES DE ORIGEN, DE EXPANSINAS MICROBIANAS”

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. Andrés de Sandozequi Mijares

TUTOR PRINCIPAL

Dra. Claudia Martínez Anaya
[Instituto de Biotecnología](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. José Luis Puente García
[Instituto de Biotecnología](#)

Dr. Mario Alberto Serrano Ortega
[Centro de Ciencias Genómicas](#)

Cuernavaca, Morelos. Agosto, 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A Roseta, por su insaciable forma de quererme,

a Mario, por darme una mente crítica.

A Martín, mi maestro de vida,

a Cinthia, por todo el amor y apoyo en estos años y en los que siguen.

Agradecimientos

A Juan Salazar, que sin su increíble ingenio tal vez nunca hubiéramos realizado este proyecto.

A Claudia, por su guía y enseñanzas que han marcado mi camino. Este logro es un reflejo de su invaluable influencia en mi vida.

A Chris Wood, por su dedicación a corregir las sandeces que escribo en inglés.

¡A todos en el laboratorio de la Dra. Claudia Martínez! ¡Gracias por aguantarme! Y saben que siempre tendrán mi apoyo cuando lo necesiten.

A Irán y a Carla, ojalá podamos continuar esta colaboración, que estoy seguro dará mucho frutos.

A mis padres, mi hermano y mi familia que me apoyó a lo largo de mi vida, para poder culminar mi formación académica.

A Cinthia, que nunca terminaré de agradecerle su más sincero amor, apoyo, paciencia, y por hacerme sonreír aún en los momentos complicados.

Y finalmente,

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico 735828 otorgado para la realización de mis estudios de Doctorado.

Agradezco al programa DGAPA-PAPIIT IN203622 y PAPIIT IN211019 por el financiamiento al presente proyecto.

Tabla de contenido

Resumen.....	1
Abstract	2
Capítulo I. Introducción.....	3
I.1 Las Expansinas en plantas	3
I.2 Expansinas en bacterias.	4
I.3 Características fisicoquímicas, estructurales y evolutivas de las expansinas en bacterias.....	5
I.4 Asociación de proteínas a la envoltura celular	9
I.4.1 Lipoproteínas.....	10
I.4.2 Anclaje a la pared de peptidoglicano mediada por sortasas	15
I.4.3 Sistema de secreción tipo IX	17
I.8 Antecedentes	19
Capítulo II. Justificación.....	22
II.1 Justificación	22
II.1 Hipótesis.....	22
II.3 Objetivo general.....	22
II.3.1 Objetivos particulares	22
Capítulo III. Métodos.....	24
III.1 Obtención y alineamiento de secuencias.....	24
III.2 Búsqueda de secuencias	24
III.4 Cálculo del punto isoeléctrico	25
III.5 Análisis de los péptidos señal.....	25
III.6 Búsqueda de motivos LPXTG-like	25
III.7 Predicción de dominios adicionales	26
III.8 Relaciones filogenéticas	26
III.10 Análisis de regiones desordenadas	27
III.11 Predicción de estructuras de proteínas	27
Capítulo IV. Resultados	28
IV.1 Expansinas en Proteobacterias, Cianobacterias, Fibrobacteres y Bacteroidetes (bacterias didermas)	33

IV.2 Expansinas en Firmicutes	36
IV.3 Expansinas en Actinobacterias.....	39
IV.4 Variedad de arquitecturas en las expansinas multimodulares.....	43
Capítulo V. Discusión, conclusiones y perspectivas.	46
V.1 Discusión	46
V.2 Conclusión.....	48
V.3 Perspectivas	49
Capítulo VI. Referencias	53
Capítulo VII. Anexos	63

Resumen

Las expansinas son un grupo de proteínas presentes en diferentes organismos, principalmente en bacterias, hongos y plantas. Aunque las expansinas presentan una conservación estructural, sus roles biológicos parecen diferir entre cada taxón. En las plantas, estas proteínas remodelan la pared celular durante el crecimiento y otros procesos. Por el contrario, la determinación de la función biológica de las expansinas bacterianas ha resultado difícil, aunque la evidencia genética de mutantes bacterianos fitopatógenos indica que las expansinas participan en las interacciones entre bacterias y plantas. Sin embargo, una gran proporción de genes de expansina se encuentran en los genomas de bacterias de vida libre, lo que sugiere roles que son independientes de la interacción con las plantas vivas. En este trabajo, analizamos todas las secuencias disponibles de expansinas procariotas en busca de correlaciones entre la carga eléctrica superficial, módulos de proteína adicionales y motivos de secuencia para la asociación con el exterior de las bacterias después de la exportación. Además, la información sobre el destino de la proteína después de la translocación a través de la membrana apunta a la existencia de una asociación de las expansinas al exterior celular bacteriano a través de seis mecanismos diferentes, tales como la fijación de una molécula lipídica para el anclaje a la membrana en especies didermas o la unión covalente a la capa de peptidoglicano en monodermas (como Bacilliales). Nuestros resultados tienen implicaciones para la clasificación de las funciones de las expansinas en el contexto de las interacciones entre bacterias y plantas y también para especies de vida libre en las que las expansinas podrían afectar las propiedades de interacción célula-célula o célula-sustrato y señalan la necesidad de reexaminar los roles actualmente considerados para estas proteínas.

Palabras clave: *Expansina, proteínas superficiales, envoltura celular de bacterias, bacterias fitopatógenas, interacción bacteria-hospedero, evolución de proteínas, análisis de secuencias.*

Abstract

Expansins are a group of proteins from diverse organisms from bacteria to plants. Although expansins show structural conservation, their biological roles seem to differ among taxa. In plants, these proteins remodel the cell wall during plant growth and other processes. Contrarily, determination of bacterial expansin activity has proven difficult, although genetic evidence of bacterial mutants indicates that expansins participate in bacteria–plant interactions. Nevertheless, a large proportion of expansin genes are found in the genomes of free-living bacteria, suggesting roles that are independent of the interaction with living plants. Here, we analyzed all available sequences of prokaryotic expansins for correlations between surface electric charge, extra protein modules, and sequence motifs for association with the bacteria exterior after export. Additionally, information on the fate of protein after translocation across the membrane also points to bacterial cell association of expansins through six different mechanisms, such as attachment of a lipid molecule for membrane anchoring in diderm species or covalent linking to the peptidoglycan layer in monoderms such as the Bacilliales. Our results have implications for expansin function in the context of bacteria–plant interactions and for free-living species in which expansins might affect cell–cell or cell–substrate interaction properties and indicate the need to re-examine the roles currently considered for these proteins.

Keywords: *Expansin, surface proteins, bacterial cell envelope, phytopathogenic bacteria, bacteria-host interaction, protein evolution, sequence analysis.*

Capítulo I. Introducción

I.1 Las Expansinas en plantas

Las expansinas han sido ampliamente estudiadas en plantas, donde se descubrieron inicialmente. Estas proteínas participan en el aflojamiento y en la remodelación de la pared celular vegetal, lo cual es importante para el crecimiento celular. Inducen la extensión irreversible de las paredes celulares, permitiendo que la célula se expanda sin romperse (Cosgrove, 2016). Existen dos familias principales de expansinas en plantas: las alfa-expansinas (EXPA) y las beta-expansinas (EXPB), ambas presentes en todas las plantas terrestres (Cosgrove, 2015; Sampedro and Cosgrove, 2005). Las EXPA se encuentran típicamente en el apoplasto (el espacio extracelular fuera de la membrana plasmática) y participan en el aflojamiento y extensión de la pared celular durante el crecimiento y desarrollo celular (Chebli and Geitmann, 2017). Las EXPB se encuentran en la pared celular y participan en la remodelación y aflojamiento de las microfibrillas de celulosa, que son el componente principal de la pared celular (T., Wang et al., 2016). En particular, se ha descubierto que las EXPB participan en el crecimiento del tubo polínico, un proceso imperativo para la fecundación en plantas con flores (Cosgrove, 2017a; Georgelis et al., 2015; Valdivia et al., 2009). También, existen en las plantas otras dos sub-familias relacionadas a las expansinas A y B (por sus siglas en inglés “Expansin-Like proteins”, EXLA y EXLB), aunque su función biológica exacta es aún desconocida. Trabajos recientes de transcriptómica muestran su sobreexpresión durante el estrés abiótico (Bashline et al., 2014; Jin et al., 2020; Shin et al., 2005). Por otra parte, se ha encontrado una gran variedad de microorganismos que poseen genes de expansinas, y aunque su divergencia evolutiva es considerable, se agruparon en la familia X de expansinas EXLX (Cosgrove, 2017b; Georgelis et al., 2011; Sampedro and Cosgrove, 2005). Estas están presentes en bacterias, hongos y oomycetos. La familia EXLX ha sido

blanco de extensos estudios tratando de elucidar su función biológica, su importancia en la patogénesis de plantas, e incluso, su posible uso en diversos procesos industriales (Cosgrove, 2017b; Georgelis et al., 2011; Georgelis et al., 2014; Lohoff et al., 2020).

I.2 Expansinas en bacterias.

A diferencia de las expansinas de las plantas, el papel de las expansinas bacterianas sigue siendo desconocido. La función biológica de las expansinas bacterianas solo se ha inferido a partir de experimentos genéticos que muestran su participación en la colonización de las plantas por fitopatógenos o bacterias que promueven el crecimiento de las plantas (Junior et al., 2015; Narváez-Barragán et al., 2020; Nikolaidis et al., 2014a; Olarte-Lozano et al., 2014; Rocha et al., 2020; Tancos et al., 2018; Tovar-Herrera et al., 2018a). Diferentes enfoques para determinar la función y los mecanismos moleculares de las expansinas bacterianas han producido resultados marginales (Georgelis et al., 2014; Lin et al., 2013; Liu et al., 2014; Q., Wang et al., 2016). Por ejemplo, en ensayos de extensión de tejidos vegetales, las expansinas bacterianas tienen un rendimiento muy por debajo de la eficiencia de sus homólogos de plantas (Georgelis et al., 2011; Q., Wang et al., 2016); mientras que, en los ensayos de debilitamiento del papel de filtro, las expansinas bacterianas funcionan solo ligeramente mejor en comparación con los ensayos en tejido vegetales (Georgelis et al., 2014). Por otro lado, debido al interés de la industria textil, papelera y energética en enzimas y proteínas capaces de mejorar los rendimientos en el procesamiento de biopolímeros, existen ensayos que contemplan la posibilidad de que exista una sinergia de las expansinas bacterianas con diferentes enzimas hidrolíticas. Una de las más utilizadas son las celulasas (Endoglucanasas, exoglucanasas, glicosil hidrolasas, etc.), con las que se plantea la hipótesis de que la posible modificación a la pared celular de la planta por parte de las expansinas facilita el acceso de las enzimas líticas a su sustrato (Artzi et al., 2015; Bunternngsook et al., 2014;

Bunterngsook et al., 2015; Chen et al., 2016; Georgelis et al., 2014; Junior et al., 2015; Kim et al., 2009; Liu et al., 2014; Pech-Cervantes et al., 2019; Qin et al., 2013; Q., Wang et al., 2016; Wang et al., 2014). Sin embargo, los resultados son controversiales y a menudo inconclusos.

I.3 Características fisicoquímicas, estructurales y evolutivas de las expansinas en bacterias.

Las expansinas son una familia de proteínas globulares, no hidrolíticas y con una alta densidad de carga eléctrica en su superficie. Una expansina típica, *Bacillus subtilis* (BsEXLX1), que fue la primera en ser descrita en bacterias (Kerff et al., 2008), consiste en alrededor de 230 aminoácidos, comprendiendo dos dominios conservados (dominio N-terminal (D1) y terminal de las expansinas (D2)) y un péptido señal que determina su translocación a través de la membrana hacia el exterior de la célula. El dominio D1 tiene homología con los dominios glicosil hidrolasa de la familia 45 (GH45), aunque ha perdido su actividad hidrolítica, por lo que usualmente se usa el término de “pseudoactividad” a los efectos de las expansinas sobre los sustratos (Georgelis et al., 2011; Martínez-Anaya, 2016a). Así, se piensa que los D1 de las expansinas microbianas realizan su función a partir de interacciones electrostáticas entre la proteína y su sustrato. El aspartato en la posición 82 (residuo Asp82) en el D1 de BsEXLX1 (Figura 1), es considerado como el uno de lo más relevantes para la pseudoactividad de las expansinas bacterianas, y está conservado en todas las expansinas bacterianas que han sido analizadas bioquímicamente (Cosgrove, 2017b). Asimismo, en análisis donde se sustituye el Asp82 por una alanina, las expansinas mutantes pierden completamente la capacidad de extender tejidos vegetales y de debilitar papel filtro (Georgelis et al., 2014; Martínez-Anaya, 2016a).

El dominio D2 pertenece a la familia de módulos de unión a carbohidratos 63 (CBM63) y es el responsable de la unión y el alineamiento del sustrato (Georgelis et al., 2011; Lohoff et al., 2020). El D2 contiene tres residuos aromáticos (Trp125,

Trp126, Tyr157) responsables de la unión al sustrato, conformando en conjunto con el D1 una región plana, llamada superficie de unión al polisacárido (Figura 1) (Nikolaidis et al., 2014b). En esta tesis, para fines prácticos, se mencionará como un solo dominio funcional al conjunto de los dos subdominios conservados de la expansina.

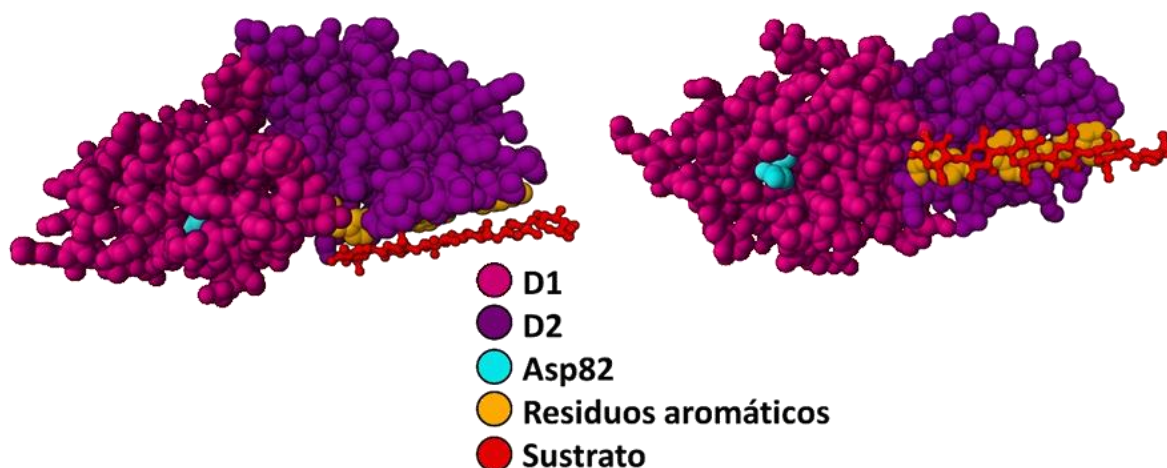


Figura 1. Estructura de BsEXLX1 con su sustrato (celohexaosa). Los dominios D1 y D2 se alinean para formar una superficie de unión al polisacárido. Los tres residuos aromáticos Trp125, Trp126, Tyr157 (Amarillo) son los responsables de unir al sustrato (Rojo, celohexaosa) y alinearlo. El Asp82 (Azul claro) es el residuo más importante para que la expansina lleve a cabo la acción de debilitamiento de los enlaces débiles entre cadenas de polisacáridos, como la celulosa, permitiendo la extensión de paredes celulares vegetales. Estructura basada en cristalografía de rayos X (PDB id: 4FER, Kerff et al., 2008), modelada con Mol*.

En estudios recientes, se ha propuesto que las expansinas en bacterias tienen su origen a partir de uno o más eventos de transferencia horizontal de genes, donde las plantas fueron las donadoras de material genético (Armijos-Jaramillo et al., 2018; Chase et al., 2020a; Nikolaidis et al., 2014b; Rocha et al., 2020). Lo que es congruente con que, las expansinas, aunque con una alta divergencia a nivel de secuencia, estén conservadas estructuralmente en todos los clados donde están presentes, además de que existen residuos estructurales y funcionales altamente

conservados. También, explica la razón por la que especies de filo bacterianos sin ancestría reciente (Figura 2), presentan genes de expansina con homología directa y similitudes de secuencia mayores al 90% en algunos casos (Chase et al., 2020a).

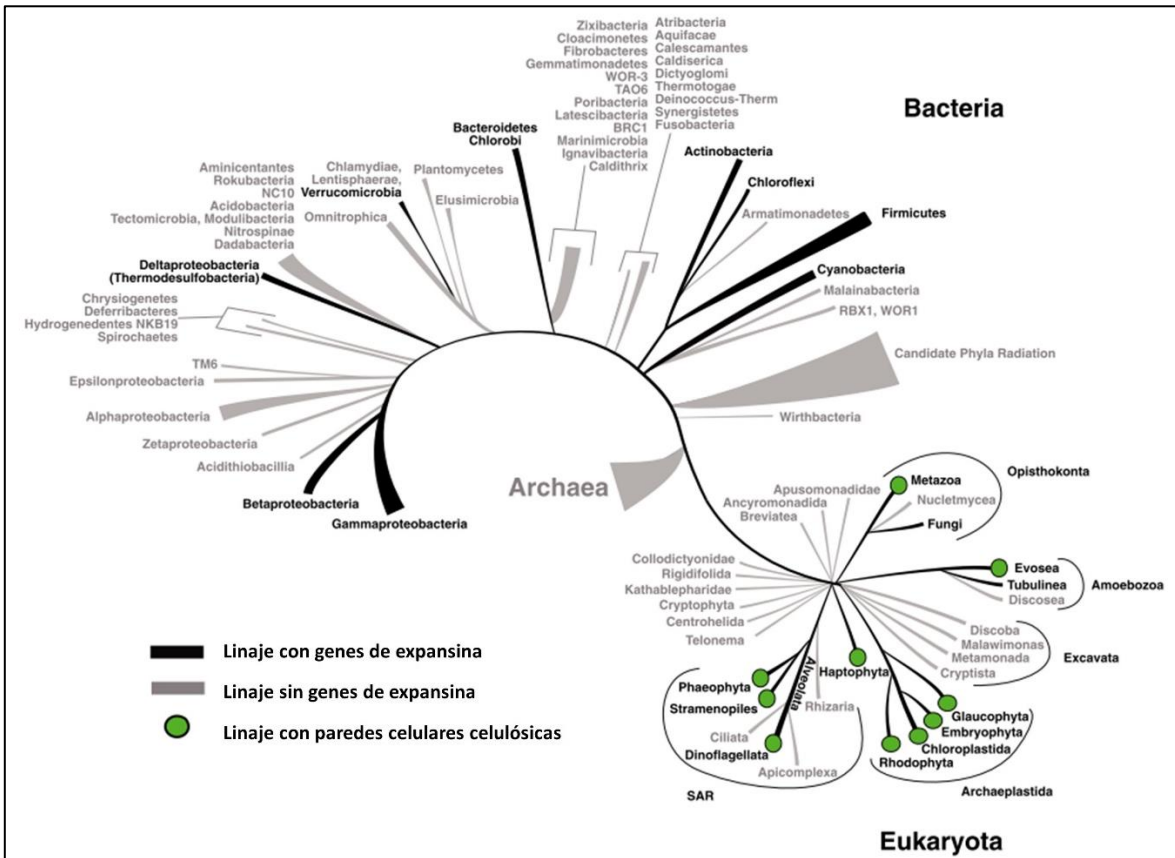


Figura 2. Relaciones filogenéticas de clados donde existen especies que presentan genes de expansinas. Las proteobacterias, los firmicutes y las actinobacterias son los filo con mayor representación de especies bacterianas con expansinas (Tomado de Chase et al., 2020).

En estudios de búsqueda de secuencias de expansinas en bacterias, se ha encontrado una gran diversidad de especies que las contienen (Chase et al., 2020a; Lohoff et al., 2020). Aunque las expansinas bacterianas que han sido estudiadas y caracterizadas bioquímicamente son relativamente pocas, en comparación con las encontradas en bases de datos de secuencias. Tomando en cuenta la diversidad de las expansinas y del estilo de vida de las bacterias que las

presentan, se ha llegado a la hipótesis de que estas proteínas son importantes para bacterias que tienen contacto constante con tejidos vegetales, ya sea, que presenten una asociación (tanto patogénica como de promoción del crecimiento) con plantas vivas, con la biomasa vegetal del suelo o sedimentos, o como parte de microbiomas intestinales de animales con dietas herbívoras y omnívoras (Chase et al., 2020b).

Aún después de casi dos décadas de investigación sobre las expansinas bacterianas, su mecanismo de acción y su función biológica sigue eludiéndonos, ya que, al ser proteínas no hidrolíticas, su estudio se complica (Martínez-Anaya, 2016a). Además, debido a la complejidad de la pared celular vegetal (Silveira and Skaf, 2016), no se ha podido caracterizar el sustrato específico en estudios *in-vivo* (Narváez-Barragán et al., 2020). Adicionalmente, la mayoría de los estudios en donde se intenta caracterizar la función de las expansinas se han utilizado proteínas recombinantes, producidas heterológicamente en ensayos hasta cierto punto artificiales, donde se observa indirectamente el efecto provocado por las expansinas. Un ejemplo es el ensayo de *creep* (palabra inglesa sin equivalente en el español, que se refiere a la capacidad de un tejido de expandirse de allí el nombre de 'expansinas'). En este caso, usando pinzas se aplica una tensión constante a una muestra de tejido vegetal incubado en un buffer con pH bajo. La velocidad de extensión aumenta inmediatamente después de la adición de la expansina (Tabuchi et al., 2011). Este método se considera el más adecuado para clasificar a una proteína como expansina; sin embargo, el dispositivo para cuantificar el efecto no se encuentra comercialmente disponible, y no es sencillo de implementar en un laboratorio ordinario de bioquímica. Otro ensayo ampliamente utilizado por su simplicidad y reproducibilidad es el debilitamiento de papel filtro inducido por las expansinas (Georgelis et al., 2011), que también ha sido utilizado para observar la posibilidad de incrementar la eficiencia de enzimas fibrolíticas, generalmente utilizando enzimas comerciales, como por ejemplo la endoglucanasa de *Trichoderma reesei* (Pech-Cervantes et al., 2019).

I.4 Asociación de proteínas a la envoltura celular

La envoltura celular de las bacterias está compuesta por una variedad de moléculas, como polisacáridos, lípidos y proteínas (Bogino et al., 2013; Silhavy et al., 2010; Viljoen et al., 2020). Las proteínas asociadas a la pared celular tienen numerosas funciones en la protección, secreción, formación de biopelículas, captación de nutrientes, metabolismo y virulencia (Cole et al., 2021; Kovacs-Simon et al., 2011; Narita and Tokuda, 2017). Las bacterias utilizan diferentes mecanismos para asociar las proteínas a su envoltura celular, que se pueden clasificar en tres grandes grupos: modificaciones post-traduccionales (Konovalova et al., 2017; Veith et al., 2017); unión covalente al peptidoglicano por parte de enzimas sortasas (Dramsi et al., 2008); y por las características intrínsecas de ciertos dominios proteicos (Konovalova et al., 2017; Wilkens, 2015).

En primer lugar, las bacterias tienen sofisticados mecanismos para translocar proteínas extracelulares a través de la envoltura celular, los cuales se encuentran tanto en bacterias didermas y como en monodermas (Frain et al., 2019; Goosens et al., 2014; Lee et al., 2006). La translocación de proteínas es esencial para todos los organismos y es especialmente crucial para los patógenos, ya que secretan factores de virulencia indispensables para la infección del huésped (Jackson et al., 2015). La exportación de proteínas en bacterias ha sido estudiada exhaustivamente (Chatzi et al., 2014), identificándose nueve sistemas de secreción (Tipos I a IX), dos sistemas de translocación a través de la membrana interna (Sec y Tat), el complejo de *sorting* de proteínas en bacterias monodermas, y un inyectosoma (Costa et al., 2015; Green and Mecsas, 2016). Las proteínas secretadas por las bacterias pueden tener tres destinos principales dictados por señales de clasificación contenidas en su secuencia: (a) pueden ser dirigidas al medio extracelular como proteínas solubles o como parte de la matriz extracelular; (b) pueden permanecer ancladas a la superficie celular en la membrana como lipoproteínas o unidas covalentemente a la capa de peptidoglicano por sortasas; o (c) pueden convertirse en parte de superestructuras celulares como los flagelos, la

matriz extracelular, o el celulosoma (De Diego et al., 2016; Green and Mecsas, 2016; Malik and Kim, 2019; Tseng et al., 2009). En las bacterias fitopatógenas, las proteínas expuestas en la superficie frecuentemente cumplen funciones fisiológicas e invasivas (Cole et al., 2021), y diferentes estudios muestran la implicación de la superficie bacteriana en el proceso de patogénesis (Bogino et al., 2013; Desvaux et al., 2006; Desvaux et al., 2018; Pfeilmeier et al., 2016; Rodríguez-Ortega, 2018). Aunque la literatura principalmente se centra en patógenos humanos, hay un interés creciente en cómo la superficie bacteriana está involucrada en el proceso de infección por bacterias fitopatógenas. En el caso de las expansinas, poco se sabe de su destino biológico una vez secretada. Comúnmente se piensa que las expansinas actúan en los tejidos vegetales de las plantas hospedadoras, pero esto no tendría sentido para bacterias de vida libre. Por lo que investigar los posibles destinos de estas proteínas es relevante para entender su función biológica. Estas señales nos ayudarán a entender como las bacterias procesan y dirigen a la expansinas.

I.4.1 Lipoproteínas

Las bacterias han desarrollado diversos métodos para asociar proteínas a su envoltura celular. Dentro de la secuencia de las proteínas, pueden existir dominios, regiones o motivos conservados que señalizan su procesamiento y su destino celular. Estas señales son reconocidas por enzimas, poros membranales y complejos proteicos para procesar y dirigir a las proteínas, por ejemplo, para su exportación a través de un sistema de secreción específico. Uno de los mecanismos más utilizados por las bacterias son los llamados péptidos señal, que son motivos conservados de alrededor de 20 a 30 aminoácidos (Figura 3A), localizados en el extremo N-terminal de las proteínas (Cole et al., 2021; Von Heijne, 1989).

Presentes tanto en bacterias monodermas como en didermas, los péptidos señal son reconocidos por alguno de los sistemas de translocación a través de la

membrana citoplásmica (Sec y Tat). Estos sistemas están encargados de externalizar a las proteínas y escindir al péptido señal para dar lugar a la proteína madura, y en algunos casos, encaminar a las proteínas blanco a otros sistemas de secreción (Dufresne and Paradis-Bleau, 2015). Existen tres principales tipos de péptidos señal en las bacterias (Almagro Armenteros et al., 2019), que indican el sistema de translocación que los reconoce, y señalar si las proteínas serán encaminadas a ser modificadas postraduccionalmente para convertirse en lipoproteínas (Figura 3B), la cuales son proteínas a las que se les agrega un lípido (triacil-glicerol) capaz de anclarse en la membrana. Dependiendo de la secuencia del péptido señal, la proteína será translocada por el sistema Sec si tiene el péptido señal SP (Sec/SPI) o por el sistema Tat si tiene el péptido señal TAT (Tat/SPI), que también señala si el péptido será cortado por la peptidasa de señales I o por la II (Frain et al., 2019; Hooda and Moraes, 2018; Lee et al., 2006; Van Wely et al., 2001).

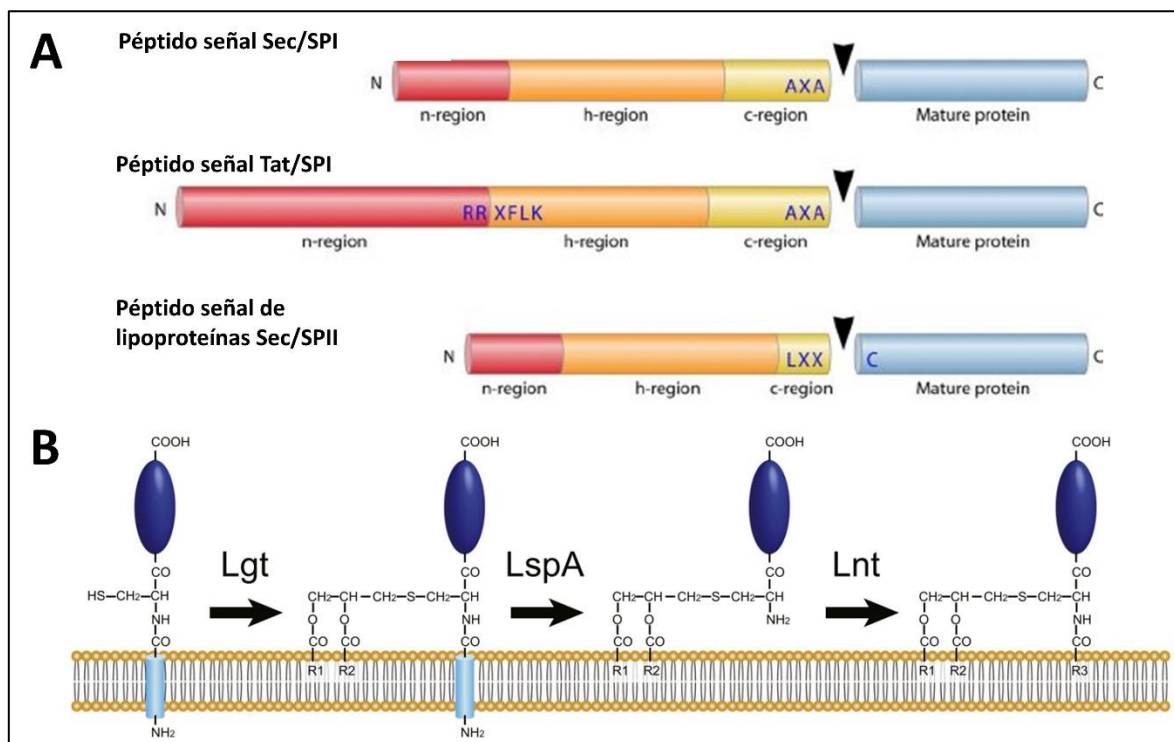


Figura 3. Los péptidos señal presentes en proteínas bacterianas. A. Los péptidos señal son reconocidos por los sistemas de translocación a través de la membrana Sec o Tat. La región N-terminal es reconocida por Sec o Tat, la región C-terminal es reconocida por la

peptidasa de señales (SP), ya sea SPI o SPII (LspA). B. En el caso de aquellas proteínas que contengan el péptido señal Sec/SPII (con el motivo LXX-C en la región C-terminal), estas proteínas serán encaminadas a ser modificadas postraduccionalmente a lipoproteínas. La enzima transferasa Lgt le une una molécula di-acilglicerol a la cisteína N-terminal, posteriormente, la peptidasa de señales LspA escinde el péptido señal. Por último, la transferasa Lnt le une otro grupo acilo. Dando como resultado una “pro-lipoproteína” triacilada embebida en la membrana (tomados de Dalbey et al., 2012; Narita and Tokuda, 2017).

Particularmente, los péptidos señal Sec/SPII de lipoproteínas, que son translocados por Sec y procesados por la peptidasa de señales II (Figura 3A), difieren en el sitio de corte (región C-terminal), al cual se le conoce como *lipobox*, y encaminan a las proteínas a ser transformadas en lipoproteínas (Cole et al., 2021; Narita and Tokuda, 2017; El Rayes et al., 2021). El *lipobox* está altamente conservado, y contiene una cisteína al extremo C-terminal, que será el residuo N-terminal de la proteína madura (Figura 3B). Esta cisteína es indispensable para su procesamiento, ya que, la acil-transferasa Lgt une, a través de un enlace tioéster, una molécula di-acilglicerol al grupo tiol de esta cisteína (Figura 3B). Posteriormente el péptido señal es escindido por la peptidasa Lsp entre la glicina y la cisteína, así como la transferencia de un grupo acilo por la acil-transferasa Lnt (Narita and Tokuda, 2017; Okuda and Tokuda, 2011). El resultado es una llamada pro-lipoproteína, embebida en la membrana en la cara periplasmática. Esta lipoproteína, tiene cuatro destinos finales posibles: quedar embebida en cualquiera de las membranas en su cara periplasmática; quedar embebida en la membrana externa (en el caso de las didermas); quedar expuesta hacia el medio extracelular; o ser liberada al medio extracelular (Lorenz et al., 2019; Okuda and Tokuda, 2011; El Rayes et al., 2021).

Las lipoproteínas son importantes para la supervivencia de las bacterias y están involucradas en casi todos los procesos celulares. Están involucradas en el mantenimiento de la integridad de la envoltura celular bacteriana (Dubuisson et al., 2005; Hellman et al., 2002; Lazzaroni and Portalier, 1992), en la morfología celular

(Paulsson et al., 2021), en la multiresistencia de antibióticos mediada por las bombas de eflujo (Lauren E. Brooks, a Sabah Ul-Hasan, a Benjamin K. Chan, 2018; Li et al., 1995), en la captación de hierro y en la secreción de proteínas (Alvarez-Martinez et al., 2021; Green and Mecsas, 2016; El Rayes et al., 2021). Algunas lipoproteínas forman parte de apéndices celulares como los flagelos y los pili (Jin et al., 2001), mientras que otras son componentes integrales de la maquinaria de biosíntesis de peptidoglicano y de sistemas de transporte de proteínas y lipopolisacáridos (Lorenz et al., 2019; Okuda and Tokuda, 2011). En algunos casos, las lipoproteínas son enzimas con actividades proteolíticas y de degradación de la pared celular de plantas (Liao et al., 2022). Curiosamente, la primera lipoproteína descubierta es la proteína más abundante en *E. coli*, Lpp, también llamada lipoproteína muréica o lipoproteína de Braun (Braun and Rehn, 1969). En una sola célula puede haber hasta un millón de moléculas de Lpp y también es una de las más estudiadas a lo largo de los años (Asmar and Collet, 2018). Lpp parece existir en dos formas: una forma 'unida', donde alrededor de un tercio de las moléculas de Lpp están unidas covalentemente a la PG y controlan el ancho del espacio periplásmico; y el resto se presentan en forma 'libre' expuesta en la superficie celular, cuyo papel aún se desconoce (Asmar et al., 2017; Choi et al., 1986; Cowles et al., 2011; Dramsi et al., 2008).

Además de los roles descritos para las lipoproteínas en bacterias modelo y patógenos humanos, las bacterias fitopatógenas a menudo dependen de las lipoproteínas para adherirse a sus sustratos biológicos preferidos, lo que es un paso importante en la colonización del huésped. El sistema LolABCDE juega un papel crítico en la correcta localización celular de lipoproteínas periplásmicas y expuestas en la superficie (Konovalova et al., 2017; Narita and Tokuda, 2017; Okuda and Tokuda, 2011; Wilson and Bernstein, 2016). Un componente particular de este sistema es la lipoproteína LolB, que une la lipoproteína objetivo a la bicapa lipídica externa (Hooda and Moraes, 2018; Wilson and Bernstein, 2016). Un estudio reciente mostró que una cepa mutante de *Xanthomonas campestris* pv. *campestris*, incapaz de producir LolB, tiene fenotipos pleiotrópicos, que van desde

la disminución de la producción de biofilm hasta la disminución de la virulencia (Liao et al., 2022). LolB es necesario para la correcta unión de las lipoproteínas a la capa externa, por lo que estos síntomas podrían explicarse por el defecto general en la integridad de la membrana celular y la perturbación de la homeostasis de las lipoproteínas. LolB también es necesario para la supervivencia *in-planta* en el fitopatógeno invasor del xilema *Ralstonia solanacearum* (Su et al., 2021).

Hay lipoproteínas con funciones inusuales en las bacterias fitopatógenas. Por ejemplo, las dos pequeñas lipoproteínas expuestas en la superficie, EcnA y EcnB, modulan la agregación celular, la formación de biofilm, la motilidad, la liberación de vesículas de la membrana externa y la resistencia a especies reactivas de oxígeno (Bishop et al., 1998; Granato et al., 2019; Schubiger et al., 2015). Las lipoproteínas también participan en las interacciones hospedero-patógeno. Por ejemplo, la lipoproteína VacJ es un componente de la maquinaria de mantenimiento de la asimetría lipídica y, cuando está ausente en *Xanthomonas citri*, resulta en una disminución de la formación de biofilm en tubos de vidrio y en las superficies de hojas, disminución de la motilidad por swarming, reducción del crecimiento bacteriano *in-planta* y reducción de la virulencia (Li and Yu, 2020). Otra lipoproteína de *X. citri*, OmlA, es importante para la resistencia a múltiples fármacos y puede estar implicada en interacciones proteína-proteína y mantenimiento de la integridad de la membrana externa (Fuangthong et al., 2008; Vanini et al., 2008). Las lipoproteínas también son importantes para la producción y exportación de exopolisacáridos en algunas bacterias. En *Rhizobium leguminosarum*, la lipoproteína PssN es parte del complejo de polimerización y exportación de exopolisacáridos PssTNOP, que se necesitan para la colonización exitosa de su hospedero (Marczak et al., 2006; Wielbo et al., 2004). En conclusión, las lipoproteínas son esenciales para la supervivencia y la interacción entre la bacteria y su entorno, ya sea con otros seres vivos o con el sustrato donde habitan.

I.4.2 Anclaje a la pared de peptidoglicano mediada por sortasas

Las bacterias monodermas pueden fijar proteínas a la capa de peptidoglicano de una manera única: a través del sistema de *sorting* de proteínas (Dai et al., 2019; Dramsi et al., 2008; Nielsen et al., 2019; Susmitha et al., 2021). En este sistema, se utilizan a las enzimas sortasas, que son proteínas ligadas al peptidoglicano y que reconocen un motivo conservado en la región C-terminal de las proteínas blanco (Novick, 2000). El motivo LPXTG (Figura 4) es el más estudiado, y es reconocido por la sortasa SrtA, conservada en múltiples clados de bacterias monodermas (Budzik et al., 2007; Comfort and Clubb, 2004; Dai et al., 2019; Douillard et al., 2014; Egan et al., 2010; Leigh et al., 2010; Liew et al., 2004; Malik and Kim, 2019; L., Si et al., 2016). Existen diferentes variaciones de este motivo, por ejemplo, los motivos LPXTA, LAXTG, LPXXXD, NP[QK]T[NG], entre otros, que son reconocidos por otras familias de sortasas (Desvaux et al., 2006; Fimereli et al., 2012; Litou et al., 2008; Malik and Kim, 2019).

La sortasa SrtA reconoce y corta estos motivos entre los residuos de treonina y glicina (LPXT / G). Posteriormente, el lípido II, que es el bloque de construcción de la capa de peptidoglicano en crecimiento, lleva a cabo un ataque nucleofílico al complejo sortasa-proteína, formando el intermediario proteína-lípido II (Desvaux et al., 2018; Dramsi et al., 2008; Leigh et al., 2010). La molécula lipídica II que acarrea a la proteína, a través de la maquinaria de biogénesis de peptidoglicano, será anclada a la pared celular (Figura 4).

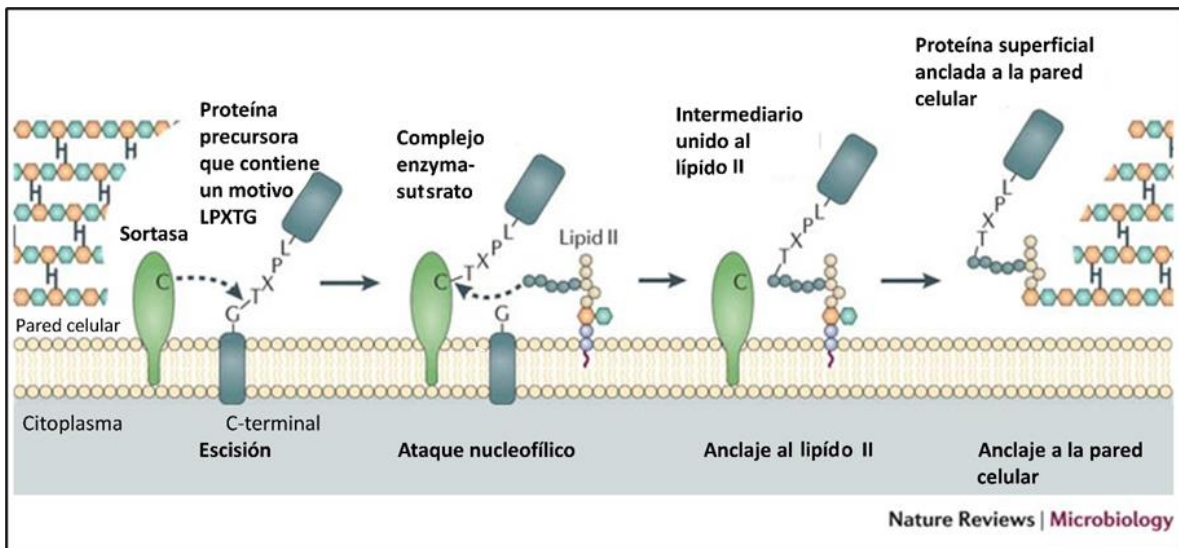


Figura 4. Mecanismo de anclaje por parte de las sortasas a la pared de peptidoglicano de las proteínas que contienen un motivo LPXTG-like (adaptado de Leigh et al., 2010).

La fijación de proteínas en la superficie mediada por sortasas es importante para algunos patógenos bacterianos como los estafilococos, estreptococos, enterococos y en particular en especies patógenas como *Listeria monocytogenes* (Dai et al., 2019; Egan et al., 2010; Hendrickx et al., 2009; Hendrickx et al., 2011). En algunas bacterias fitopatógenas, como la bacteria patógena de tomate y papa, *Clavibacter michiganensis* ssp. *michiganensis*, la sortasa es esencial para la formación de ampollas, que es una característica del proceso infeccioso, pero no para los síntomas de marchitamiento. Esto sugiere que es importante para la adhesión de las bacterias a la superficie de la hoja (Chalupowicz et al., 2017). Además, la proteasa celular unida a la pared Pat-1, un factor de virulencia importante y sustrato de la sortasa está ausente en algunos aislados no virulentos de especies de *Clavibacter*. Por lo tanto, se ha asumido que es relevante en el proceso de infección (Gartemann et al., 2003). En especies del género *Bacillus*, que en muchos casos son bacterias que promueven el crecimiento de plantas y en algunos casos son patógenos oportunistas, las sortasas son vitales para la colonización de las plantas y están involucradas en la síntesis de pili y la formación de endosporas (Budzik et al., 2007; Hendrickx et al., 2011). La inhibición de las sortasas en patógenos multirresistentes como objetivo terapéutico ha sido

un tema muy estudiado y podría resultar útil para tratar bacterias fitopatógenas (Cascioferro et al., 2014; Cossart and Jonquieres, 2000; Kudryavtsev et al., 2021; Ouyang et al., 2018; T., Y., Si et al., 2016).. Sin embargo, se necesita más investigación para comprender todas las implicaciones que las sortasas y sus sustratos tienen en las enfermedades de las plantas

I.4.3 Sistema de secreción tipo IX

Dentro de los nueve tipos de sistemas de secreción descubiertos hasta ahora, el más recientemente descrito es el sistema de secreción tipo IX (T9SS). El T9SS se encontró por primera vez en la bacteria *Porphyromonas gingivalis*, y se relacionó con la producción de porfirinas, una molécula que le da un característico color negro a las colonias cuando esta bacteria crece en medio sangre. Por lo tanto, el T9SS fue primero referido como el sistema de secreción 'Por' (refiriéndose a las porfirinas, no al nombre de la especie como podría pensarse) (De Diego et al., 2016; Gorasia et al., 2015; Veith et al., 2017). Este sistema de secreción está presente en una gran parte de las especies pertenecientes al superfilum de bacterias didermas Fibrobacteres-Chlorobi-Bacteroidetes. Las bacterias pertenecientes a este superfilo, también llamado grupo FCB, están ampliamente distribuidas en los suelos, en los sedimentos y en los océanos, así como parte del rumen y de la microbiota intestinal, y patógenos de animales (Veith et al., 2017). El T9SS tiene una peculiaridad que lo distingue de los otros ocho sistemas de secreción, modificando postraduccionalmnte las proteínas secretadas al unir covalentemente un lipopolisacárido aniónico (A-LPS). Los A-LPS son parte integral de la membrana externa (Walker et al., 2004), por lo que las proteínas unidas a estos lipopolisacáridos son ancladas a la cara exterior de la membrana externa (Figura 5). El T9SS reconoce un dominio llamado Dominio C-Terminal (CTD) dentro de la secuencia de la proteína, que es escindido (De Diego et al., 2016). Posteriormente la proteína se conjuga a un A-LPS anclado en la membrana externa de la bacteria (Gorasia et al., 2015). Las proteínas que contienen estos dominios reconocidos por el T9SS tienen bastantes funciones, algunos incluyen

enzimas implicadas en el metabolismo de biopolímeros, dentro de las cuales se pueden encontrar proteinasas, celulasas, quitinasas, nucleasas y lipasas y procesos de motilidad como el “gliding” (Abby and Rocha, 2017). Otros ejemplos de sustratos del T9SS llevan a cabo funciones estructurales y de adhesión, como las adhesinas, las hemaglutininas o las proteínas de la capa virulenta (también conocida como la capa superficial electrodensa) (Veith et al., 2017).

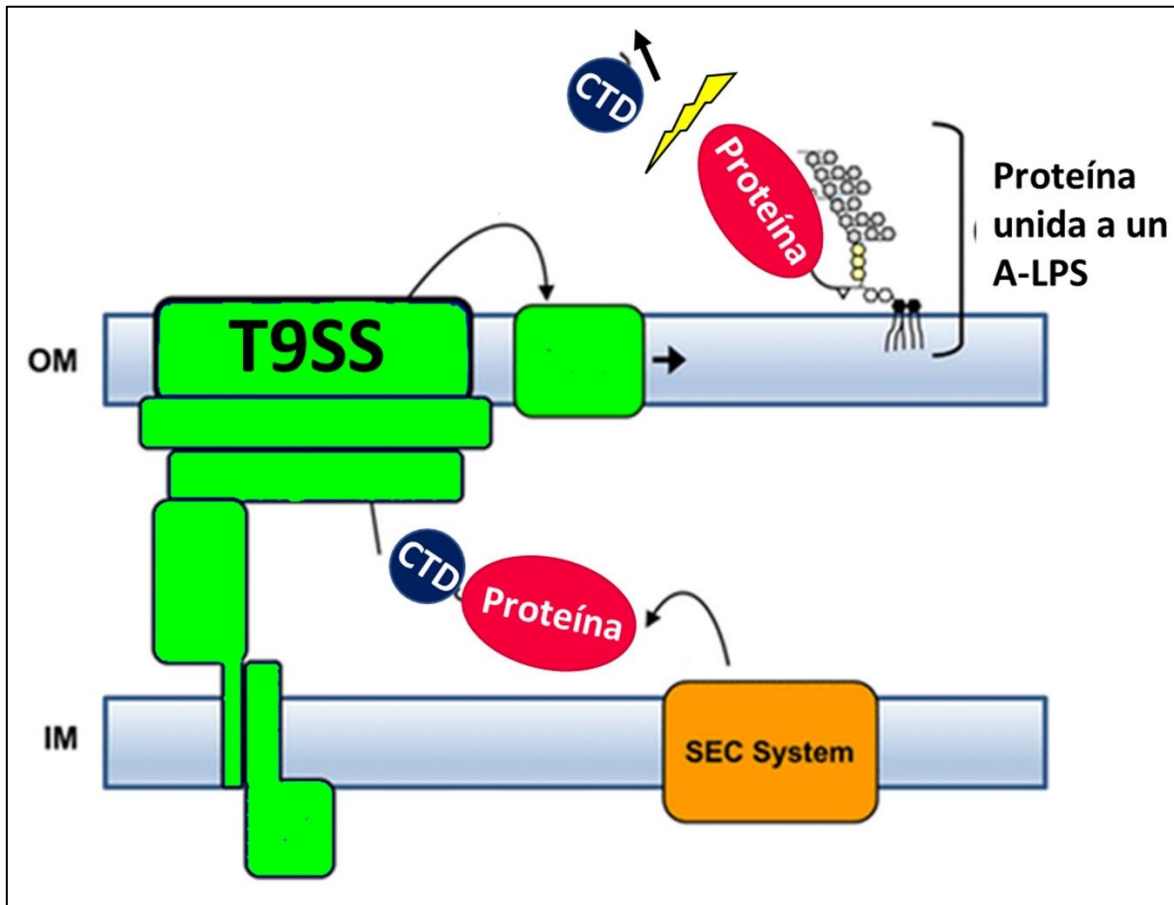


Figura 5. Secreción de proteínas por el sistema de secreción tipo IX. Las proteínas que contienen un CTD son reconocidas por el T9SS, después de ser translocadas por el sistema Sec. Posteriormente, se escinde el CTD y son enlazadas a un lipopolisacárido aniónico (A-LPS), quedando ancladas a la membrana externa de las bacterias (adaptado de Veith et al., 2017).

I.8 Antecedentes

Recientemente en nuestro laboratorio, se ha asumido el enfoque de encontrar la función biológica de las expansinas, en el contexto del proceso de infección de bacterias fitopatógenas. Para tratar de entender su participación en este proceso, se han realizado diferentes experimentos, por ejemplo, la capacidad de las expansinas de *Pectobacterium brasiliense* (PbExl1) y de *P. atrosepticum* (PaExl1) para solubilizar un polisacárido, aún no identificado, del tejido vascular aislado de la acelga suiza (Tovar-Herrera et al., 2018a). El alcance de esta observación a otras expansinas bacterianas y su relevancia biológica no son claras, aunque se sugirió una función en la colonización de las plantas por bacterias. Asimismo, se encontró que las expansinas inducen la respuesta inmune de las plantas (Narvéez-Barragán et al., 2020) y que se localizan en el espacio intracelular de las células circundantes de los haces vasculares en el xilema de apio (Tovar-Herrera et al., 2018a). Estas evidencias, junto con el hecho de que al observar que la ausencia de las expansinas afecta el proceso de colonización de plantas en especies de *Bacillus*, *Ralstonia*, *Xanthomonas*, *Clavibacter*, *Pectobacterium*, *Erwinia*, etc. (Georgelis et al., 2011; Junior et al., 2015; Martínez-Anaya, 2016b; Olarte-Lozano et al., 2014; Rocha et al., 2020; Tancos et al., 2018), sugiere que las expansinas están participando en la interacción planta-hospedero; pudiendo facilitar la colonización de la planta o la penetración hacia los tejidos internos, así como localizar y señalar el sitio de adherencia, como en casos de bacterias que infectan el xilema (por ejemplo, *R. solanacearum*). Además de los métodos experimentales, las herramientas bioinformáticas ayudan a identificar características clave en las secuencias y estructuras de las proteínas, como la cantidad de carga eléctrica en su superficie (Pastor et al., 2015). Esto es importante porque la carga superficial de las expansinas afecta la unión a diferentes tipos de polisacáridos dentro de la pared celular de la planta. Por ejemplo, la expansina básica BsEXLX1 (con un punto isoeléctrico teórico (pI) de 9.04) se une a los polisacáridos negativamente cargados en la mayoría de los tipos celulares de los pecíolos de apio, mientras que en las mismas condiciones,

PbExl1, una expansina ácida (pI de 4.78), se localiza en las esquinas celulares entre las células vasculares y las células circundantes (Tovar-Herrera et al., 2018a). Además, se han encontrado módulos de proteínas adicionales en las expansinas microbianas que podrían arrojar luz sobre su actividad. Por ejemplo, la expansina de *Clostridium clariflavum* (*CcEXL1*) contiene un dominio dockerina, que, debido a su asociación con la scaffoldina ScaG, la une a la superficie celular (Artzi et al., 2015; Artzi et al., 2016). Esto podría sugerir una participación en la utilización de celulosa en esta especie y una función de acercamiento de la célula al sustrato. Otra característica relevante contenida en la secuencia de todas las expansinas es la presencia de un péptido señal en el extremo N-terminal, que las encamina a ser secretadas al medio extracelular. En el caso de las expansinas bacterianas, desde su descubrimiento, se ha asumido que la proteína madura es liberada fuera de la célula para unirse a un ligando externo (es decir, a los polisacáridos de la pared celular de la planta) (Cosgrove, 2017a). De hecho, las expansinas producidas de manera heteróloga provenientes de *B. subtilis*, *Pectobacterium brasiliensis*, *Hahella chejuensis*, entre otras, son solubles, e interactúan con las paredes celulares de las plantas y la celulosa (Kerff et al., 2008; Lee et al., 2013; Olarte-Lozano et al., 2014). Sin embargo, *BsEXLX1* también une con alta afinidad al peptidoglicano (Kerff et al., 2008), y *PbExl1* producida endógenamente por *P. brasiliense* se encuentra principalmente en el pellet celular (al crecerse en medio líquido LB), lo que sugiere que permanece unida a las bacterias después de su procesamiento (Narváez-Barragán et al., 2020). Por lo tanto, existe la posibilidad de que las expansinas bacterianas se asocien a las propias envolturas celulares de las bacterias que las expresan, donde estarían llevando a cabo su función biológica.

En este estudio, se realizó una búsqueda actualizada de proteínas expansinas en proteomas bacterianos anotados y exploramos la relación de ciertas características estructurales y secuenciales: relaciones filogenéticas, punto isoeléctrico y modularidad de dominios proteicos. También se estudió sus señales peptídicas de secreción que, curiosamente, indican que una vez translocadas,

muchas expansinas pueden permanecer asociadas con la superficie bacteriana. Estos hallazgos aumentan nuestro conocimiento sobre cómo la actividad de las expansinas bacterianas podría ser interpretada en el contexto de las interacciones planta-bacteria o para la industria energética, y sugieren posibles roles para las expansinas en bacterias de vida libre.

Capítulo II. Justificación

II.1 Justificación

Las expansinas en bacterias han sido objeto de estudio durante casi 20 años, y aún no está claro su mecanismo de acción ni su función biológica, llegando a un estancamiento. Sin embargo, se ha encontrado que son importantes en el proceso de colonización e infección de las bacterias patógenas de plantas, así como para algunas bacterias promotoras del crecimiento. Por lo tanto, es necesario un cambio en el paradigma del estudio de las expansinas, tomando en cuenta su secuencia, las características fisiológicas de cada especie, y la posibilidad de que la función de las expansinas esté relacionada a la envoltura celular bacteriana.

II.1 Hipótesis

Las expansinas se encuentran asociadas a la envoltura celular de las bacterias a través de diversos métodos.

II.3 Objetivo general

Identificar mediante técnicas bioinformáticas patrones que permitan obtener indicios sobre la actividad de las expansinas.

II.3.1 Objetivos particulares

- Generar una base de datos de secuencias de expansinas a partir de los proteomas de bacterias depositados en UniProt.
- Analizar y clasificar a los péptidos señal de las expansinas bacterianas.

- Determinar si existen otras señales de clasificación en la secuencia de las expansinas (motivos LPXTG-like, T9SS-CTDs, dominios adicionales).
- Combinar toda esta información en busca de conjeturas sobre la actividad de las expansinas en bacterias.

Capítulo III. Métodos

III.1 Obtención y alineamiento de secuencias

Las secuencias de expansinas fueron adquiridas de la base de datos de UniProt, basándose en aquellas identificadas y descritas en la literatura. Para alinear las secuencias de proteínas, se utilizó el programa ClustalO con parámetros por defecto. Este programa está disponible para su uso en el servidor público <https://www.ebi.ac.uk/Tools/msa/clustalo/> (Sievers et al., 2011).

III.2 Búsqueda de secuencias

Para buscar secuencias de expansinas, se utilizó el método de perfiles ocultos de Markov (HMM), que se especializa en encontrar secuencias homólogas distantes gracias a la información de cada una de las posiciones de aminoácidos. Esta información se obtiene a partir del alineamiento de varias secuencias, para lo cual se utilizó el alineamiento de 44 secuencias de expansinas bacterianas que ya han sido caracterizadas y confirmadas en la literatura. La selección de las secuencias de expansinas se basó en aquellas con análisis bioquímicos donde se comprueba su pseudoactividad y que contengan los residuos conservados descritos como indispensables. Se construyó un perfil HMM usando el programa *hmmbuild* del paquete HMMER v3.3.2 (Eddy, 2009) utilizando el alineamiento de expansinas como base. Para la búsqueda de secuencias de expansinas, se utilizó el programa *hmmsearch*, del mismo paquete, en contra de proteomas de bacterias contenidos en la base de datos de proteomas UniProt (Bateman et al., 2021), utilizando la versión de noviembre del 2021. Se tomó un límite de puntaje por secuencia resultado ('bit score per-target output') de 50, el cual fue determinado por la aparición de falsos positivos (a menor puntaje, mayor la proporción de falsos positivos). Las secuencias encontradas fueron curadas manualmente, comprobando la longitud de la secuencia del dominio de la expansina (~220 aa), y

la presencia de los aminoácidos conservados. Se construyó una base de datos con la información de cada secuencia y de la especie de origen.

III.4 Cálculo del punto isoeléctrico

El punto isoeléctrico teórico de cada secuencia fue determinado con el programa IPC v2.0 disponible para su uso en <http://isoelectric.org/>. Se tomó el valor dado por el método “IPC protein” (Kozlowski, 2016).

III.5 Análisis de los péptidos señal

Para analizar los péptidos señales se utilizó el programa SignalP v5.0 (Almagro Armenteros et al., 2019), disponible para su uso en <https://services.healthtech.dtu.dk/services/SignalP-5.0/>. Este programa utiliza la secuencia de cada péptido señal para determinar el sistema de translocación a través de la membrana que la reconoce. SignalP-5.0 clasifica los péptidos señal en cuatro diferentes tipos: a) ‘SP(Sec/SPI)’, que es translocado por el sistema Sec y cortado por la peptidasa de señales I; b) ‘TAT(Tat/SPI)’, que es translocado por el sistema Tat y cortado por la peptidasa de señales I; c) ‘LIPO(Sec/SPII)’, que es translocado por el sistema Sec y cortado por la peptidasa de señales II, adicionalmente se encamina a la proteína ser modificada a lipoproteína; y d) ‘OTHER’, refiriéndose a que contiene un péptido señal, pero que no fue reconocido por el programa. Esta información se agregó a la base de datos de secuencias de expansinas.

III.6 Búsqueda de motivos LPXTG-like

Para la búsqueda de posibles señales de *sorting* de proteínas, se utilizó el programa FIMO v5.5 del paquete MEMESuite (Grant et al., 2011), disponible en <https://meme-suite.org/meme/tools/fimo>. Se utilizaron los motivos LPXTG, NP[Q/K]TN, [I/L][P/A]XTG, LPNTA y LAXTG que están descritos en Malik and Kim,

2019, utilizando un e-value de $1e-5$. Únicamente se analizaron las expansinas que provenían de bacterias monodermas, ya que las bacterias didermas no contienen sortasas en sus genomas (con excepción de algunas proteobacterias marinas como *Marinobacter* spp., que no contienen expansinas).

III.7 Predicción de dominios adicionales

Se analizaron las secuencias completas para determinar la presencia de dominios proteicos conservados adicionales a los dominios característicos de las expansinas. Para esto se utilizó el programa InterProScan v5.5 (Jones et al., 2014) disponible para su uso en línea en <https://www.ebi.ac.uk/interpro/search/sequence/>. Se agregó a la base de datos la información de la arquitectura de dominios adicionales y se les clasificó como 'expansinas multimodulares'. En el caso de las expansinas que no contenían dominios adicionales, se les clasificó como 'expansinas monomodulares'.

III.8 Relaciones filogenéticas

Para entender la evolución y las relaciones filogenéticas de las expansinas, se tomó la porción de cada secuencia que se alineó al perfil HMM construido para las expansinas, dejando fuera del análisis a cualquier dominio adicional y a los péptidos señal. Se alinearon todas las secuencias usando el método antes descrito. Se construyó un árbol por máxima verosimilitud con el programa IQ-TREE (Nguyen et al., 2015) utilizando la selección del modelo de sustitución automática, y 1000 iteraciones bootstrap ultrarrápido. El cladograma resultante fue visualizado en el servidor iTol (Letunic and Bork, 2016). Para la determinación de grupos de secuencias, se construyó una matriz de similitud 'todos contra todos' con el programa ClustalO con las banderas `--distmat-out`, `--full`, y `--percent-id`. La matriz se visualizó con el paquete heatmap2 en R.

III.10 Análisis de regiones desordenadas

Para la determinación y el análisis de regiones desordenadas presentes en las secuencias de algunas expansinas, se utilizaron los programas IUPred2 (long) y ANCHOR2 (Mészáros et al., 2018), disponibles como un paquete en <https://iupred2a.elte.hu/>.

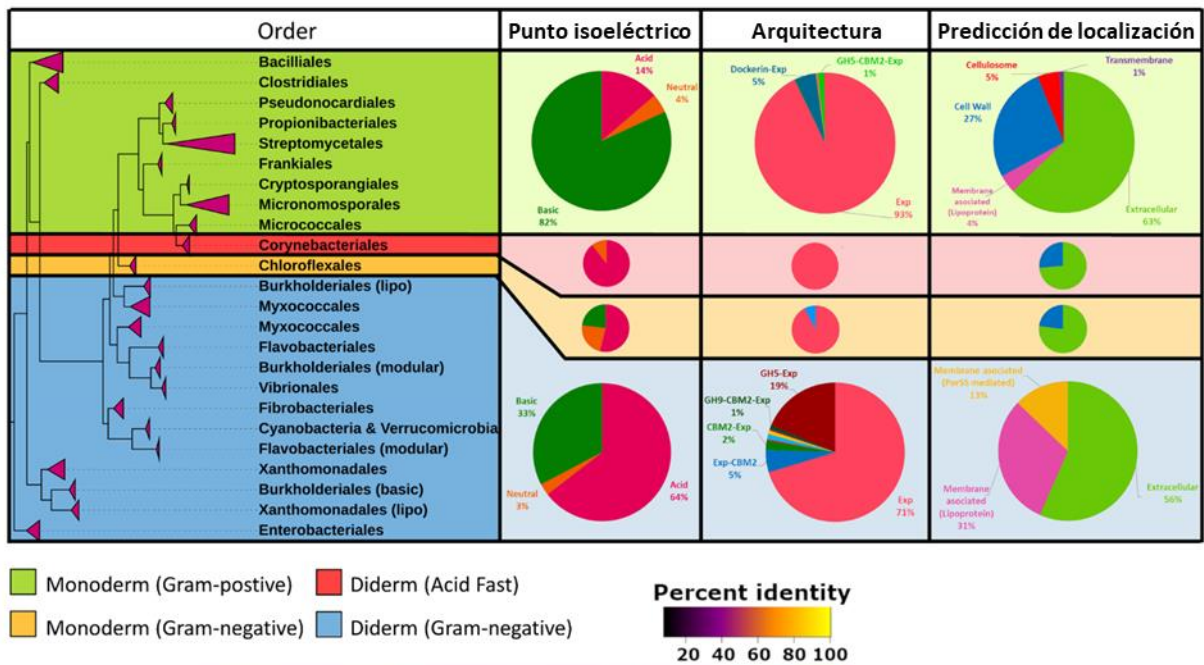
III.11 Predicción de estructuras de proteínas

Para obtener las estructuras seleccionadas como representativas de cada tipo de arquitectura de dominios, se utilizó trRosetta (Yang et al., 2020) y se visualizó con Mol* (usando la representación '*Spacefill*'). Las proteínas seleccionadas para cada arquitectura fueron: expansina monomodular canónica (UniProt Id: C6DHV9); expansina monomodular con un motivo LPXTG-like (UniProt Id: D3JXR9; PDB: 3D30); expansina modular con un dominio celulasa N-terminal (UniProt Id: Q8P513); expansina modular con un dominio CBM2 C-terminal (UniProt Id: F0Q783); expansina modular con un dominio CBM2 N-terminal (UniProt Id: A9FSQ7); expansina modular con un dominio dockerina N-terminal (UniProt Id: G8LUU0); expansina modular con un dominio CTD reconocido por el T9SS (UniProt Id: A0A2N3TEV7); expansina con un dominio transmembranal, y una región desordenada altamente polar (UniProt Id: A0A010ZNG5); y la expansina modular con un dominio GH5 y un dominio CBM2 (UniProt Id: A5CLK3).

Capítulo IV. Resultados

En este proyecto, realizamos un estudio detallado utilizando análisis bioinformáticos con secuencias de expansinas bacterianas recientemente depositadas en la base de datos de secuencias de proteínas UniProt. La alta divergencia de secuencia y la falta de una secuencia de consenso de la familia de proteínas de expansina bacteriana han dado como resultado una anotación ambigua en todas las bases de datos (por ejemplo, "lipoproteína rara A", "proteína putativa que alberga un dominio endoglucanasa C-terminal", "Proteína que contiene el dominio de unión a peptidoglicano", etc.). Por lo tanto, utilizamos un enfoque de modelos ocultos de Markov (HMMs) para encontrar expansinas en proteomas bacterianos. Construimos un perfil de HMM utilizando 44 dominios de expansina y lo buscamos en proteomas no redundantes en UniProt, obteniendo 908 expansinas putativas compuestas por los dos dominios de expansina: D1, en el extremo N-terminal con un doble psi-barril beta que contiene el aspartato pseudocatalítico en la posición equivalente al D82 de *B. subtilis*; también estas secuencias deben de contener el dominio D2 que es un dominio de unión a polisacáridos que contiene tres residuos aromáticos para la unión al sustrato (Kerff et al., 2008). Asimismo, se descartaron secuencias truncas, pero se incluyeron las secuencias que contienen dominios proteicos adicionales al dominio de la expansina. Nuestros resultados de búsqueda son similares a otros informados anteriormente, en los que encontramos genes de expansina en nueve filo bacterianos y 43 órdenes. La mayoría de las secuencias encontradas pertenecen a Actinobacteria (47.7%), Proteobacteria (29.9%) y Firmicutes (14.6%), y el 7.8% restante pertenece a Fibrobacteres, Bacteroidetes, Chloroflexi y Cyanobacteria. También encontramos tres y dos secuencias de expansina en los filo Verrucomicrobia y Spirochaetes, respectivamente. Debido a la sobrerrepresentación de genomas secuenciados de clados con mayor interés dentro de las bases de datos de secuencias (como las gamma-Proteobacterias y las Streptomyetales), no es del todo claro si estas proporciones representan la distribución de las expansinas en bacterias.

A



B

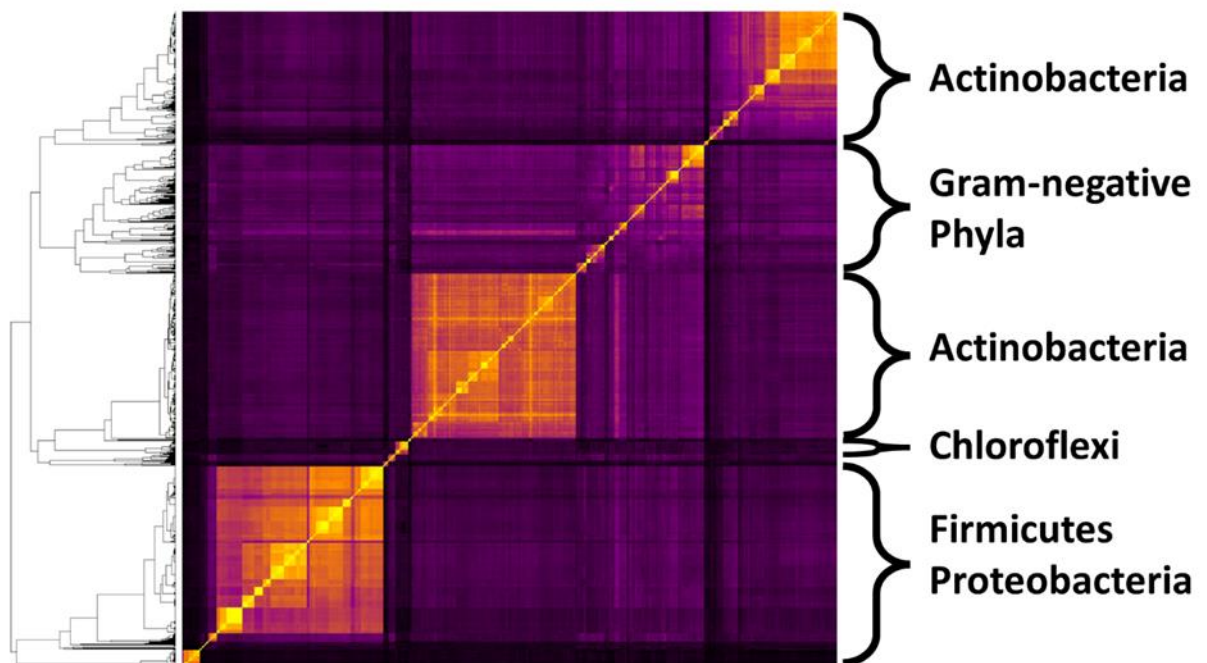


Figura 6. Características de las expansinas bacterianas. A, cladograma de los órdenes de bacterias que contienen expansinas (basado en el árbol filogenético representado en la Figura 7). Las gráficas circulares exponen la proporción de características de las secuencias de expansinas analizadas en este estudio. “Punto isoelectrico”, se calculó el punto isoelectrico teórico de cada secuencia y se clasifico en:

Ácida ($\text{pH} < 6.5$), Neutral ($6.5 < \text{pH} < 7.5$), Básica ($\text{pH} > 7.5$). “Arquitectura”, se analizó la arquitectura de dominios de cada secuencia y se clasificaron dependiendo de los dominios presentes y su orden N-terminal a C-terminal. “Predicción de localización”, la localización celular de las expansinas fue predicha usando motivos conservados, señales, y dominios funcionales contenidos en su secuencia. **B**, mapa de calor construido a partir de una matriz de similitud entre cada par de secuencias, en el cual se forman cuatro grandes grupos de secuencias con al menos 50% de similitud.

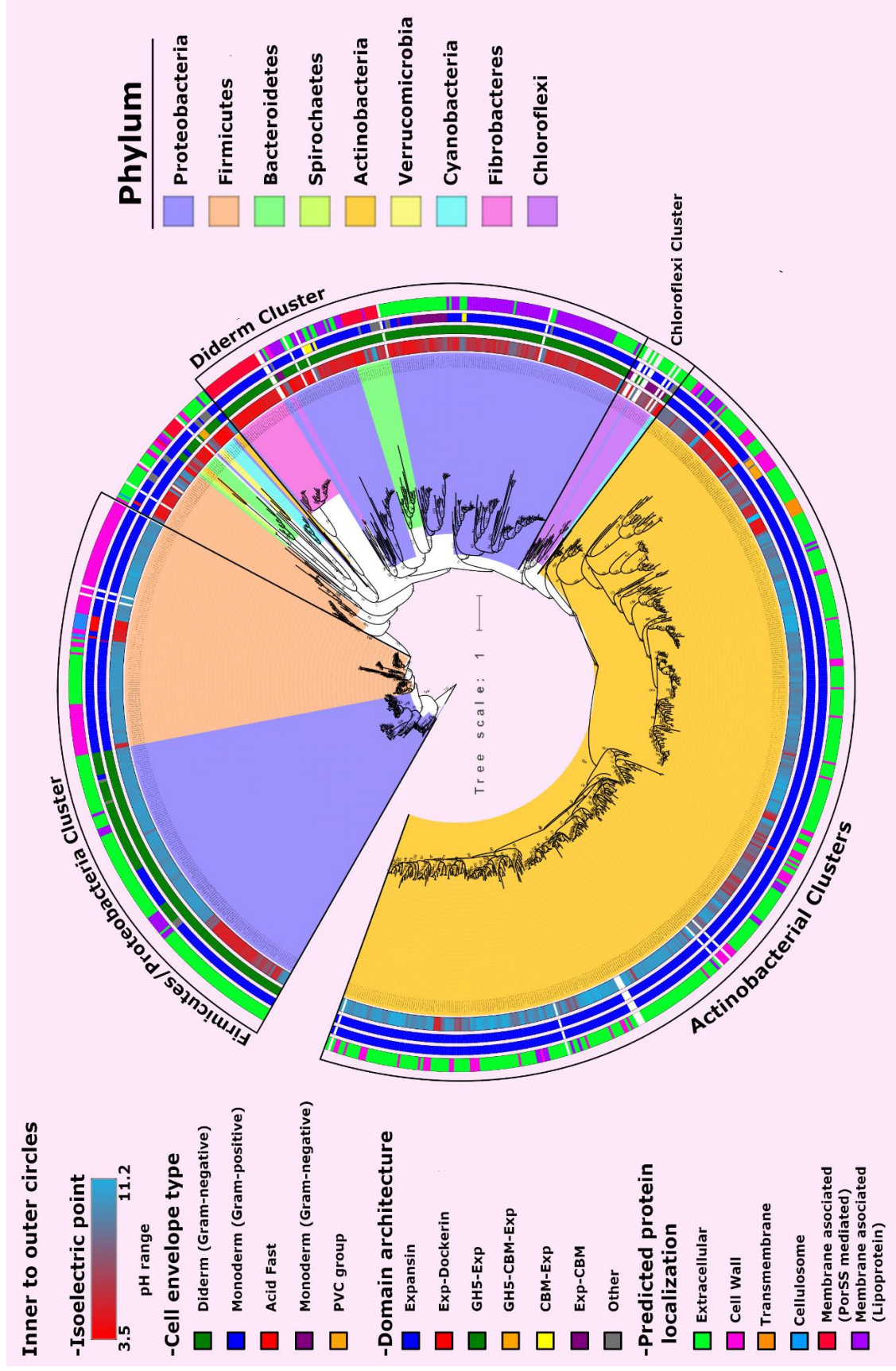


Figura 7. Árbol filogenético inferido a partir de máxima parsimonia. El árbol debe considerarse sin raíz y circular para fines de presentación. Los valores de Bootstrap se calcularon con 1000 iteraciones. Relaciones de clúster determinadas por PCA y por el heatmap de similitud en la Figura 6B. La localización de proteínas predicha se basa en motivos de secuencia y dominios adicionales. Exp, dominio expansina. CBM, módulo de unión a carbohidratos. GH5, familia 5 glicosil-hidrolasa.

De estas secuencias, el 84.5% son expansinas canónicas (expansinas monomodulares) y el 15.5% son proteínas multimodulares con dominios funcionales adicionales, como un dominio de glicosil hidrolasa de familia 5 (GH5) o un dominio de módulo de unión a carbohidratos de familia 2 (CBM2). Un mapa de calor (Figura 6B) construido a partir de una matriz de similitud por pares de secuencias, reveló que hay cuatro grupos principales de expansinas (con al menos el 50% de identidad): el grupo Firmicutes/Proteobacteria, el grupo de filidermos y dos grupos separados de Actinobacterias. Un grupo menor con 13 secuencias de expansina de especies de Chloroflexi, así como las secuencias que mostraron una alta divergencia de las secuencias de expansina canónicas no se definieron como parte de un grupo.

Analizamos diferentes características de las expansinas en una búsqueda de información que pudiera proporcionar pistas sobre su localización. Las expansinas son proteínas altamente cargadas con residuos ácidos o básicos en la superficie y esto se refleja en su punto isoeléctrico (pI); anteriormente hemos estudiado la correlación del pI con el tipo de bacteria: las bacterias monodermas contenían principalmente expansinas básicas, mientras que la mayoría de las bacterias didermas contenían expansinas ácidas (Pastor et al., 2015).

Nuestro análisis de 908 expansinas mostró una situación más compleja, con valores de pI que van desde 3.5 hasta 11 en diferentes órdenes bacterianos (Figura 7); los clados específicos muestran un continuo de valores de pI, como Burkholderiales (Figura 8), Clostridiales (Figura 9) o Streptomycetales (Figura 10). Razonamos que esto podría estar relacionado con las características y el estilo de vida específicas de cada bacteria o relacionado con la presencia de dominios adicionales. Para probar esta hipótesis, agrupamos las bacterias según su envoltura celular como 'Acid-fast' (Corynebacteriales), monodermas (Gram-positivas), monodermas (Gram-negativas), didermas (Gram-negativas). Posteriormente, se estudió la variedad de arquitecturas de dominios proteicos para distinguir las expansinas monomodulares de las expansinas multimodulares

(Figura 6A). También analizamos a las expansinas en función del tipo de péptido señal, y la presencia de motivos y dominios de clasificación. Con esta información, clasificamos las expansinas en seis grupos según la predicción de su localización celular: (a) "expansinas extracelulares", que solo contienen un péptido señal Sec o Tat, y debido a que no se pudo inferir más información acerca de su localización, se asume que son secretadas al medio extracelular; (b) "expansinas asociadas a la pared celular bacteriana", que contienen un péptido señal Sec o Tat y un motivo LPXTG-like, el cual es reconocido por las sortasas; (c) "expansinas transmembranales", que contienen un dominio transmembranal N-terminal; (d) "expansinas del celulosoma", que contienen un péptido señal Sec o Tat y un dominio dockerina; (e) "expansinas asociadas a la membrana (T9SS)", que contienen un péptido señal Sec y un dominio C-terminal (CTD) reconocido por el T9SS; y (f) "expansinas asociadas a la membrana (lipoproteína)" (a partir de ahora denominadas lipo-expansinas) que contienen un péptido señal Sec y un motivo de escisión de la señal por la peptidasa II, y son procesadas para convertirse en una lipoproteína incrustada en la/las membranas (Figura 1). Las bacterias didermas tienen la variedad de arquitecturas de dominio más abundante con ocho expansinas multimodulares diferentes (Figura 8), y se pudo predecir que el 44% de estas expansinas permanecen asociadas a la membrana como lipoproteínas o unidas a un A-LPS.

IV.1 Expansinas en Proteobacterias, Cianobacterias, Fibrobacteres y Bacteroidetes (bacterias didermas)

Aunque el 54% de las expansinas de Proteobacteria son proteínas monomodulares ácidas sin información de secuencia sobre su localización más allá de la requerida para la translocación a través de la membrana interna (Figura 6^a, "extracelulares"), hay casos que proporcionan pistas interesantes sobre la diversidad de expansinas en este grupo. Quizás los más informativos son los de Burkholderiales (Figura 8), compuestos por tres grupos bien definidos: las expansinas extracelulares básicas de *Ralstonia spp.* (Figura 8, subíndice a), las

lipo-expansinas ácidas y las CBM-expansinas modulares ácidas. Ambos grupos ácidos pertenecen a *Acidovorax spp.* y especies relacionadas, y se agrupan (Figura 6B, 7, grupos Gram-negativos), pero las expansinas básicas pertenecientes a las especies de *Ralstonia* tienen una alta similitud de secuencia con expansinas de Firmicutes, Xanthomonadales y Enterobacteriales (Figura 6B, grupo Firmicutes/Proteobacteria). Estas últimas (Enterobacteriales) son expansinas monomodulares ácidas, predichas para ser secretadas a través del sistema Sec y cortadas por la enzima SPI (Figura 8): la expansina ExI1 de especies de *Pectobacterium* es un representante de este grupo. Y, aunque todas las demás expansinas en este grupo son básicas, es interesante que las expansinas de Enterobacteriales sean ácidas, lo que sugiere una presión evolutiva desconocida que impulsa esta característica.

Las expansinas de *Ralstonia* (monomodulares) son básicas (Figura 8, subíndice a) de Xanthomonadales (fusionadas a un dominio GH5) (Figura 8, subíndice b), y están estrechamente relacionadas con las expansinas provenientes de especies del filo Firmicutes (Figura 6B, >60% de similitud). Esto indica una presión selectiva (posiblemente funcional) para que estas secuencias en bacterias de clados lejanos se encuentren conservadas. Por otro lado, aunque la mayoría de las expansinas presentes en Proteobacterias son monomodulares (Figura 8), parece que las expansinas multimodulares con mayor abundancia evolucionaron análogamente en tres órdenes diferentes: en Xanthomonadales, en Burkholderiales y en Myxococcales.

El 95% de las expansinas de Fibrobacteres y Bacteroidetes tienen un péptido señal Sec necesario para su translocación a través de la membrana interna y un dominio CTD que funciona como señal de reconocimiento para el T9SS (Figura 5). Este sistema de secreción puede anclar la proteína secretada que contiene el dominio CTD a la membrana externa, mediante la escisión del CTD y la posterior unión covalente a un A-LPS. El T9SS secreta proteínas relacionadas con la motilidad, adhesinas y exporta factores de virulencia a la superficie celular en el

superfilo Fibrobacteres-Chlorobi-Bacteroidetes. Se encontraron muy pocas expansinas en Cianobacterias, pero casi todas (cinco de las seis expansinas) son multimodulares y cada una de las arquitecturas modulares es única (Figura 8).

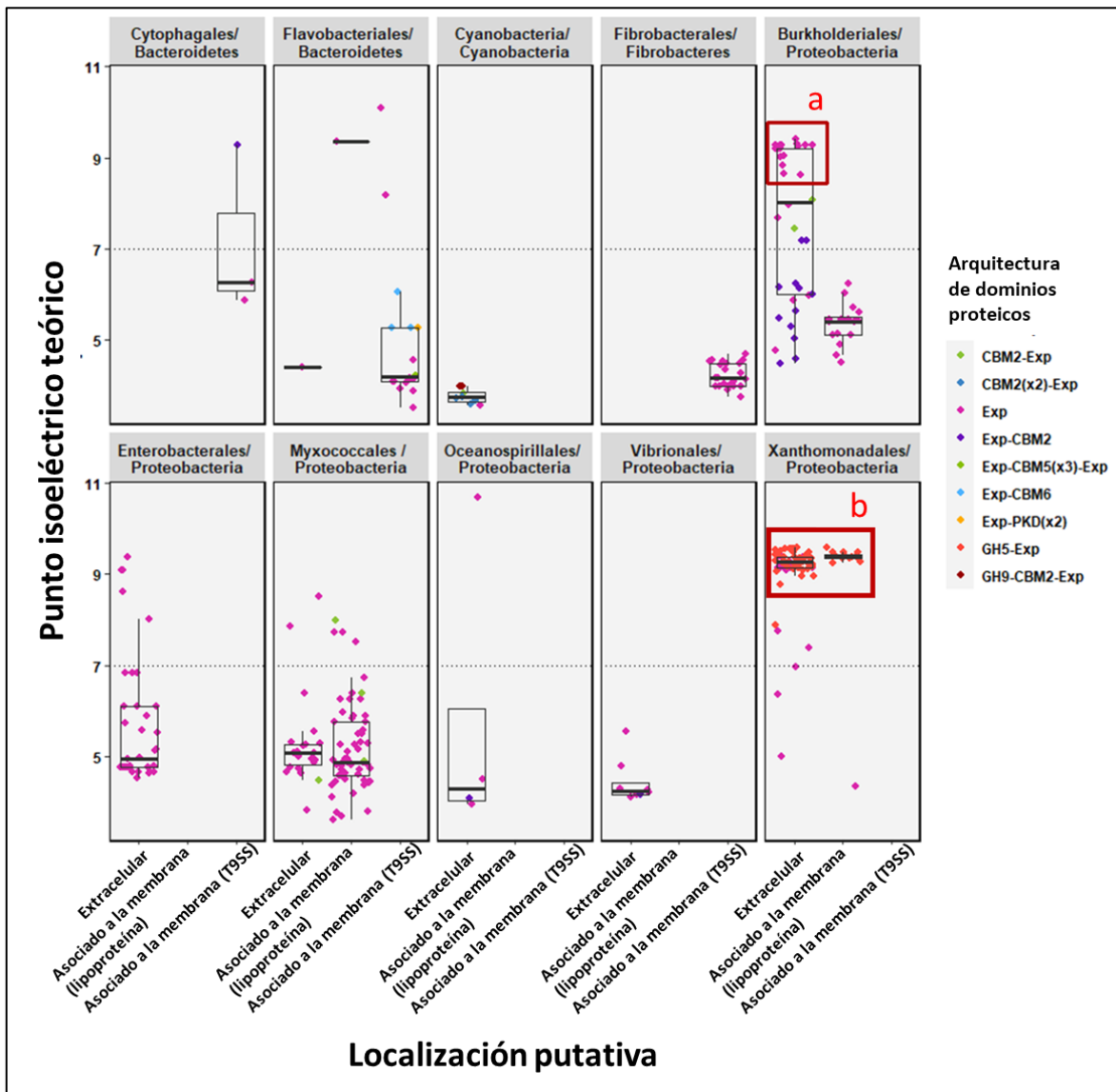


Figura 8. Expansinas provenientes de bacterias didermas. Cada punto representa una secuencia de expansina. Utilizando el punto isoelectrico teórico, graficado en el eje y, se clasifican las expansinas como básicas ($pH > 7$) o ácidas ($pH < 7$). En el eje x, se separan por su localización celular inferida a partir de señales de clasificación en su secuencia. El color de cada punto representa la arquitectura modular de cada secuencia de expansina.

a, expansinas monomodulares básicas conservadas en el género de *Ralstonia*. **b**, expansinas multimodulares básicas (GH5-Expansina) conservadas dentro del género de *Xanthomonas*.

IV.2 Expansinas en Firmicutes

La expansina de *B. subtilis* (BsEXLX1), que quizás sea la más estudiada, es una expansina monomodular básica (Figura 1). BsEXLX1 pertenece al clúster de Firmicutes y es altamente similar (>60% de similitud) a algunas expansinas provenientes de gamma- y beta-Proteobacterias (Figura 6 y 7), lo que apoya la idea de que han sido intercambiadas horizontalmente como se ha sugerido anteriormente en otros estudios (Chase et al., 2020b; Cosgrove, 2017b; Nikolaidis et al., 2014a). También, es posible que las expansinas requieran una naturaleza básica (Figura 9A) para realizar su función biológica dentro del contexto fisiológico de estas bacterias. Curiosamente, el 75% de las secuencias de especies de Bacilliales contienen un motivo similar a LPXTG y una región hidrofóbica en la región C-terminal (Figura 9B). El primer motivo de *sorting* descubierto fue el LPXTG, pero desde entonces, se han encontrado otros motivos conservados y llamados motivos LPXTG-like (como los motivos LPNTA, [I/L][P/A]XTG o LAXTG).

Las expansinas de Firmicutes contienen un motivo LPXTG-like (Figura 9B), que aunque no fue reconocido por algunos programas bioinformáticos como CW-PRED, la similitud del motivo, la presencia de una "triple G" (LPXTGGG) característico de algunos motivos caracterizados, así como la presencia de una cola hidrofóbica y de un "ancla" cargada positivamente al final de la secuencia (Dai et al., 2019; Dramsi et al., 2008; Marraffini et al., 2006), sugieren una alta posibilidad de que pudieran ser reconocidas por una sortasa, aunque se requiere de la confirmación experimental.

Las bacterias pertenecientes al orden de las Clostridiales presentan complejos multienzimáticos especializados en degradar celulosa y otros polisacáridos de la

pared celular de las plantas, llamados celulosomas (Figura 9C). Estos están organizados mediante un tipo especial de subunidad llamada 'scaffoldina', que está compuesta por una serie de módulos de 'cohesinas'. Cada módulo de cohesina interactúa de manera selectiva con un tipo complementario de dominio, llamado dockerina. Las proteínas y enzimas que forman parte de los celulosomas contienen un dominio dockerina, para poder asociarse al andamiaje de las scaffoldinas.

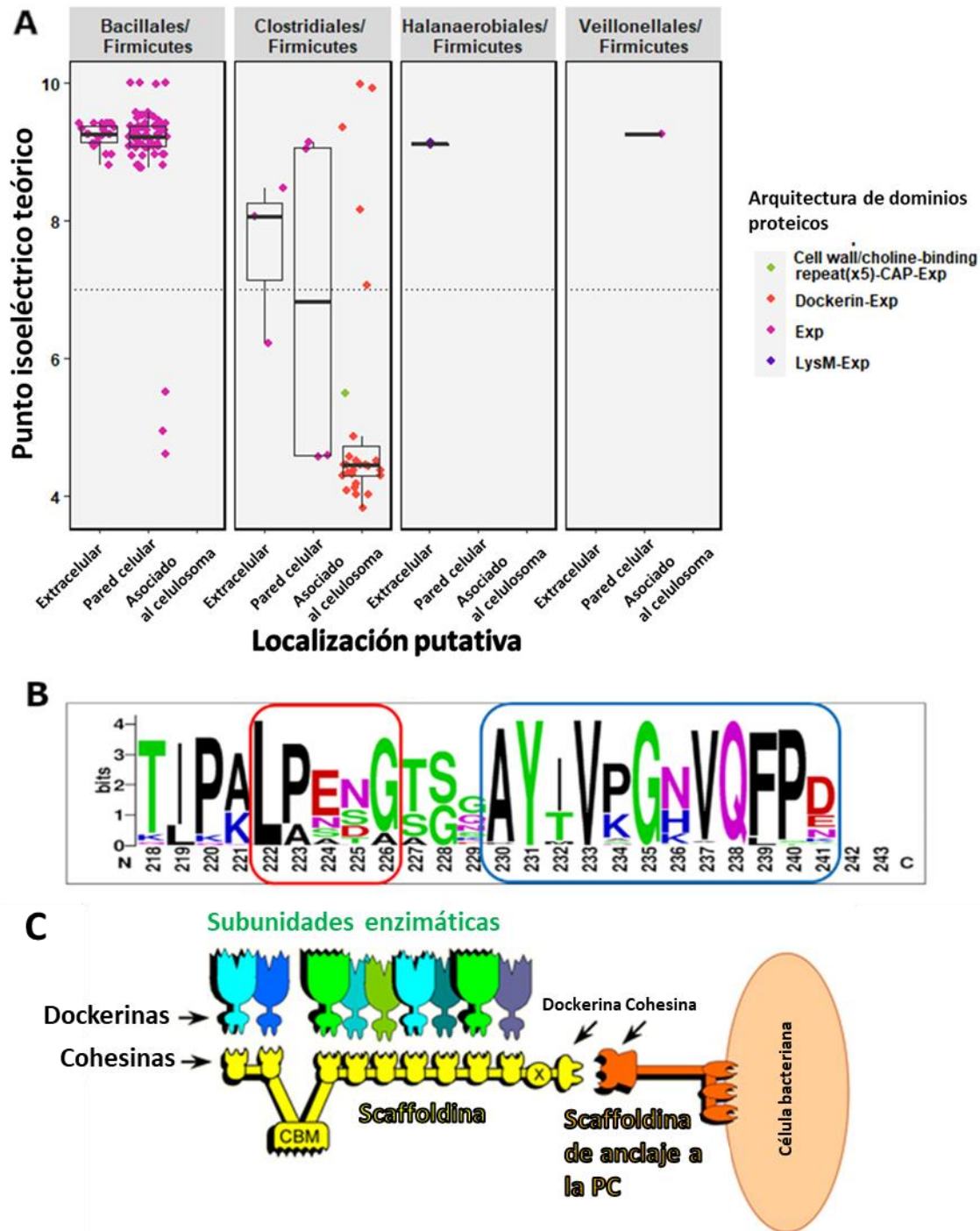


Figura 9. Expansinas provenientes de bacterias del filo Firmicutes. **A**, cada punto representa una secuencia de expansina. Utilizando el punto isoelectrico teórico, graficado en el eje y, se clasifican las expansinas como básicas ($\text{pH} > 7$) o ácidas ($\text{pH} < 7$). En el eje x, se separan por su localización celular inferida a partir de señales de clasificación en su secuencia. El color de cada punto representa la arquitectura modular de cada secuencia

de expansina. **B**, Logo del alineamiento de la región C-terminal de 96 expansinas en Baciliales, resaltando el posible motivo LPXTG-like y su respectiva ancla transmembranal. **C**, diagrama representativo de los complejos proteicos llamados celulosomas, las subunidades enzimáticas contienen dominios dockerina que se ensamblan con las cohesinas presentes en las scaffoldinas (Adaptado de Artzi, et al. 2017). La mayoría de las expansinas en el Orden de Clostridiales contienen un dominio dockerina.

El 73% de las secuencias de expansina provenientes de Clostridiales están fusionadas a un dominio dockerina. De hecho, la expansina dockerina-Exl1 de *Clostridium clariflavum* (CcExl1) interactúa con cohesinas del Tipo I que son parte de diferentes proteínas de andamiaje. En Artzi et al., 2016, se analizó la afinidad del dominio dockerina de CcExl1 con diferentes módulos de cohesina, demostrando que la interacción más fuerte fue con la cohesina de la scaffoldina monovalente ScaG, en comparación con ScaA y ScaD (celulosoma extracelular y celulosoma asociado a la pared celular, respectivamente). Podemos suponer que las expansinas/dockerina tienen un mecanismo de acción diferente en comparación con otras bacterias con diferentes estilos de vida, como fitopatógenos o habitantes del suelo. Por lo tanto, si el tipo de carga eléctrica de las expansinas es importante para su función biológica en diferentes especies, los resultados encontrados para Clostridiales pueden explicar la diferencia en el punto isoeléctrico de sus expansinas con respecto a las de otras bacterias del mismo filo (Figuras 9A). Vale la pena señalar que la scaffoldina ScaG tiene un módulo de unión a la superficie celular y se encuentra anclado a la célula. Asimismo, encontramos una alta similitud de secuencia entre las expansinas de Clostridiales y otras del filo Firmicutes (Figura 6A), algunas incluso con un posible motivo LPXTG-like.

IV.3 Expansinas en Actinobacterias

Las expansinas en Actinobacteria se agrupan en dos grupos separados: el primero está formado por Streptomycetales, Pseudonocardiales y Propionobacteriales y el segundo comprende los órdenes Micromonosporales,

Micrococcales, Corynebacteriales, Cryptosporangiales y Frankiales (Figura 6B, Figura 7). La mayoría de las expansinas de Actinobacteria (75%) son básicas, monomodulares y carecen de evidencia derivada de la secuencia que apunte a su asociación a la superficie celular.

Sin embargo, encontramos que una porción de expansinas en Corynebacteriales (26%), Micromonosporales (10%), Micrococcales (39%), Pseudonocardiales (54%) y Streptomycetales (18%) contienen motivos de sortasa putativos. Además, los órdenes de Frankiales (84%), Micromonosporales (4,4%) y Streptomycetales (4%) contienen expansinas con péptidos señal para localización en la membrana.

Es interesante señalar que las Corynebacteriales, que son similares a las bacterias didermas debido a una segunda membrana celular compuesta de ácidos micólicos, contienen expansinas ácidas (Figura 10). Este hallazgo es consistente con la correlación de las bacterias didermas que contienen expansinas cargadas negativamente. Mientras que la estructura de la superficie celular de las monodermas favorece a las expansinas cargadas positivamente. Sin embargo, hay algunas especies que contienen una expansina cargada negativamente, como es el caso de las expansinas de *Streptomyces acidiscabies* y *Kutzneria sp.* 744. De manera similar, las expansinas del orden Cryptosporangiales también son proteínas ácidas, pero se predice que son proteínas transmembranales, ya que están fusionadas con un dominio transmembranal N-terminal. También presentan una región desordenada C-terminal que se predice que es extracelular (Figura 11), lo sugiere que podría unirse a otra proteína, ya que se predice una región de unión (Figura 11A).

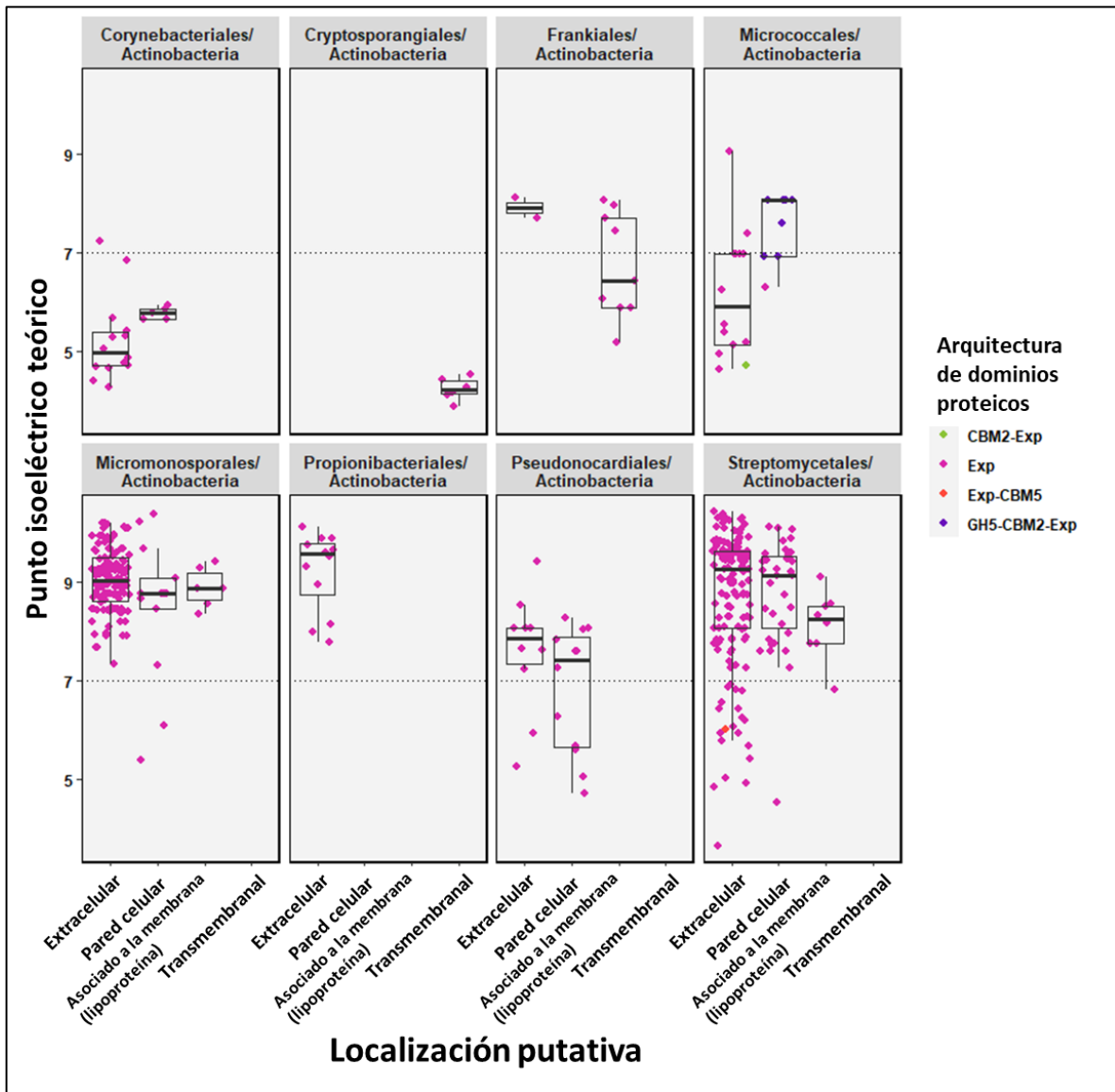


Figura 10. Expansinas provenientes de bacterias del filo Actinobacteria. A, cada punto representa una secuencia de expansina. Utilizando el punto isoeléctrico teórico, graficado en el eje y, se clasifican las expansinas como básicas (> pH 7) o ácidas (< pH 7). En el eje x, se separan por su localización celular inferida a partir de señales de clasificación en su secuencia. El color de cada punto representa la arquitectura modular de cada secuencia de expansina.

Otra posibilidad es que el dominio desordenado exista para estabilizar a la proteína, ya que se ha reportado que las regiones desordenadas en proteínas transmembranales proporcionan flexibilidad estructural para estabilizar las

interacciones con las cabezas cargadas negativamente de las moléculas de fosfolípidos (Tusnády et al., 2015), lo que parece plausible considerando que estas son expansinas cargadas negativamente. Las especies de *Clavibacter* del Orden Micrococcales contienen expansinas multimodulares GH5-CBM2 (Figura 10) y han sido objeto de investigación por su participación en la patogénesis de *Clavibacter michiganensis*, y por un posible aumento sinérgico de la degradación de celulosa por celulasas comerciales, donde muestran un efecto sinérgico con la celulasa de *Trichoderma reesei* en la hidrólisis de polisacáridos (Georgelis et al., 2015).

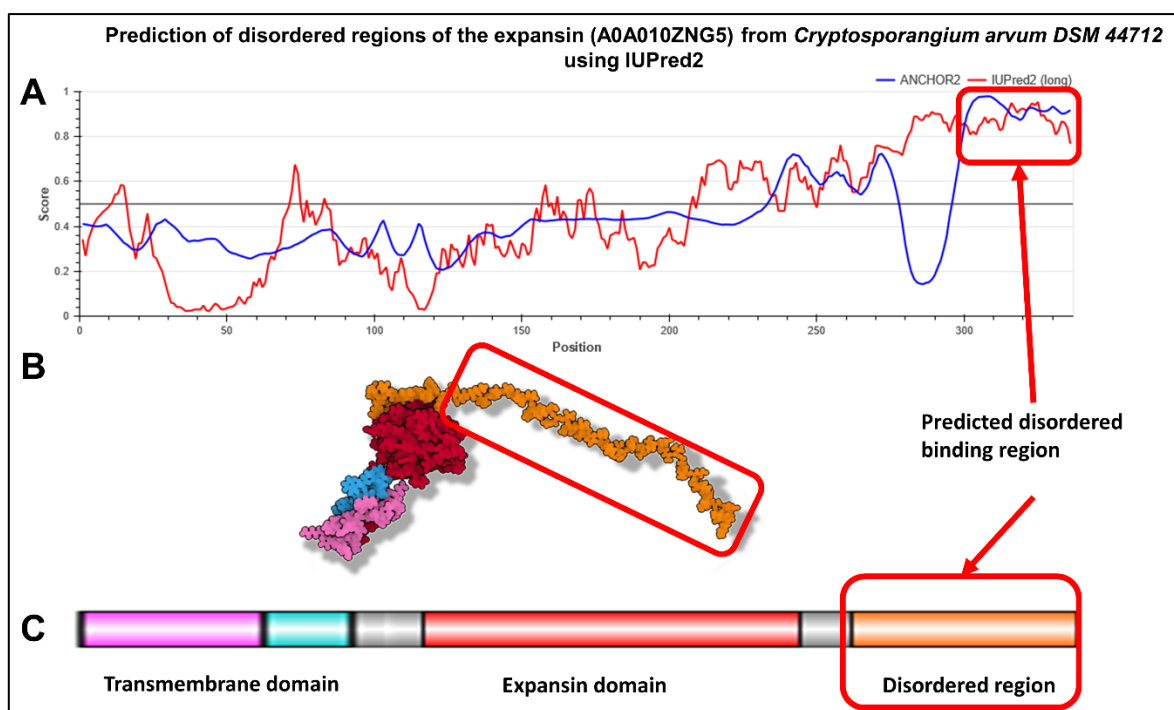


Figura 11. Predicción de regiones desordenadas en la expansina de *Cryptosporangium arzum* (UniProt id: A0A010ZNG5). **A**, grafica del puntaje de probabilidad de que la región sea desordenada (rojo, IUPred2), y de la posibilidad de que algún ligando externo pueda tener interacción con la región desordenada (azul, ANCHOR2). **B**, predicción de la estructura de esta expansina. **C**, arquitectura modular de la expansina. Recuadros rojos, región desordenada.

Streptomycetales y Micromonosporales son los órdenes con más expansinas (198 y 134 expansinas, respectivamente), con el 99,5% siendo expansinas monomodulares y el 86% predicho como expansinas básicas. La abundancia

relativa de expansinas en estos órdenes podría deberse a la sobrerrepresentación de cepas en las bases de datos de secuencias (Doroghazi and Metcalf, 2013), ya que son importantes productores de metabolitos secundarios, pero es interesante que especies fitopatógenas como *Streptomyces scabiei*, *Streptomyces stelliscabiei*, *Streptomyces turgidiscabies* y *Streptomyces ipomoeae* contengan genes de expansina en sus genomas.

IV.4 Variedad de arquitecturas en las expansinas multimodulares.

Para entender la influencia de los módulos adicionales en la evolución de las expansinas, también buscamos relaciones entre el orden bacteriano, la carga eléctrica de la proteína y la diversidad de arquitecturas modulares (Figura 12). La tendencia general de las expansinas ácidas en bacterias didermas difiere en algunos casos, como (a) las expansinas básicas modulares del orden didermo Xanthomonadales (Figura 8, asterisco y Figura 12C), (b) las expansinas ácidas de los dockerina-expansinas del celulosoma de las monodermas Clostridiales (Figura 9), (c) las expansinas monomodulares básicas en especies de *Ralstonia* (Figura 8, subíndice a), y (d) las expansinas ácidas de Corynebacteriales (Figura 10). Aunque los datos experimentales son inconclusos para la mejora enzimática significativa entre las expansinas y las enzimas degradadoras de polisacáridos, las fusiones entre expansinas y dominios relacionados con el metabolismo de los polisacáridos son comunes en las bacterias, y se conservan dentro de los grupos donde aparecen (Figura 7).

De hecho, las expansinas modulares más abundantes son: las fusionadas al dominio de la familia 5 de hidrolasas glicosil-GH5 de Xanthomonadales (GH5-Exp, Figura 12C); al dominio de unión a carbohidratos expansina en el extremo N o C, en varios órdenes (Figura 12C-D); al dominio expansina-dockerina en Clostridiales (Figura 12F); y las expansinas con ambos dominios GH5 y CMB2 del género *Clavibacter* (Figura 12I), que se relacionan directamente con el metabolismo de la celulosa, como se ha informado anteriormente.

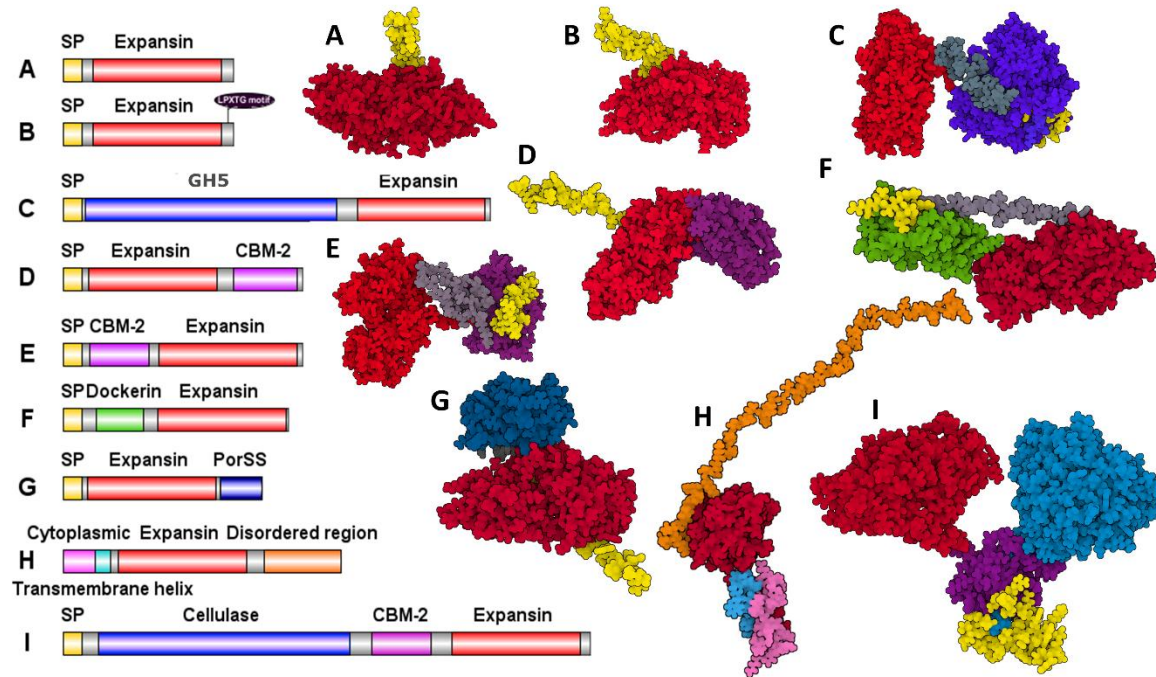


Figura 12. Ejemplos esquemáticos de estructuras primarias y terciarias de las arquitecturas modulares de las expansinas multimodulares más abundantes (N- a C-terminal). **A**, expansina monomodular (UniProt Id: C6DHV9). **B**, expansina monomodular con un motivo de clasificación similar a LPXTG (UniProt Id: D3JXR9; PDB: 3D30). **C**, expansina modular con un dominio de celulasa N-terminal (UniProt Id: Q8P513). **D**, expansina modular con un dominio CBM2 C-terminal (UniProt Id: F0Q783). **E**, expansina modular con un dominio CBM2 N-terminal (UniProt Id: A9FSQ7). **F**, expansina modular con un dominio dockerina N-terminal (UniProt Id: G8LUU0). **G**, expansina con un dominio T9SS (sistema de secreción tipo IX) C-terminal (UniProt Id: A0A2N3TEV7). **H**, expansina con dominio citoplásmico N-terminal, una hélice transmembranal y una región desordenada altamente polar C-terminal (UniProt Id: A0A010ZNG5). **I**, expansina modular con dominio GH5 N-terminal y un dominio CBM2 (UniProt Id: A5CLK3). Las estructuras terciarias se predijeron con trRosetta. El péptido señal (en amarillo) se escinde tras la translocación de la membrana, pero se incluyó en la predicción de la estructura. CBM2, módulo de unión a carbohidratos de la familia 2; GH5, glicosil hidrolasa de la familia 5.

Complementando los hallazgos de Chase et al., 2020a nuestros resultados muestran otras combinaciones de dominios relacionados con el metabolismo de

los polisacáridos (GH9, CBM5, CBM6) y similares a los dominios de unión a galactosa, fusionados con expansinas que se encuentran con menos frecuencia, que posiblemente hayan ocurrido recientemente, o que la ventaja evolutiva que confieren a cada especie es mínima en comparación con las expansinas modulares más conservadas. Es interesante que a pesar de encontrar solo algunas secuencias en Cyanobacteria y Verrucomicrobia, la mayoría fueron modulares, con arquitecturas de dominio raras o incluso únicas: las expansinas en Verrucomicrobia que se fusionaron con un dominio similar al dominio de unión a galactosa; el GH9-CBM2-expansina en Nostocales; una fusión de dominio doble CBM5 (CBM5-CBM5-expansina) en Pleurocapsales; una arquitectura de dominio única con tres repeticiones de CBM5 y dos dominios expansina adyacentes (expansina-CBM5 [x3]-expansina) en una especie de Flavobacteria; y una arquitectura de dominio única compuesta por una expansina fusionada a varios dominios de unión a la pared celular en una especie de *Ruminococcus*. En conjunto, nuestros resultados sugieren roles específicos para las expansinas en diferentes especies, donde podrían estar influyendo en la fisiología única de cada bacteria, lo que a su vez explicaría la gran diversidad de expansinas existentes.

Capítulo V. Discusión, conclusiones y perspectivas.

V.1 Discusión

Nuestro análisis actualizado de secuencias de expansinas bacterianas sugiere que la evolución de las expansinas podría ser impulsada por las características del tipo de envoltura celular bacteriana y por la fusión de dominios, lo que puede ofrecer diferentes funciones u objetivos según el nicho ecológico y la fisiología de cada especie. La información derivada de las secuencias de péptidos señal y los motivos de localización de las expansinas, que han sido pasados por alto en otros estudios (Chase et al., 2020a; Lohoff et al., 2020), fueron muy informativos ya que arrojaron luz sobre las vías de exportación de las expansinas entre las bacterias. Es coherente esperar que las expansinas sean proteínas exportadas al exterior si pensamos en que estas proteínas han sido mayormente estudiadas en fitopatógenos, y debido a su homología con las expansinas de las plantas de las que se sabe que su sustrato es algún componente de la pared celular vegetal; además se ha demostrado experimentalmente que las expansinas bacterianas tienen afinidad por tejidos vegetales y celulosa. Sin embargo, la diversidad de bacterias que codifican expansinas es mucho más que solo fitopatógenos o especies que tienen relación con las plantas, por lo que el sustrato de la mayoría de las especies todavía se desconoce, y por lo tanto tampoco se sabe dónde está localizado.

Nuestros datos indican que las expansinas endógenas bacterianas podrían no separarse de la superficie de la célula después de la secreción y en su lugar permanecer asociadas con las bacterias a través de diferentes mecanismos (Figura 13): (a) mediante el procesamiento que convierte a las expansinas en lipoproteínas, (b) por la presencia de dominios de clasificación en Fibrobacteres y Bacteroidetes que predicen la unión de las expansinas a la superficie celular a través del T9SS mediante la unión covalente de un A-LPS a la proteína; (c) la presencia de motivos similares a LPXTG en la región C-terminal y una cola

hidrofóbica en la región C-terminal que son sustratos potenciales para sortasas que unirían a las proteínas al peptidoglicano en 168 expansinas monodermas; y (d) por fusión de las expansinas a dominios que las anclan a complejos multiproteicos más grandes, como a la scaffoldina ScaG. Todos los mecanismos combinados predicen que hasta el 39% de las expansinas se localizan en la superficie celular después de la translocación de la membrana, lo que sugiere que su función es necesaria en la superficie o en la proximidad de la célula.

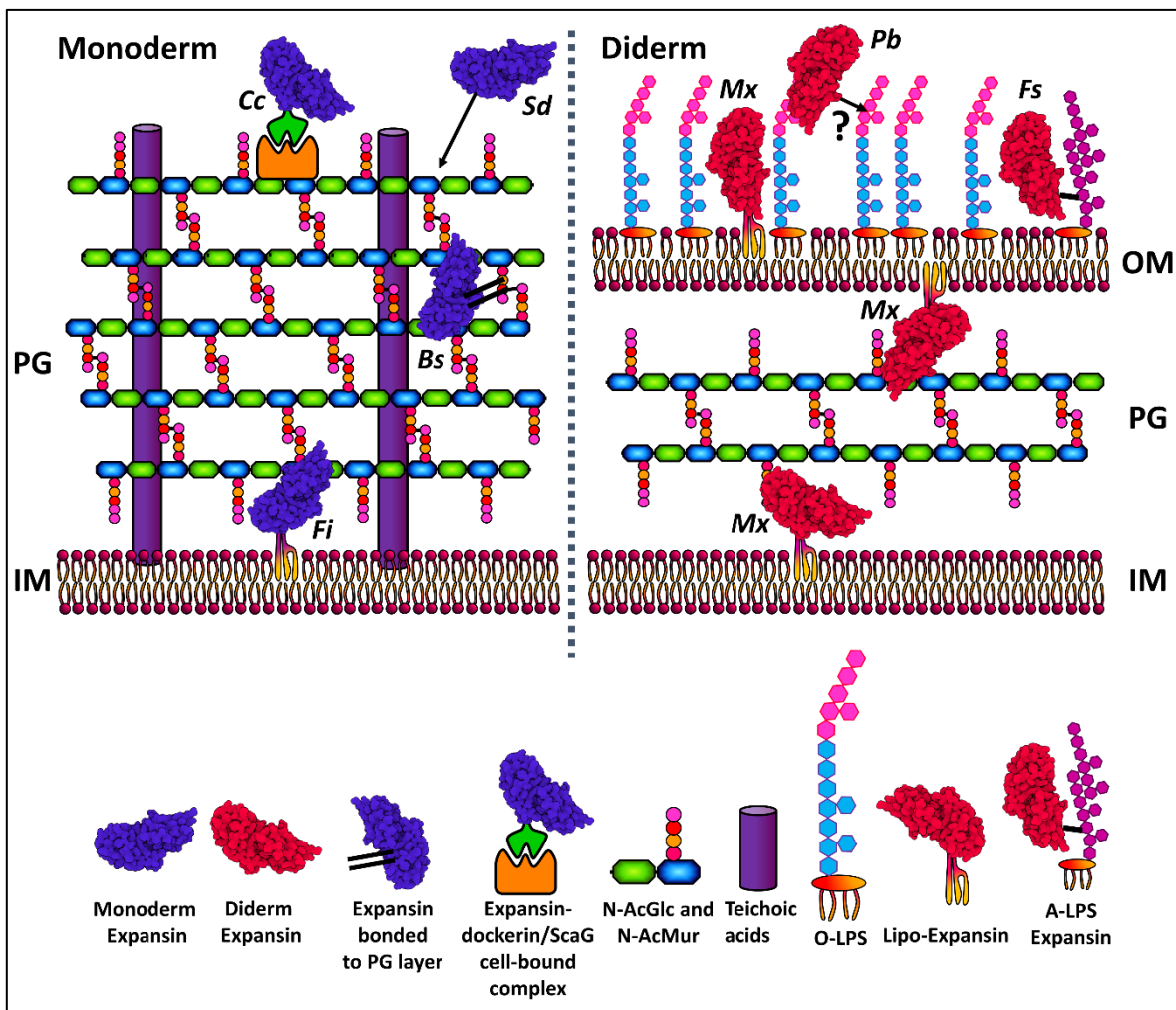


Figura 13. Representación esquemática de diversas estrategias para asociar expansinas a la superficie celular que utilizan las bacterias. Expansinas representativas de: **Bs**, *B. subtilis*; **Pb**, *Pectobacterium brasiliense* (cuyas propiedades de unión y su ligando aún se desconoce); **Fs**, *Fibrobacter succinogenes* (el T9SS une un A-LPS a la expansina); **Mx**, *Myxococcus xanthus* (las lipo-expansinas podrían estarse adhiriendo a cualquiera de las

membranas de las didermas); **Cc**, *Clostridium clariflavum* (asociada con la scaffoldina asociada a la envoltura celular ScaG); **Sd**, *Streptomyces davaonensis* (las expansinas tienen gran afinidad por el peptidoglicano); **Fi**, *Frankia ineficax*. OM, membrana externa. PG, pared de peptidoglicano. IM, membrana interna.

Es importante destacar que estas predicciones e interpretaciones deben investigarse experimentalmente, ya que, la divergencia de secuencias entre bacterias relacionadas y la falta de una mejor comprensión de todas los péptidos señal y señales de clasificación podrían oscurecer la predicción precisa de secuencias o dar falsos positivos debido a que todos los programas disponibles actualmente están entrenados con un subconjunto de organismos modelo. Para las secuencias sin información de localización en la superficie, nuestras observaciones anteriores muestran que *PbExl1* se retiene en el sedimento de los cultivos bacterianos en medio LB. Es decir, Exl1 podría estar uniéndose a un exopolisacárido, un lipopolisacárido u otro ligando en la envoltura celular desconocido a través de su superficie de unión a polisacáridos o mediante interacciones electrostáticas. Esto aún debe confirmarse, pero también abre la posibilidad de que la localización de las expansinas en la superficie celular sea un fenómeno más frecuente e inherente a las características de la proteína.

V.2 Conclusión

La función que desempeñan las expansinas en la envoltura celular aún no se ha respondido, pero proponemos dos posibilidades: una primera en la que se requiere la actividad de las expansinas para remodelar, estabilizar la pared celular bacteriana de algunas especies en un proceso similar o equivalente a las expansinas de las plantas. O una segunda, en la que la densidad de las expansinas contribuye a la modulación de las propiedades electrostáticas o hidrodinámicas de la célula debido a la naturaleza altamente cargada e hidrofílica de las expansinas. Las propiedades eléctricas del exterior de la célula podrían afectar los procesos bacterianos o tener implicaciones para las bacterias que interactúan con las plantas en el caso de patógenos o especies colonizadoras de

plantas como *B. subtilis*. También para las bacterias de vida libre, como los Myxococcales que exhiben un comportamiento multicelular dependiente de la producción de exopolisacáridos (Wu et al., 2007). Asimismo, dependiendo del grupo taxonómico, las expansinas podrían desempeñar diferentes roles combinados: actuar en su propia pared celular, actuar en la pared celular de la planta, formar parte del celulosoma para la digestión de biomasa vegetal, o contribuir a la modulación de las propiedades eléctricas de la envoltura celular bacteriana.

V.3 Perspectivas

Estudiar expansinas con diferentes características, pero provenientes de bacterias estrechamente relacionadas (como aquellas de Burkholderiales, por ejemplo) podría ayudar a identificar las causas subyacentes de las preferencias por tipos específicos de expansinas; o la alta similitud de secuencia entre expansinas de Proteobacteria y Firmicutes (atribuida a eventos de transferencia horizontal de genes (Chase et al., 2020a; Nikolaidis et al., 2014a)), lo que podría proporcionar una ventaja a las especies receptoras (Rocha et al., 2020). Se debe determinar si más de una función afecta la fisiología bacteriana, ya que las mencionadas no son mutuamente exclusivas. Por ejemplo, hemos observado anteriormente que Exl1, una expansina ácida soluble de *P. brasiliense* y EXLX1, una expansina básica soluble de *B. subtilis*, tienen patrones de unión distintivos en las paredes celulares de apio (Tovar-Herrera et al., 2018b), donde Exl1 se unió a las esquinas de la pared celular que rodean los vasos del xilema, abriendo la posibilidad de que la localización de las bacterias sea impulsada por la expresión de expansinas, es decir, que la afinidad de las expansinas por un ligando del espacio intracelular de los tejidos vasculares determine la localización bacteriana, lo que hasta ahora sigue siendo una pregunta abierta. Esto también explicaría la recurrencia de las expansinas en los fitopatógenos invasores del xilema.

Además, en trabajos anteriores de nuestro laboratorio (Narváez-Barragán et al., 2020), se estudió a una cepa mutante nula *exl1* de *Pectobacterium atrosepticum* menos infectiva y se encontró que presentaba un defecto en la motilidad por *swarming* (enjambre, en español). El *swarming* es un proceso independiente de la asociación con la planta, es dependiente de una alta hidratación y, por lo tanto, es susceptible a la densidad de moléculas hidrofílicas, como las expansinas, en la superficie celular. En apoyo de esto, Valdivia y colaboradores (Valdivia et al., 2009) informaron que entre diferentes fenotipos de una mutante de expansina *expb1* del polen de maíz, ocurre un defecto de deshidratación después de la liberación de la antera. Estudiar más a fondo el papel de las expansinas de *Pectobacterium* en el *swarming* y en la capacidad de migrar dentro de la planta, podría dar información acerca de cómo las expansinas están participando en estos procesos. El único fenotipo atribuido a la mutante nula de expansina Δ EXLX1 de *B. subtilis* que es independiente de las plantas es la mayor susceptibilidad de las bacterias a la autólisis. Esto se debe posiblemente a la reducción de la estabilidad de la pared de peptidoglicano, a la que la expansina podría estar unida covalentemente por sortasas (Kerff et al., 2008). También encontramos que la mayoría (74%) de las expansinas pertenecientes a Myxococcales se predicen como lipo-expansinas mono-modulares y ácidas, que corresponde al número más alto de lipo-expansinas en todos los grupos (Figura 8), lo que sugiere que las expansinas de este Orden necesitan estar incrustadas en la membrana. Sería interesante estudiar el papel biológico de las lipo-expansinas en un contexto social bacteriano, como el de las especies de vida libre *Myxococcus xanthus* (lipo-expansina ácida) y *Sorangium cellulosum* (lipo-CBM2-expansina ácida), que exhiben un comportamiento multicelular. Esto para determinar si las expansinas influyen en la hidrodinámica de las colonias o en las interacciones célula-célula/célula-superficie, y si este efecto es el mismo durante la motilidad por *swarming* de *P. brasiliense*. Asimismo, se ha encontrado que unas proteínas relacionadas a las expansinas en las amebas del género *Dictyostelium* influyen en la formación del cuerpo fructífero (el cual es edificado por un comportamiento social), produciendo fenotipos aberrantes cuando no se expresa o cuando se

sobreexpresa (Ogasawara et al., 2009). Un caso interesante, son las lipo-expansinas de Frankiales, que son Actinobacterias fijadoras de nitrógeno y simbióticas de las plantas (Gtari et al., 2020). Estas expansinas podrían estar desempeñando un papel importante en la colonización de las plantas, similar a otras bacterias promotoras del crecimiento de plantas como *B. subtilis* (Kerff et al., 2008), por lo que estudiarlas brindaría información relevante sobre la función de las lipo-expansinas en bacterias monodermas.

Al analizar un gran número de secuencias de expansinas, encontramos una correlación entre la carga eléctrica de las expansinas y el tipo de bacteria de origen (expansinas ácidas en didermas y básicas en monodermas), aunque esta relación no es perfecta, lo que sugiere que también depende de otros factores. Curiosamente, las expansinas básicas pertenecientes a las especies Xanthomonadales (diderma) y al género *Clavibacter* (monoderma) están fusionadas a un dominio de glicosil hidrolasa (GH5) (Figura 12C y 12I, respectivamente), por lo que estas expansinas podrían requerir de una naturaleza básica para la funcionalidad de los dominios GH5, explicando la presencia de expansinas básicas en bacterias didermas. Estas expansinas han sido estudiadas por su capacidad de aumentar sinérgicamente la eficiencia enzimática de la hidrólisis de celulosa (Bunterngsook et al., 2014; Bunterngsook et al., 2015; Junior et al., 2015; Kim et al., 2009; Lee et al., 2013; Pech-Cervantes et al., 2019) pero su mecanismo de acción *in vivo* debe ser analizado bajo diferentes condiciones, ya que no está claro si las expansinas modulares mejoran la actividad de las enzimas adyacentes o si proporcionan otra funcionalidad desconocida, como la selectividad del sustrato, alineación del sustrato, o adhesión celular al sustrato como en el caso de las dockerina-expansinas de especies de *Clostridium* (Artzi et al., 2016).

El análisis de la actividad de expansinas monomodulares y multimodulares proporcionaría información adicional si se estudia bajo condiciones naturales y en el momento en que la actividad fuera necesaria en cada especie, ya que es posible que el requerimiento de expansinas ocurra en etapas de desarrollo

específicas, por lo que los fenotipos de mutantes nulos de expansinas podrían ser críticos y específicos de la especie. Por lo tanto, es necesario desarrollar nuevos procedimientos experimentales para la investigación de expansinas bacterianas considerando que podrían estar asociadas con el exterior de la bacteria. Por ejemplo, analizar cómo *Xanthomonas* o *Clavibacter* utilizan diferentes polisacáridos como única fuente de carbono, en una comparación de mutantes nulos de expansinas versus *wild-type*, podría indicar si las expansinas modulares que contienen un dominio de celulasa ayudan a las bacterias a metabolizar carbohidratos complejos de manera más eficiente. Al mismo tiempo, la determinación del mecanismo en fenotipos no relacionados con la infección de plantas de mutantes nulos y cepas que sobreexpresan expansinas como la susceptibilidad a la autólisis en *B. subtilis* (Kerff et al., 2008) y la motilidad de enjambre de *P. atrosepticum* y *P. carotovorum* (Narváez-Barragán et al., 2020), podrían arrojar luz sobre la función de las expansinas en estos microorganismos. Se necesita un estudio experimental integral que corrobore nuestros hallazgos, teniendo en cuenta todos los tipos de expansinas bacterianas, las señales de localización, y la composición de la envoltura celular en diferentes bacterias.

VI. Referencias

- Abby, S.S. & Rocha, E.P.C. (2017) Identification of protein secretion systems in bacterial genomes using MacSyFinder. *Methods in Molecular Biology*, 1615, 1-21. https://doi.org/10.1007/978-1-4939-7033-9_1.
- Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37, 420-423. <https://doi.org/10.1038/s41587-019-0036-z>.
- Alvarez-Martinez, C.E., Sgro, G.G., Araujo, G.G., Paiva, M.R.N., Matsuyama, B.Y., Guzzo, C.R., et al. (2021) Secrete or perish: The role of secretion systems in Xanthomonas biology. *Computational and Structural Biotechnology Journal*, 19, 279-302. <https://doi.org/10.1016/j.csbj.2020.12.020>.
- Armijos-Jaramillo, V., Santander-Gordón, D., Tejera, E. & Perez-Castillo, Y. (2018) The dilemma of bacterial expansins evolution. The unusual case of Streptomyces acidiscabies and Kutzneria sp. 744. *Communicative and Integrative Biology*, 11, 1-10. <https://doi.org/10.1080/19420889.2018.1539612>.
- Artzi, L., Morag, E., Barak, Y., Lamed, R. & Bayer, E.A. (2015) Clostridium clariflavum: Key cellulosome players are revealed by proteomic analysis. *mBio*, 6, 1-12. <https://doi.org/10.1128/mBio.00411-15>.
- Artzi, L., Morag, E., Shamsoum, M. & Bayer, E.A. (2016) Cellulosomal expansin: Functionality and incorporation into the complex. *Biotechnology for Biofuels*, 9. <https://doi.org/10.1186/s13068-016-0474-5>.
- Asmar, A.T. & Collet, J.F. (2018) Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. *FEMS Microbiology Letters*, 365, 1-8. <https://doi.org/10.1093/femsle/fny199>.
- Asmar, A.T., Ferreira, J.L., Cohen, E.J., Cho, S.H., Beeby, M., Hughes, K.T., et al. (2017) Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS Biology*, 15, 1-16. <https://doi.org/10.1371/journal.pbio.2004303>.
- Bashline, L., Lei, L., Li, S. & Gu, Y. (2014) Cell Wall , Cytoskeleton , and Cell Expansion in Higher Plants CELLULOSE AND NON-CELLULOSIC. *Molecular Plant*, 7, 586-600. <https://doi.org/10.1093/mp/ssu018>.
- Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021) UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49, D480-D489. <https://doi.org/10.1093/nar/gkaa1100>.
- Bishop, R.E., Leskiw, B.K., Hodges, R.S., Kay, C.M. & Weiner, J.H. (1998) The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death. *Journal of Molecular Biology*, 280, 583-596. <https://doi.org/10.1006/jmbi.1998.1894>.
- Bogino, P.C., Oliva, M. de las M., Sorroche, F.G. & Giordano, W. (2013) The role of bacterial biofilms and surface components in plant-bacterial associations. *International Journal of Molecular Sciences*, 14, 15838-15859. <https://doi.org/10.3390/ijms140815838>.
- Braun, V. & Rehn, K. (1969) Chemical Characterization, Spatial Distribution and Function of a Lipoprotein (Murein-Lipoprotein) of the E. coli Cell Wall: The Specific Effect of Trypsin on the Membrane Structure. *European Journal of Biochemistry*, 10, 426-438. <https://doi.org/10.1111/j.1432-1033.1969.tb00707.x>.
- Budzik, J.M., Marraffini, L.A. & Schneewind, O. (2007) Assembly of pili on the surface of Bacillus cereus vegetative cells. *Molecular Microbiology*, 66, 495-510. <https://doi.org/10.1111/j.1365-2958.2007.05939.x>.
- Bunterngsook, B., Eurwilaichitr, L., Thamchaipenet, A. & Champreda, V. (2015) Binding characteristics and

- synergistic effects of bacterial expansins on cellulosic and hemicellulosic substrates. *Bioresource Technology*, 176, 129-135. <https://doi.org/10.1016/j.biortech.2014.11.042>.
- Bunterngrsook, B., Mhuanong, W., Champreda, V., Thamchaiphenet, A. & Eurwilaichitr, L. (2014) Identification of novel bacterial expansins and their synergistic actions on cellulose degradation. *Bioresource Technology*, 159, 64-71. <https://doi.org/10.1016/j.biortech.2014.02.004>.
- Cascioferro, S., Totsika, M. & Schillaci, D. (2014) Sortase A: An ideal target for anti-virulence drug development. *Microbial Pathogenesis*, 77, 105-112. <https://doi.org/10.1016/j.micpath.2014.10.007>.
- Chalupowicz, L., Barash, I., Reuven, M., Dror, O., Sharabani, G., Gartemann, K.H., et al. (2017) Differential contribution of *Clavibacter michiganensis* ssp. *michiganensis* virulence factors to systemic and local infection in tomato. *Molecular Plant Pathology*, 18, 336-346. <https://doi.org/10.1111/mpp.12400>.
- Chase, W.R., Zhaxybayeva, O., Rocha, J., Cosgrove, D.J. & Shapiro, L.R. (2020a) Global cellulose biomass, horizontal gene transfers and domain fusions drive microbial expansin evolution. *New Phytologist*, 226, 921-938. <https://doi.org/10.1111/nph.16428>.
- Chase, W.R., Zhaxybayeva, O., Rocha, J., Cosgrove, D.J. & Shapiro, L.R. (2020b) Global cellulose biomass, horizontal gene transfers and domain fusions drive microbial expansin evolution. *New Phytologist*, 226, 921-938. <https://doi.org/10.1111/nph.16428>.
- Chatzi, K.E., Sardis, M.F., Economou, A. & Karamanou, S. (2014) SecA-mediated targeting and translocation of secretory proteins. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843, 1466-1474. <https://doi.org/10.1016/j.bbamcr.2014.02.014>.
- Chebli, Y. & Geitmann, A. (2017) Cellular growth in plants requires regulation of cell wall biochemistry. *Current Opinion in Cell Biology*, 44, 28-35. <https://doi.org/10.1016/j.ceb.2017.01.002>.
- Chen, C., Cui, Z., Song, X., Liu, Y.J., Cui, Q. & Feng, Y. (2016) Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation. *Applied Microbiology and Biotechnology*, 100, 2203-2212. <https://doi.org/10.1007/s00253-015-7071-6>.
- Choi, D.S., Yamada, H., Mizuno, T. & Mizushima, S. (1986) Trimeric structure and localization of the major lipoprotein in the cell surface of *Escherichia coli*. *Journal of Biological Chemistry*, 261, 8953-8957. [https://doi.org/10.1016/s0021-9258\(19\)84474-5](https://doi.org/10.1016/s0021-9258(19)84474-5).
- Cole, G.B., Bateman, T.J. & Moraes, T.F. (2021) The surface lipoproteins of gram-negative bacteria: Protectors and foragers in harsh environments. *Journal of Biological Chemistry*, 296, 100147. <https://doi.org/10.1074/jbc.REV120.008745>.
- Comfort, D. & Clubb, R.T. (2004) A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infection and Immunity*, 72, 2710-2722. <https://doi.org/10.1128/IAI.72.5.2710-2722.2004>.
- Cosgrove, D.J. (2016) Catalysts of plant cell wall loosening. *F1000Research*, 5, 119. <https://doi.org/10.12688/f1000research.7180.1>.
- Cosgrove, D.J. (2017a) Microbial Expansins. *Annual Review of Microbiology*, 71, 479-497. <https://doi.org/10.1146/annurev-micro-090816-093315>.
- Cosgrove, D.J. (2017b) Microbial Expansins. *Annual Review of Microbiology*, 71, 479-497. <https://doi.org/10.1146/annurev-micro-090816-093315>.
- Cosgrove, D.J. (2015) Plant expansins: Diversity and interactions with plant cell walls. *Current Opinion in Plant*

- Biology*, 25, 162-172. <https://doi.org/10.1016/j.pbi.2015.05.014>.
- Cossart, P. & Jonquieres, R. (2000) Sortase, a universal target for therapeutic agents against Gram-positive bacteria? *Proceedings of the National Academy of Sciences of the United States of America*, 97, 5013-5015. <https://doi.org/10.1073/pnas.97.10.5013>.
- Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., et al. (2015) Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nature Reviews Microbiology*, 13, 343-359. <https://doi.org/10.1038/nrmicro3456>.
- Cowles, C.E., Li, Y., Semmelhack, M.F., Cristea, I.M. & Silhavy, T.J. (2011) The free and bound forms of Lpp occupy distinct subcellular locations in *Escherichia coli*. *Molecular Microbiology*, 79, 1168-1181. <https://doi.org/10.1111/j.1365-2958.2011.07539.x>.
- Dai, X., Böker, A. & Glebe, U. (2019) Broadening the scope of sortagging. *RSC Advances*, 9, 4700-4721. <https://doi.org/10.1039/c8ra06705h>.
- Dalbey, R.E., Wang, P. & Dijk, J.M. van (2012) Membrane Proteases in the Bacterial Protein Secretion and Quality Control Pathway. *Microbiology and Molecular Biology Reviews*, 76, 311-330. <https://doi.org/10.1128/mubr.05019-11>.
- Desvaux, M., Candela, T. & Serron, P. (2018) Surfaceome and proteosurfaceome in parietal monoderm bacteria: Focus on protein cell-surface display. *Frontiers in Microbiology*, 9, 1-28. <https://doi.org/10.3389/fmicb.2018.00100>.
- Desvaux, M., Dumas, E., Chafsey, I. & Hébraud, M. (2006) Protein cell surface display in Gram-positive bacteria: From single protein to macromolecular protein structure. *FEMS Microbiology Letters*, 256, 1-15. <https://doi.org/10.1111/j.1574-6968.2006.00122.x>.
- Diego, I. De, Ksiazek, M., Mizgalska, D., Koneru, L., Golik, P., Szmigielski, B., et al. (2016) The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal β -sandwich domain. *Scientific Reports*, 6, 1-17. <https://doi.org/10.1038/srep23123>.
- Doroghazi, J.R. & Metcalf, W.W. (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics*, 14, 611. <https://doi.org/10.1186/1471-2164-14-611>.
- Douillard, F.P., Rasinkangas, P., Ossowski, I. Von, Reunanen, J., Palva, A. & Vos, W.M. De (2014) Functional identification of conserved residues involved in *Lactobacillus rhamnosus* strain GG sortase specificity and pilus biogenesis. *Journal of Biological Chemistry*, 289, 15764-15775. <https://doi.org/10.1074/jbc.M113.542332>.
- Dramsı, S., Magnet, S., Davison, S. & Arthur, M. (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiology Reviews*, 32, 307-320. <https://doi.org/10.1111/j.1574-6976.2008.00102.x>.
- Dubuisson, J.F., Vianney, A., Hugouvieux-Cotte-Pattat, N. & Lazzaroni, J.C. (2005) Tol-Pal proteins are critical cell envelope components of *Erwinia chrysanthemi* affecting cell morphology and virulence. *Microbiology*, 151, 3337-3347. <https://doi.org/10.1099/mic.0.28237-0>.
- Dufresne, K. & Paradis-Bleau, C. (2015) Biology and assembly of the bacterial envelope.
- Eddy, S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome informatics. International Conference on Genome Informatics*, 23, 205-211.
- Egan, S.A., Kurian, D., Ward, P.N., Hunt, L. & Leigh, J.A. (2010) Identification of sortase A (SrtA) substrates in *Streptococcus uberis*: Evidence for an additional hexapeptide (LPXXXD) sorting motif. *Journal of*

- Proteome Research*, 9, 1088-1095. <https://doi.org/10.1021/pr901025w>.
- Fimereli, D.K., Tsigirgos, K.D., Litou, Z.I., Liakopoulos, T.D., Bagos, P.G. & Hamodrakas, S.J. (2012) Artificial Intelligence: Theories and Applications. CW-PRED: A HMM-Based Method for the Classification of Cell Wall-Anchored Proteins of Gram-Positive Bacteria. *Hellenic Conference on Artificial Intelligence*, 7297, 285-290.
- Frain, K.M., Robinson, C. & Dijk, J.M. van (2019) Transport of Folded Proteins by the Tat System. *Protein Journal*, 38, 377-388. <https://doi.org/10.1007/s10930-019-09859-y>.
- Fuangthong, M., Sallabhan, R., Atichartpongkul, S., Rangkadilok, N., Sriprang, R., Satayavivad, J., et al. (2008) The omlA gene is involved in multidrug resistance and its expression is inhibited by coumarins in *Xanthomonas campestris* pv. *phaseoli*. *Archives of Microbiology*, 189, 211-218. <https://doi.org/10.1007/s00203-007-0310-1>.
- Gartemann, K.H., Kirchner, O., Engemann, J., Gräfen, I., Eichenlaub, R. & Burger, A. (2003) *Clavibacter michiganensis* subsp. *michiganensis*: First steps in the understanding of virulence of a Gram-positive phytopathogenic bacterium. *Journal of Biotechnology*, 106, 179-191. <https://doi.org/10.1016/j.jbiotec.2003.07.011>.
- Georgelis, N., Nikolaidis, N. & Cosgrove, D.J. (2015) Bacterial expansins and related proteins from the world of microbes. *Applied Microbiology and Biotechnology*, 99, 3807-3823. <https://doi.org/10.1007/s00253-015-6534-0>.
- Georgelis, N., Nikolaidis, N. & Cosgrove, D.J. (2014) Biochemical analysis of expansin-like proteins from microbes. *Carbohydrate Polymers*, 100, 17-23. <https://doi.org/10.1016/j.carbpol.2013.04.094>.
- Georgelis, N., Tabuchi, A., Nikolaidis, N. & Cosgrove, D.J. (2011) Structure-function analysis of the bacterial expansin EXLX1. *Journal of Biological Chemistry*, 286, 16814-16823. <https://doi.org/10.1074/jbc.M111.225037>.
- Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., et al. (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1854, 1019-1037. <https://doi.org/10.1016/j.bbapap.2015.04.015>.
- Goosens, V.J., Monteferrante, C.G. & Dijk, J.M. Van (2014) The Tat system of Gram-positive bacteria. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843, 1698-1706. <https://doi.org/10.1016/j.bbamcr.2013.10.008>.
- Gorasia, D.G., Veith, P.D., Chen, D., Seers, C.A., Mitchell, H.A., Chen, Y.Y., et al. (2015) *Porphyromonas gingivalis* Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism. *PLoS Pathogens*, 11, 1-31. <https://doi.org/10.1371/journal.ppat.1005152>.
- Granato, L.M., Picchi, S.C., Oliveira Andrade, M. De, Martins, P.M.M., Takita, M.A., Machado, M.A., et al. (2019) The ECNA antitoxin is important not only for human pathogens: Evidence of Its Role in the Plant Pathogen *Xanthomonas citri* subsp. *Citri*. *Journal of Bacteriology*, 201. <https://doi.org/10.1128/JB.00796-18>.
- Grant, C.E., Bailey, T.L. & Noble, W.S. (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, 27, 1017-1018. <https://doi.org/10.1093/bioinformatics/btr064>.
- Green, E.R. & Meccas, J. (2016) Bacterial secretion systems: An overview. *Virulence Mechanisms of Bacterial*

- Pathogens*, 4, 213-239. <https://doi.org/10.1128/9781555819286.ch8>.
- Gtari, M., Ghodhbane-Gtari, F. & Nouioui, I. (2020) *Frankia soli* sp. nov., an actinobacterium isolated from soil beneath *Ceanothus jepsonii*. *International Journal of Systematic and Evolutionary Microbiology*, 70, 1203-1209. <https://doi.org/10.1099/ijsem.0.003899>.
- Heijne, G. Von (1989) The structure of signal peptides from bacterial lipoproteins. *Protein Engineering, Design and Selection*, 2, 531-534. <https://doi.org/10.1093/protein/2.7.531>.
- Hellman, J., Roberts, J.D., Tehan, M.M., Allaire, J.E. & Shaw Warren, H. (2002) Bacterial peptidoglycan-associated lipoprotein is released into the bloodstream in Gram-negative sepsis and causes inflammation and death in mice. *Journal of Biological Chemistry*, 277, 14274-14280. <https://doi.org/10.1074/jbc.M109696200>.
- Hendrickx, A.P.A., Budzik, J.M., Oh, S.Y. & Schneewind, O. (2011) Architects at the bacterial surface-sortases and the assembly of pili with isopeptide bonds. *Nature Reviews Microbiology*, 9, 166-176. <https://doi.org/10.1038/nrmicro2520>.
- Hendrickx, A.P.A., Willems, R.J.L., Bonten, M.J.M. & Schaik, W. van (2009) LPxTG surface proteins of enterococci. *Trends in Microbiology*, 17, 423-430. <https://doi.org/10.1016/j.tim.2009.06.004>.
- Hooda, Y. & Moraes, T.F. (2018) Translocation of lipoproteins to the surface of gram negative bacteria. *Current Opinion in Structural Biology*, 51, 73-79. <https://doi.org/10.1016/j.sbi.2018.03.006>.
- Jackson, R.W., Vinatzer, B.A. & Arnold, D. (2015) Bacteria-Plant Interactions: Advanced Research and Future Trends. *Bacteria-Plant Interactions: Advanced Research and Future Trends*, 2-5. <https://doi.org/10.21775/9781908230584>.
- Jin, K.M., Zhuo, R.Y., Xu, D., Wang, Y.J., Fan, H.J., Huang, B.Y., et al. (2020) Genome-wide identification of the expansin gene family and its potential association with drought stress in moso bamboo. *International Journal of Molecular Sciences*, 21, 1-21. <https://doi.org/10.3390/ijms21249491>.
- Jin, S., Joe, A., Lynett, J., Hani, E.K., Sherman, P. & Chan, V.L. (2001) JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Molecular Microbiology*, 39, 1225-1236. <https://doi.org/10.1046/j.1365-2958.2001.02294.x>.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30, 1236-1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Junior, A.T., Dolce, L.G., Oliveira Neto, M. de & Polikarpov, I. (2015) Xanthomonas campestris expansin-like X domain is a structurally disordered beta-sheet macromolecule capable of synergistically enhancing enzymatic efficiency of cellulose hydrolysis. *Biotechnology Letters*, 37, 2419-2426. <https://doi.org/10.1007/s10529-015-1927-9>.
- Kerff, F., Amoroso, A., Herman, R., Sauvage, E., Petrella, S., Filée, P., et al. (2008) Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 16876-16881. <https://doi.org/10.1073/pnas.0809382105>.
- Kim, E.S., Lee, H.J., Bang, W.G., Choi, I.G. & Kim, K.H. (2009) Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering*, 102, 1342-1353. <https://doi.org/10.1002/bit.22193>.

- Konovalova, A., Kahne, D.E. & Silhavy, T.J. (2017) Outer Membrane Biogenesis. *Annual Review of Microbiology*, 71, 539-556. <https://doi.org/10.1146/annurev-micro-090816-093754>.
- Kovacs-Simon, A., Titball, R.W. & Michell, S.L. (2011) Lipoproteins of bacterial pathogens. *Infection and Immunity*, 79, 548-561. <https://doi.org/10.1128/IAI.00682-10>.
- Kozlowski, L.P. (2016) IPC - Isoelectric Point Calculator. *Biology Direct*, 11, 1-16. <https://doi.org/10.1186/s13062-016-0159-9>.
- Kudryavtsev, K. V., Fedotcheva, T.A. & Shimanovsky, N.L. (2021) Inhibitors of Sortases of Gram-Positive Bacteria and their Role in the Treatment of Infectious Diseases (Review). *Pharmaceutical Chemistry Journal*, 55, 751-756. <https://doi.org/10.1007/s11094-021-02488-9>.
- Lauren E. Brooks,^a Sabah Ul-Hasan,^a Benjamin K. Chan, ^b M.J.S. (2018) Quantifying the evolutionary conservation of genes encoding multidrug efflux pumps in the ESKAPE pathogens to identify antimicrobial drug targets. *mSystems*, 3, e00024-18. <https://doi.org/10.1128/mSystems.00024-18>.
- Lazzaroni, J. -C & Portalier, R. (1992) The excC gene of Escherichia coli K-12 required for cell envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL). *Molecular Microbiology*, 6, 735-742. <https://doi.org/10.1111/j.1365-2958.1992.tb01523.x>.
- Lee, H.J., Kim, I.J., Kim, J.F., Choi, I.-G. & Kim, K.H. (2013) An expansin from the marine bacterium Hahella chejuensis acts synergistically with xylanase and enhances xylan hydrolysis. *Bioresource Technology*, 149, 516-519. <https://doi.org/10.1016/j.biortech.2013.09.086>.
- Lee, P.A., Tullman-Ercek, D. & Georgiou, G. (2006) The Bacterial Twin-Arginine Translocation Pathway. *Annual Review of Microbiology*, 60, 373-395. <https://doi.org/10.1146/annurev.micro.60.080805.142212>.
- Leigh, J.A., Egan, S.A., Ward, P.N., Field, T.R. & Coffey, T.J. (2010) Sortase anchored proteins of Streptococcus uberis play major roles in the pathogenesis of bovine mastitis in dairy cattle. *Veterinary Research*, 41. <https://doi.org/10.1051/vetres/2010036>.
- Letunic, I. & Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, 44, W242-W245. <https://doi.org/10.1093/nar/gkw290>.
- Li, X.Z., Nikaido, H. & Poole, K. (1995) Role of MexA-MexB-OprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 39, 1948-1953. <https://doi.org/10.1128/AAC.39.9.1948>.
- Li, Y. & Yu, Q. (2020) Role of the lipoprotein VacJ/MlaA in the physiology and virulence of the plant pathogen Xanthomonas citri subsp. citri. *Physiological and Molecular Plant Pathology*, 112, 1-6. <https://doi.org/10.1016/j.pmpp.2020.101546>.
- Liao, C.T., Li, C.E., Chang, H.C., Hsu, C.H., Chiang, Y.C. & Hsiao, Y.M. (2022) The lolB gene in Xanthomonas campestris pv. campestris is required for bacterial attachment, stress tolerance, and virulence. *BMC Microbiology*, 22, 1-13. <https://doi.org/10.1186/s12866-021-02416-7>.
- Liew, C.K., Smith, B.T., Pilpa, R., Suree, N., Ilangovan, U., Connolly, K.M., et al. (2004) Localization and mutagenesis of the sorting signal binding site on sortase A from Staphylococcus aureus. *FEBS Letters*, 571, 221-226. <https://doi.org/10.1016/j.febslet.2004.06.070>.
- Lin, H., Shen, Q., Zhan, J.M., Wang, Q. & Zhao, Y.H. (2013) Evaluation of Bacterial Expansin EXLX1 as a Cellulase Synergist for the Saccharification of Lignocellulosic Agro-Industrial Wastes. *PLoS ONE*, 8, 2-9.

- <https://doi.org/10.1371/journal.pone.0075022>.
- Litou, Z.I., Bagos, P.G., Tsigirgos, K.D., Liakopoulos, T.D. & Hamodrakas, S.J. (2008) Prediction of cell wall sorting signals in gram-positive bacteria with a hidden markov model: Application to complete genomes. *Journal of Bioinformatics and Computational Biology*, 6, 387-401.
<https://doi.org/10.1142/S0219720008003382>.
- Liu, X., Liu, C., Ma, Y., Hong, J. & Zhang, M. (2014) Heterologous expression and functional characterization of a novel cellulose-disruptive protein LeEXP2 from lycopersicum esculentum. *Journal of Biotechnology*, 186, 148-155. <https://doi.org/10.1016/j.jbiotec.2014.07.013>.
- Lohoff, C., Buchholz, P.C.F., Roes-Hill, M. Le & Pleiss, J. (2020) Expansin Engineering Database: A navigation and classification tool for expansins and homologues.
- Lorenz, C., Dougherty, T.J. & Lory, S. (2019) Correct sorting of lipoproteins into the inner and outer membranes of pseudomonas aeruginosa by the escherichia coli lolcde transport system. *mBio*, 10, 1-13. <https://doi.org/10.1128/MBIO.00194-19>.
- Malik, A. & Kim, S.B. (2019) A comprehensive in silico analysis of sortase superfamily. *Journal of Microbiology*, 57, 431-443. <https://doi.org/10.1007/s12275-019-8545-5>.
- Marczak, M., Mazur, A., Król, J.E., Gruszecki, W.I. & Skorupska, A. (2006) Lipoprotein PssN of Rhizobium leguminosarum bv. trifolii: Subcellular localization and possible involvement in exopolysaccharide export. *Journal of Bacteriology*, 188, 6943-6952. <https://doi.org/10.1128/JB.00651-06>.
- Marraffini, L.A., DeDent, A.C. & Schneewind, O. (2006) Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews*, 70, 192-221. <https://doi.org/10.1128/mubr.70.1.192-221.2006>.
- Martinez-Anaya, C. (2016a) Understanding the structure and function of bacterial expansins: a prerequisite towards practical applications for the bioenergy and agricultural industries. *Microbial Biotechnology*, 9, 727-736. <https://doi.org/10.1111/1751-7915.12377>.
- Martinez-Anaya, C. (2016b) Understanding the structure and function of bacterial expansins: a prerequisite towards practical applications for the bioenergy and agricultural industries. *Microbial Biotechnology*, 9, 727-736. <https://doi.org/10.1111/1751-7915.12377>.
- Mészáros, B., Erdős, G. & Dosztányi, Z. (2018) IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, 46, W329-W337. <https://doi.org/10.1093/nar/gky384>.
- Narita, S. ichiro & Tokuda, H. (2017) Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1862, 1414-1423. <https://doi.org/10.1016/j.bbalip.2016.11.009>.
- Narváez-Barragán, D.A., Tovar-Herrera, O.E., Torres, M., Rodríguez, M., Humphris, S., Toth, I.K., et al. (2020) Expansin-like ExI1 from Pectobacterium is a virulence factor required for host infection, and induces a defence plant response involving ROS, and jasmonate, ethylene and salicylic acid signalling pathways in Arabidopsis thaliana. *Scientific Reports*, 10, 1-14. <https://doi.org/10.1038/s41598-020-64529-9>.
- Nguyen, L.T., Schmidt, H.A., Haeseler, A. Von & Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268-274. <https://doi.org/10.1093/molbev/msu300>.

- Nielsen, H., Tsirigos, K.D., Brunak, S. & Heijne, G. von (2019) A Brief History of Protein Sorting Prediction. *Protein Journal*, 38, 200-216. <https://doi.org/10.1007/s10930-019-09838-3>.
- Nikolaidis, N., Doran, N. & Cosgrove, D.J. (2014a) Plant expansins in bacteria and fungi: Evolution by horizontal gene transfer and independent domain fusion. *Molecular Biology and Evolution*, 31, 376-386. <https://doi.org/10.1093/molbev/mst206>.
- Nikolaidis, N., Doran, N. & Cosgrove, D.J. (2014b) Plant expansins in bacteria and fungi: Evolution by horizontal gene transfer and independent domain fusion. *Molecular Biology and Evolution*, 31, 376-386. <https://doi.org/10.1093/molbev/mst206>.
- Novick, R.P. (2000) Sortase: The surface protein anchoring transpeptidase and the LPXTG motif. *Trends in Microbiology*, 8, 148-151. [https://doi.org/10.1016/S0966-842X\(00\)01741-8](https://doi.org/10.1016/S0966-842X(00)01741-8).
- Ogasawara, S., Shimada, N. & Kawata, T. (2009) Role of an expansin-like molecule in Dictyostelium morphogenesis and regulation of its gene expression by the signal transducer and activator of transcription protein Dd-STATA: Original Article. *Development Growth and Differentiation*, 51, 109-122. <https://doi.org/10.1111/j.1440-169X.2009.01086.x>.
- Okuda, S. & Tokuda, H. (2011) Lipoprotein sorting in bacteria. *Annual Review of Microbiology*, 65, 239-259. <https://doi.org/10.1146/annurev-micro-090110-102859>.
- Olarte-Lozano, M., Mendoza-Nuñez, M.A., Pastor, N., Segovia, L., Folch-Mallol, J. & Martínez-Anaya, C. (2014) PcEx11 a novel acid expansin-like protein from the plant pathogen Pectobacterium carotovorum, binds cell walls differently to BsEXLX1. *PLoS ONE*, 9. <https://doi.org/10.1371/journal.pone.0095638>.
- Ouyang, P., He, X., Yuan, Z.W., Yin, Z.Q., Fu, H., Lin, J., et al. (2018) Erianin against staphylococcus aureus infection via inhibiting sortase A. *Toxins*, 10, 1-14. <https://doi.org/10.3390/toxins10100385>.
- Pastor, N., Dávila, S., Pérez-Rueda, E., Segovia, L. & Martínez-Anaya, C. (2015) Electrostatic analysis of bacterial expansins. *Proteins: Structure, Function, and Bioinformatics*, 83, 215-223. <https://doi.org/10.1002/prot.24718>.
- Paulsson, M., Kragh, K.N., Su, Y.C., Sandblad, L., Singh, B., Bjarnsholt, T., et al. (2021) Peptidoglycan-Binding Anchor Is a Pseudomonas aeruginosa OmpA Family Lipoprotein With Importance for Outer Membrane Vesicles, Biofilms, and the Periplasmic Shape. *Frontiers in Microbiology*, 12, 1-13. <https://doi.org/10.3389/fmicb.2021.639582>.
- Pech-Cervantes, A.A., Ogunade, I.M., Jiang, Y., Irfan, M., Arriola, K.G., Amaro, F.X., et al. (2019) An expansin-like protein expands forage cell walls and synergistically increases hydrolysis, digestibility and fermentation of livestock feeds by fibrolytic enzymes. *PLoS ONE*, 14, 1-23. <https://doi.org/10.1371/journal.pone.0224381>.
- Pfeilmeier, S., Caly, D.L. & Malone, J.G. (2016) Bacterial pathogenesis of plants: future challenges from a microbial perspective: Challenges in Bacterial Molecular Plant Pathology. *Molecular plant pathology*, 17, 1298-1313. <https://doi.org/10.1111/mpp.12427>.
- Qin, Y.M., Tao, H., Liu, Y.Y., Wang, Y.D., Zhang, J.R. & Tang, A.X. (2013) A novel non-hydrolytic protein from Pseudomonas oryzihabitans enhances the enzymatic hydrolysis of cellulose. *Journal of Biotechnology*, 168, 24-31. <https://doi.org/10.1016/j.jbiotec.2013.07.028>.
- Rayes, J. El, Rodríguez-Alonso, R. & Collet, J.F. (2021) Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles. *Current Opinion in Microbiology*, 61, 25-

34. <https://doi.org/10.1016/j.mib.2021.02.003>.
- Rocha, J., Shapiro, L.R. & Kolter, R. (2020) A horizontally acquired expansin gene increases virulence of the emerging plant pathogen *Erwinia tracheiphila*. *Scientific Reports*, 10, 1-15. <https://doi.org/10.1038/s41598-020-78157-w>.
- Rodríguez-Ortega, M.J. (2018) The Surfaceome.
- Sampedro, J. & Cosgrove, D.J. (2005) The expansin superfamily. *Genome Biology*, 6, 1-11. <https://doi.org/10.1186/gb-2005-6-12-242>.
- Schubiger, C.B., Orfe, L.H., Sudheesh, P.S., Cain, K.D., Shah, D.H. & Calla, D.R. (2015) Entericidin is required for a probiotic treatment (*Enterobacter* sp. Strain C6-6) to protect trout from cold-water disease challenge. *Applied and Environmental Microbiology*, 81, 658-665. <https://doi.org/10.1128/AEM.02965-14>.
- Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 2498-2504. <https://doi.org/10.1101/gr.1239303.metabolite>.
- Shin, J.-H., Jeong, D.-H., Park, M.C. & An, G. (2005) Characterization and Transcriptional Expression of the a-Expansin Gene Family in Rice. *Molecules and Cells*, 20, 210-218.
- Si, L., Li, P., Liu, X. & Luo, L. (2016) Chinese herb medicine against Sortase A catalyzed transformations, a key role in gram-positive bacterial infection progress. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 31, 184-196. <https://doi.org/10.1080/14756366.2016.1178639>.
- Si, T.Y., Yuan, J.H., Wu, Y.L. & Tang, J.X. (2016) Physical biology of bacterial motility. *Wuli Xuebao/Acta Physica Sinica*, 65. <https://doi.org/10.7498/aps.65.178703>.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7. <https://doi.org/10.1038/msb.2011.75>.
- Silhavy, T.J., Kahne, D. & Walker, S. (2010) The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, 2, a000414-a000414. <https://doi.org/10.1101/cshperspect.a000414>.
- Silveira, R.L. & Skaf, M.S. (2016) Molecular dynamics of the *Bacillus subtilis* expansin EXLX1: Interaction with substrates and structural basis of the lack of activity of mutants. *Physical Chemistry Chemical Physics*, 18, 3510-3521. <https://doi.org/10.1039/c5cp06674c>.
- Su, Y., Xu, Y., Liang, H., Yuan, G., Wu, X. & Zheng, D. (2021) Genome-Wide Identification of *Ralstonia solanacearum* Genes Required for Survival in Tomato Plants. *mSystems*, 6. <https://doi.org/10.1128/msystems.00838-21>.
- Susmitha, A., Bajaj, H. & Madhavan Nampoothiri, K. (2021) The divergent roles of sortase in the biology of Gram-positive bacteria. *The Cell Surface*, 7, 100055. <https://doi.org/10.1016/j.tcs.2021.100055>.
- Tabuchi, A., Li, L.C. & Cosgrove, D.J. (2011) Matrix solubilization and cell wall weakening by β -expansin (group-1 allergen) from maize pollen. *Plant Journal*, 68, 546-559. <https://doi.org/10.1111/j.1365-313X.2011.04705.x>.
- Tancos, M.A., Lowe-Power, T.M., Peritore-Galve, F.C., Tran, T.M., Allen, C. & Smart, C.D. (2018) Plant-like bacterial expansins play contrasting roles in two tomato vascular pathogens. *Molecular Plant Pathology*, 19, 1210-1221. <https://doi.org/10.1111/mpp.12611>.

- Tovar-Herrera, O.E., Rodríguez, M., Olarte-Lozano, M., Sampedro-Guerrero, J.A., Guerrero, A., Pinto-Cámara, R., et al. (2018a) Analysis of the Binding of Expansin ExI1, from *Pectobacterium carotovorum*, to Plant Xylem and Comparison to EXLX1 from *Bacillus subtilis*. *ACS Omega*, 3, 7008–7018. <https://doi.org/10.1021/acsomega.8b00406>.
- Tovar-Herrera, O.E., Rodríguez, M., Olarte-Lozano, M., Sampedro-Guerrero, J.A., Guerrero, A., Pinto-Cámara, R., et al. (2018b) Analysis of the Binding of Expansin ExI1, from *Pectobacterium carotovorum*, to Plant Xylem and Comparison to EXLX1 from *Bacillus subtilis*. *ACS Omega*, 3, 7008-7018. <https://doi.org/10.1021/acsomega.8b00406>.
- Tseng, T.T., Tyler, B.M. & Setubal, J.C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, 9, 1-9. <https://doi.org/10.1186/1471-2180-9-S1-S2>.
- Tusnády, G.E., Dobson, L. & Tompa, P. (2015) Disordered regions in transmembrane proteins. *Biochimica et Biophysica Acta - Biomembranes*, 1848, 2839–2848. <https://doi.org/10.1016/j.bbamem.2015.08.002>.
- Valdivia, E.R., Stephenson, A.G., Durachko, D.M. & Cosgrove, D. (2009) Class B β -expansins are needed for pollen separation and stigma penetration. *Sexual Plant Reproduction*, 22, 141-152. <https://doi.org/10.1007/s00497-009-0099-y>.
- Vanini, M.M.T., Spisni, A., Sforça, M.L., Pertinhez, T.A. & Benedetti, C.E. (2008) The solution structure of the outer membrane lipoprotein OmlA from *Xanthomonas axonopodis* pv. *citri* reveals a protein fold implicated in protein-protein interaction. *Proteins: Structure, Function and Genetics*, 71, 2051-2064. <https://doi.org/10.1002/prot.21886>.
- Veith, P.D., Glew, M.D., Gorasia, D.G. & Reynolds, E.C. (2017) Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Molecular Microbiology*, 106, 35-53. <https://doi.org/10.1111/mmi.13752>.
- Viljoen, A., Foster, S.J., Fantner, G.E., Hobbs, J.K. & Dufrêne, Y.F. (2020) Scratching the surface: bacterial cell envelopes at the nanoscale. *mBio*, 11, 1-12. <https://doi.org/10.1128/mBio.03020-19>.
- Walker, S.L., Redman, J.A. & Elimelech, M. (2004) Role of cell surface lipopolysaccharides in *Escherichia coli* K12 adhesion and transport. *Langmuir*, 20, 7736-7746. <https://doi.org/10.1021/la049511f>.
- Wang, Q., Chen, L., Lin, H., Yu, D., Shen, Q., Wan, L., et al. (2016) The binding, synergistic and structural characteristics of BsEXLX1 for loosening the main components of lignocellulose: Lignin, xylan, and cellulose. *Enzyme and Microbial Technology*, 92, 67-75. <https://doi.org/10.1016/j.enzmictec.2016.06.017>.
- Wang, T., Chen, Y., Tabuchi, A., Cosgrove, D.J. & Hong, M. (2016) The Target of β -Expansin EXPB1 in Maize Cell Walls from Binding and Solid-State NMR Studies. *Plant Physiology*, 172, 2107-2119. <https://doi.org/10.1104/pp.16.01311>.
- Wang, W., Liu, C., Ma, Y., Liu, X., Zhang, K. & Zhang, M. (2014) Improved production of two expansin-like proteins in *Pichia pastoris* and investigation of their functional properties. *Biochemical Engineering Journal*, 84, 16-27. <https://doi.org/10.1016/j.bej.2013.12.018>.
- Wely, K.H.M. Van, Swaving, J., Freudl, R. & Driessen, A.J.M. (2001) Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiology Reviews*, 25, 437-454. [https://doi.org/10.1016/S0168-6445\(01\)00062-6](https://doi.org/10.1016/S0168-6445(01)00062-6).

- Wielbo, J., Mazur, A., Król, J.E., Marczak, M. & Skorupska, A. (2004) Environmental modulation of the pssTNOP gene expression in *Rhizobium leguminosarum* bv. trifolii. *Canadian Journal of Microbiology*, 50, 201-211. <https://doi.org/10.1139/w04-004>.
- Wilkins, S. (2015) Structure and mechanism of ABC transporters. *F1000Prime Reports*, 7, 1-9. <https://doi.org/10.12703/P7-14>.
- Wilson, M.M. & Bernstein, H.D. (2016) Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. *Trends in Microbiology*, 24, 198-208. <https://doi.org/10.1016/j.tim.2015.11.006>.
- Wu, Y., Jiang, Y., Kaiser, D. & Alber, M. (2007) Social interactions in myxobacterial swarming. *PLoS Computational Biology*, 3, 2546-2558. <https://doi.org/10.1371/journal.pcbi.0030253>.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. & Baker, D. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 1496-1503. <https://doi.org/10.1073/pnas.1914677117>.

Capítulo VII. Anexos

Artículo Principal

<https://doi.org/10.1002/pro.4315>

Prevalent association with the bacterial cell envelope of prokaryotic expansins revealed by bioinformatics analysis

Andrés de Sandozequi  | Juan José Salazar-Cortés | Irán Tapia-Vázquez  |
Claudia Martínez-Anaya 

Departamento de Ingeniería Celular y Biotatálisis, Instituto de Biotecnología, UNAM, Av. Universidad 2001, Chamilpa, Cuernavaca, Morelos, Mexico

Correspondence

Claudia Martínez-Anaya, Departamento de Ingeniería Celular y Biotatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Av. Universidad 2001, Col. Chamilpa 62250, Cuernavaca, Morelos, Mexico.
Email: claudia.martinez@ibt.unam.mx

Funding information

Consejo Nacional de Ciencia y Tecnología, Grant/Award Number: 252551; Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, Grant/Award Number: IN211019 and IN203622

Review editor: Nir Ben-Tal

Abstract

Expansins are a group of proteins from diverse organisms from bacteria to plants. Although expansins show structural conservation, their biological roles seem to differ among kingdoms. In plants, these proteins remodel the cell wall during plant growth and other processes. Contrarily, determination of bacterial expansin activity has proven difficult, although genetic evidence of bacterial mutants indicates that expansins participate in bacteria–plant interactions. Nevertheless, a large proportion of expansin genes are found in the genomes of free-living bacteria, suggesting roles that are independent of the interaction with living plants. Here, we analyzed all available sequences of prokaryotic expansins for correlations between surface electric charge, extra protein modules, and sequence motifs for association with the bacteria exterior after export. Additionally, information on the fate of protein after translocation across the membrane also points to bacterial cell association of expansins through six different mechanisms, such as attachment of a lipid molecule for membrane anchoring in diderm species or covalent linking to the peptidoglycan layer in monoderms such as the Bacilliales. Our results have implications for expansin function in the context of bacteria–plant interactions and also for free-living species in which expansins might affect cell–cell or cell–substrate interaction properties and indicate the need to re-examine the roles currently considered for these proteins.

KEYWORDS

bacterial expansin, cell envelope, isoelectric point, protein domain, signal peptide

1 | INTRODUCTION

Unlike plant expansins, the biological role of bacterial expansins is still unknown.¹ In plants, expansins participate in the physiological remodeling of the cell wall in processes such as tissue expansion or pollen tube growth, among others.² The role of bacterial expansins has been only inferred from genetic experiments showing their involvement in plant colonization by phytopathogens or

plant-growth-promoting bacteria.^{3–7} Different approaches to determine the bacterial expansin function and molecular mechanisms have produced marginal results; in cellulose creep assays, bacterial expansins perform far below the efficiency of their plant homologues, whereas in filter paper weakening assays bacterial expansins performed only slightly better.^{8–10} The reports of bacterial expansin synergism with hydrolytic enzymes (such as cellulases, due to modification of the plant cell wall to facilitate

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

enzyme access to the substrate) are controversial, as results exist with positive and negative outcomes.^{11,12} We recently reported the ability of expansin Ex11 from *Pectobacterium brasiliense* to solubilize a yet unidentified polysaccharide from isolated vascular tissue from Swiss chard,⁵ but the extent of this observation to other bacterial expansins and its biological relevance are unclear, although a function in plant colonization by bacteria was suggested. Besides experimental methods, bioinformatic tools help in identifying key features in bacterial expansins sequences and predicted structures, such as the amount of electric charge on their surface.¹³ This is important because protein surface charge affects expansin binding to polysaccharides of different types within the plant cell wall; for instance, basic expansin EXLX1 from *Bacillus subtilis* binds to negatively charged polysaccharides in most cell types of celery petioles, meanwhile under the same conditions acidic Ex11 from *P. brasiliense* localizes to the cell corners between xylem vessels and surrounding cells.^{5,14} Additionally, extra protein modules have been bioinformatically predicted in microbial expansins that could shed light on their activity, for example, Ex11 from *Clostridium clariflavum* contains a dockerin module that attaches it to the cell surface due to its association with scaffoldin ScaG, suggesting a participation in cellulose utilization in this species.¹⁵ All prokaryote and eukaryote expansins contain a signal peptide for secretion at the N-terminus. In the case of bacteria, it has been assumed that the mature protein is released outside the cell for binding to an external ligand (i.e., to the plant cell wall polysaccharides). Indeed, heterologously produced expansins from *B. subtilis*, *Pectobacterium carotovorum* (now *P. brasiliense*), *Hahella chejuensis*, among others, are soluble and interact with plant cell walls and cellulose.^{14,16,17} However, EXLX1 from *B. subtilis* also binds peptidoglycan with high affinity,¹⁸ and endogenously produced Ex11 from *P. brasiliense* is mainly found in the pellet of cells growing in LB medium, suggesting that it remains bound to the bacteria after processing. Thus, the possibility exists that bacterial expansins associate with an unknown substrate in their own cell envelopes.

Several mechanisms to translocate extracellular proteins across the cell envelope are found in both diderm and monoderm bacteria. Protein translocation is essential to all organisms and is especially crucial for pathogens since they secrete virulence factors involved in host infection. Protein export in bacteria has been exhaustively studied,^{19,20} with the identification of nine secretion systems (Types I to IX), two inner membrane translocation systems (Sec and Tat), one protein sorting complex in monoderm bacteria, and one injectosome. Proteins secreted by bacteria can have

three main fates dictated by their signal peptides and other sorting signals: (a) they can be directed to the extracellular milieu as soluble proteins or as part of the extracellular matrix; (b) they can remain anchored to the cell surface on the membrane as lipoproteins or covalently bound to the peptidoglycan layer by sortases; or (c) they become part of cell superstructures like flagella or the cellulosome.^{19,21–23}

In this study, we conducted an up-to-date screening for expansin proteins in annotated bacterial proteomes and explored the relationship of certain sequence and structural characteristics: phylogenetic relationships, isoelectric point, and protein domain modularity. We also studied their secretion signal peptides that, interestingly, indicate that once translocated, many expansins may remain associated with the bacterial surface. These findings develop our knowledge of how bacterial expansin activity might be interpreted in the context of plant–bacteria interactions or for the energy industry and illuminates potential roles for expansins in free-living bacteria.

2 | RESULTS

We updated our bioinformatics analyses with prokaryotic expansin sequences recently deposited in public databases. High sequence divergence and lack of a consensus sequence of the bacterial expansin protein family have resulted in ambiguous annotation in all databases (e.g., “Rare lipoprotein A”, “Putative endoglucanase c-terminal domain harbouring protein”, “Peptidoglycan-binding domain-containing protein”, etc.). Hence, we used a hidden Markov model (HMM) approach to find expansins in bacterial proteomes. We constructed an HMM profile using 44 expansin domains (available as File SS1) and searched against nonredundant proteomes in UniProt, obtaining 908 putative expansins composed by at least the two canonical domains: D1, at the N-terminus with a double psi beta barrel fold containing the pseudocatalytic aspartate at the equivalent position of D82 of *B. subtilis*, and D2 polysaccharide binding domain containing three aromatic residues for binding to the substrate. Truncated sequences were discarded, but longer genes with additional sequences for extra domains were included. Our search results are similar to others previously reported^{24,25} in that we found expansin genes across nine bacterial phyla and 43 orders (Table SS1 and Figure SS1). Most of them belong to Actinobacteria (47.7%), Proteobacteria (29.9%), and Firmicutes (14.6%), and <2.5% each were divided between Fibrobacteres, Bacteroidetes, Chloroflexi, and Cyanobacteria (Figure 1). We also found three and two expansin proteins in the

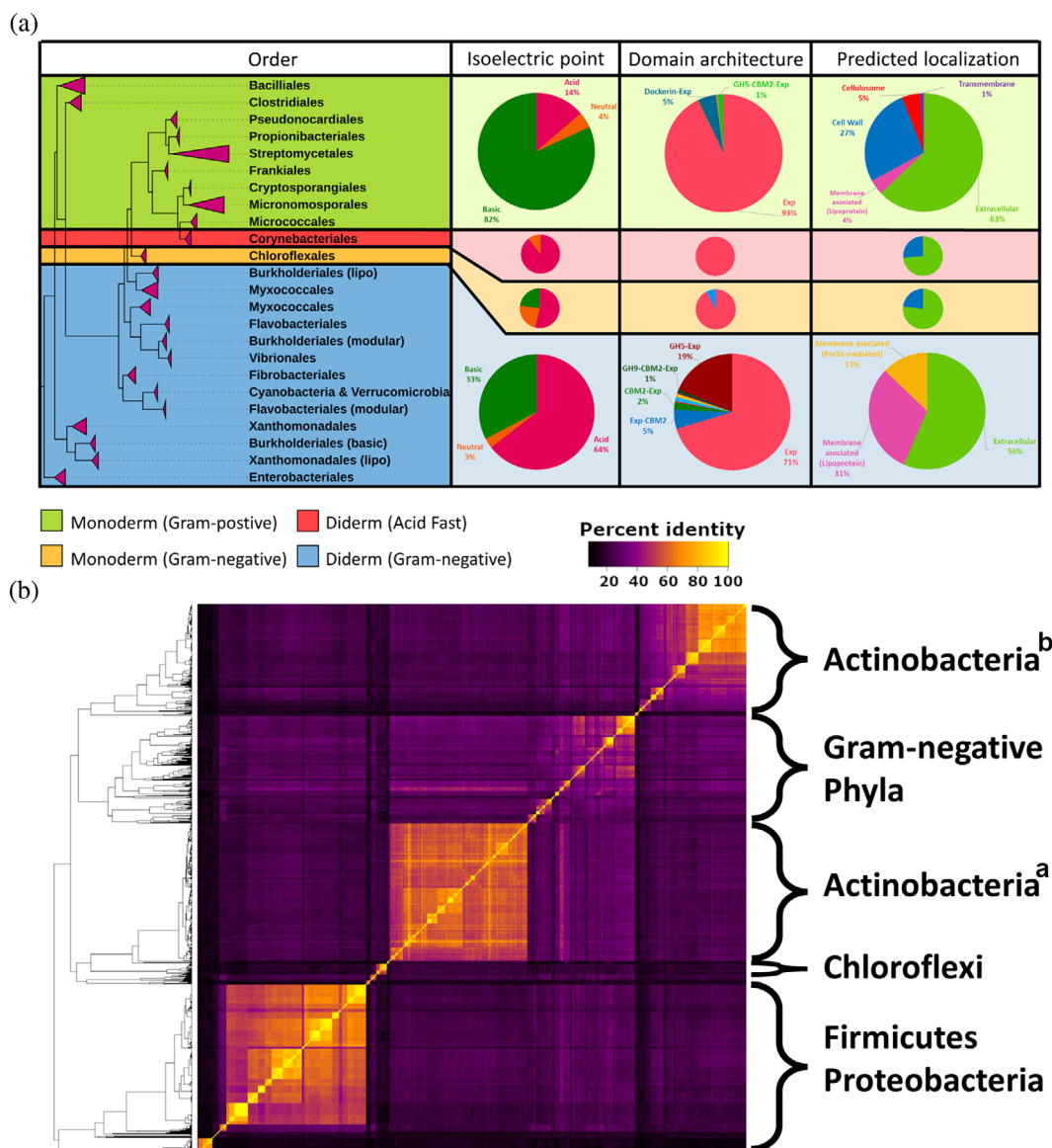


FIGURE 1 Bioinformatic analysis of 908 bacterial expansins. (a) maximum-likelihood cladogram constructed with expansin domains. The tree should be considered unrooted and was ordered according to cell envelope type. Bootstrap values were calculated with 1000 iterations. Predicted protein localization is based on sequence motifs and additional domains. Complete tree with the information of each sequence is provided in File S1. To the right of the tree, pie charts indicate the proportion sequences for each bacterial type by isoelectric point, domain architecture, and protein localization. (b) Heat map of sequence percent identity between all expansin domains. The larger group of Actinobacteria is Group A, and the other is Group B. Expansins from Firmicutes and some Proteobacteria (*Enterobacteriales*, *Ralstonia* spp., and *Xanthomonadales*) group together and have high sequence similarity. CBM, carbohydrate-binding module; Exp, expansin domain; GH5, family 5 glycoside hydrolase

Phyla Verrucomicrobia and Spirochaetes, respectively (Figure SS1). Due to overrepresentation of sequenced clades (such as gamma-Proteobacteria and Streptomycetales), it is not entirely clear if these proportions represent the actual distribution of expansins in nature. Of these sequences, 84.5% were canonical expansins (monomodular expansins), and 15.5% were multimodular proteins with additional functional domains, such as a family 5 glycoside hydrolase (GH5) domain or a family

2 carbohydrate-binding module (CBM2) domain. A heat map (Figure 1b) constructed with an all-vs-all pairwise alignment matrix of expansin domains revealed that there are four main groups of expansins (with at least 50% identity), which are also represented in the complete cladogram (Figure SS1): the Firmicutes/Proteobacteria group, the diderm phyla group, and two Actinobacterial groups. A minor group with 13 expansin sequences from Chloroflexi species and sequences that showed high

divergence from canonical expansin sequences were not defined as a group (Figure 1b).

We analyzed different characteristics of expansins combined in a search for patterns that could provide clues on their activity. Expansins are highly charged proteins with acidic or basic residues at the surface²⁶ and this is reflected in their isoelectric point (pI); we have previously studied the correlation of the pI with bacterial type: Gram-positive bacteria contained mainly basic expansins, whereas most Gram-negative bacteria contained acidic expansins.¹³ Our analysis of 908 expansins showed a more complex situation (Figure 1), with pIs ranging from 3.5 to 11 in different bacterial orders (Figure S2), in which specific clades show a continuum of pI values, such as the Streptomycetales or the Burkholderiales. We reasoned that this could be related to the specific characteristics of their envelope or related to their domain modularity. To test this hypothesis, we grouped bacteria according to their cell envelope as acid fast, monoderm (Gram-positive), monoderm (Gram-negative), diderm (Gram-negative), and the presence of extra protein domains (Figure 1a). We also analyzed expansins based on their signal peptide type, presence of sorting motifs, and sorting domains. With this information, we classified expansins into six groups according to putative protein localizations: (a) “extracellular expansins,” which only contain a Sec or a Tat signal peptide and Signal Peptidase I cleavage motif; (b) “cell wall expansins,” containing a Sec or Tat signal peptide and a putative LPXTG-like motif; (c) “transmembrane expansins,” containing N-terminal cytoplasmic and transmembrane domains; (d) “cellulosome expansins,” containing a Sec or Tat signal peptide and a dockerin domain; (e) “membrane associated (PorSS mediated) expansins,” containing a Sec signal peptide and a Type 9 secretion system (T9SS, also known as Por secretion system) C-terminal domain (CTD); and (f) “membrane associated (lipoprotein) expansins” (from now on referred as lipo-expansins) containing a Sec signal peptide and a Signal Peptidase II cleavage motif, which are processed to become a prolipoprotein embedded to the membrane (Figure 1a).

2.1 | Expansins from Gram-negative bacteria

Although 64% expansins from Proteobacteria are monomodular acidic proteins without sequence information on localization other than that required for translocation through the inner membrane (Sec or Tat signal peptide and cleaved by SP1), there are cases that provide interesting clues on the nature of expansin diversity in this cluster. Perhaps the most informative are those from

Burkholderiales (Figure 2), composed of three well-defined groups: basic extracellular expansins from *Ralstonia* spp., acidic lipo-expansins, and acidic modular CBM-expansins. Both acidic groups belong to *Acidovorax* spp. and related species, and cluster together (Figure 1b, Gram-negative clusters), but basic expansins belonging to the species of *Ralstonia* have a high sequence similarity to expansins from Firmicutes, Xanthomonadales, and Enterobacteriales (Figure 1b, Firmicutes/Proteobacteria cluster; Figure S2). The latter are acidic monomodular expansins, predicted to be secreted through the Sec system and cleaved by SPI (Figure 2); expansin Ex11 from *Pectobacterium* species is a representative of this group (Figure 5a). And, while all other expansins in this cluster are basic, it is interesting that expansins from Enterobacteriales are acidic, suggesting an unknown evolutionary pressure that drives this characteristic.

Expansins from Xanthomonadales (fused to a GH5 domain) and *Ralstonia* spp. (monomodular) are basic with high sequence similarity closely related to Firmicutes expansins (Figure 1b, >60% similarity). This indicates a (possibly functional) selective pressure to conserve similar sequences in such divergent orders. Although most expansins from Proteobacteria are monomodular, it seems that the most abundant modular expansins appeared in three different Orders: in Xanthomonadales (Figure 5c), in Burkholderiales (Figure 5d), and in Myxococcales (Figure 5e).

Expansin from Fibrobacteres and Bacteroidetes have a Sec signal peptide sequence necessary for inner membrane translocation, and a CTD that functions as a recognition signal for T9SS. This secretion system can anchor the secreted CTD containing protein to the outer membrane by cleaving the CTD and attaching an anionic lipopolysaccharide (A-LPS). The T9SS secretes motility related proteins, adhesins, and exports virulence factors to the cell surface in the Fibrobacteres–Clostridia–Bacteroidetes super-phylum.^{22,27–29} Very few Cyanobacterial expansins were found, but almost all of them (five of the six expansins) are modular and their domain architectures are unique to the clade (Figure 2). The Gram-negative has the most abundant variety of domain architectures with eight different modular expansins, and in 44% sequence information indicates that they either remain associated with the membrane as lipoproteins or because of T9SS CTD post-translational modification.

2.2 | Expansins from Firmicutes

Perhaps the most studied bacterial expansin, EXLX1 from *B. subtilis*, is a basic monomodular expansin (Figure 5b).

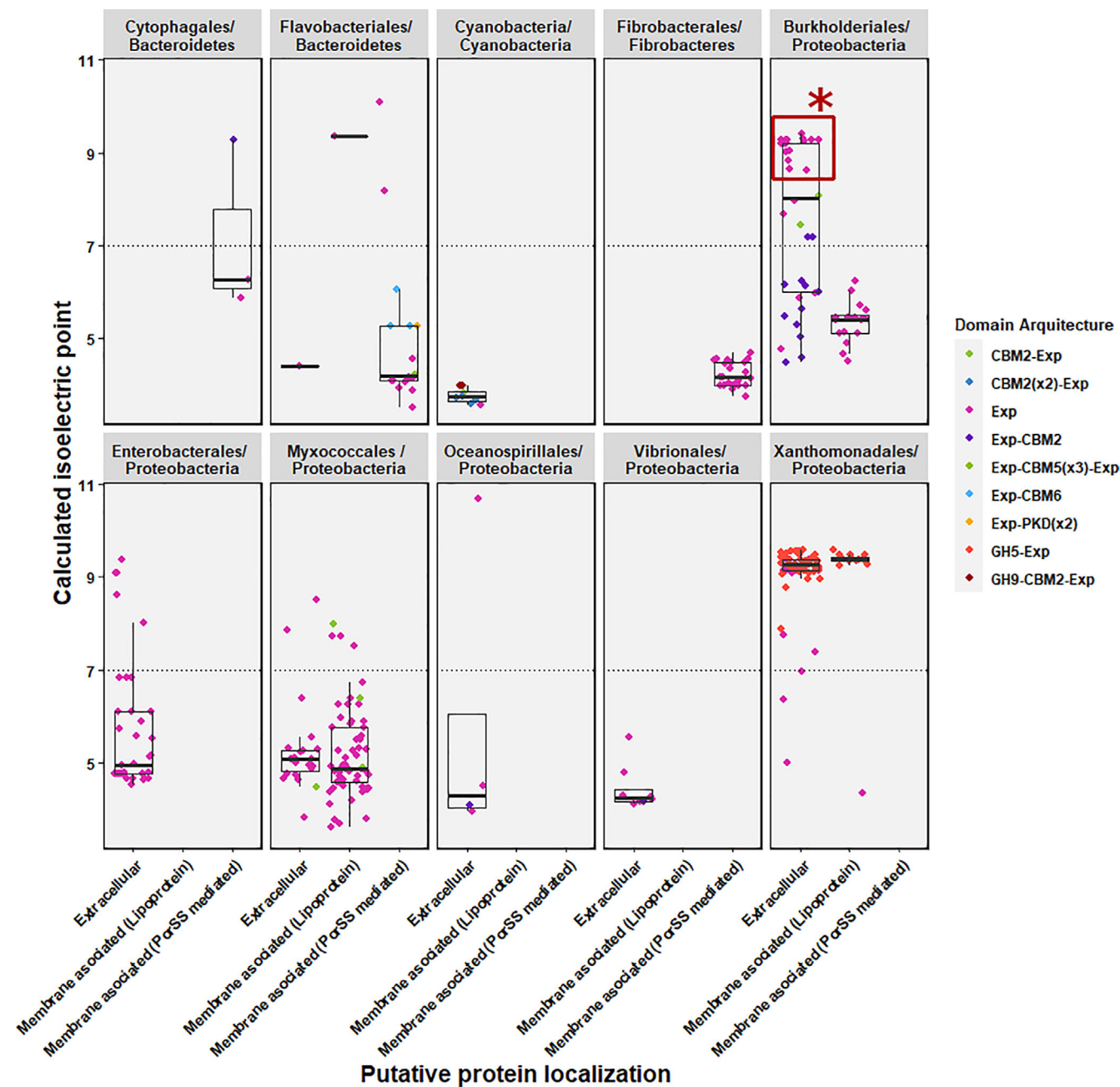


FIGURE 2 Clustering of expansins from Diderm (Gram-negative) Phyla. Distribution of expansins as defined by the isoelectric point of the expansin domain (~220 amino acids) and the putative protein localization predicted from sequence information. Each dot represents a unique sequence, and it is coloured by the entire protein domain architecture represented N- to C-terminal. *Dots in the red rectangle show basic expansins from *Ralstonia* spp. CBM, carbohydrate-binding module (families 2, 5, and 6); Exp, expansin domain; GH5, family 5 glycoside hydrolase; PKD, polycystic kidney disease domain

EXLX1 belongs to the Firmicutes cluster but is highly similar (>60% similarity) to some expansin domains from gamma- and beta-Proteobacteria (Figure 1b), supporting the idea that they have been horizontally exchanged as previously suggested.^{6,25} It is also possible that a basic nature is required to perform their biological function. Interestingly, 75% of sequences from Bacilliales species contain a putative LPXTG-like motif and a C-terminal hydrophobic region. The LPXTG was the first discovered

motif to be the substrate of sortases, but since then, some other conserved motifs have been found and named LPXTG-like motifs (such as LPNTA, [I/L][P/A]XTG or LAXTG).²³ Expansins from Firmicutes contain a somehow degenerated LPXTG motif (Figure 3b), not recognized by some bioinformatic programs like CW-PRED. But the similarity of the motif, with some sequences even having the characteristic “triple G” at the end of the LPXTG motif (Figure 3b),³⁰ the presence of a

hydrophobic tail and a positively charged clamp,³¹ suggests that they could be recognized by a sortase, although experimental confirmation is required.

In the case of Clostridiales, 73% of sequences are acidic expansins fused to a dockerin domain for attachment to the cellulosome via cohesin interactions (Figure 5f). Indeed, a dockerin-Exl1 expansin from *C. clariflavum* interacts with Type-I cohesins from various scaffoldin proteins, thus confirming experimentally that this is a Type-I dockerin.^{15,32} Artzi et al. analyzed the affinity of the dockerin domains to different cohesin modules, showing that the strongest interaction was with the monovalent scaffoldin ScaG-cohesin, in comparison with ScaA and ScaD.¹⁵ We can suppose that dockerin-expansins have a different mechanism of action in comparison with others from bacteria with different lifestyles such as phytopathogens or soil inhabitants. Thus, if the type of expansin electric charge is

important for their biological role in different species, the results found for Clostridiales may explain the difference in isoelectric point of their expansins with respect to other monoderms which are basic (Figures 3 and 4). It is worth noting that ScaG protein has a cell surface binding module and is predicted to be a cell-anchored scaffoldin.³² Interestingly, we found high sequence similarity between expansins from Clostridiales and others of the Firmicutes phylum (Figure 1), with some even containing a putative LPXTG-like motif.

2.3 | Expansins from Actinobacteria

The expansins in Actinobacteria cluster into two separate groups: the first group is formed by Streptomycetales, Pseudonocardiales, and Propionobacteriales (Figure 1b,

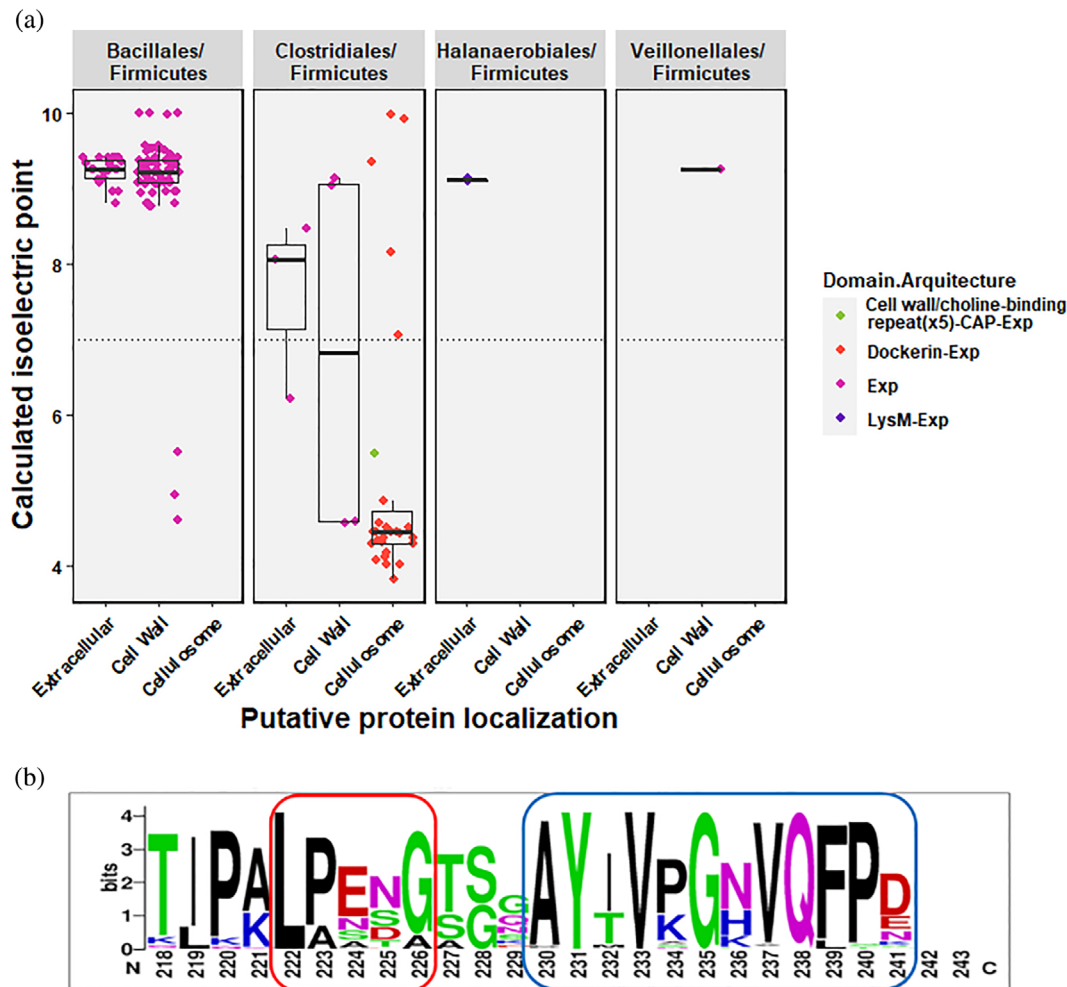


FIGURE 3 Clustering and motifs of expansins from Firmicutes. (a) Distribution of expansins as defined by the isoelectric point of the expansin domain (~220 amino acids) and the putative protein localization predicted with sequence information. Each dot represents a unique sequence and is colored by the entire protein domain architecture represented N- to C-terminal. (b) Alignment of the C-terminal region of 96 expansins belonging to Bacilliales: 75% of expansins contain LPXTG-like motifs with a significant p -value ($<1e-5$). CAP, cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 domains; Exp, expansin domain; LysM, lysin motif domain

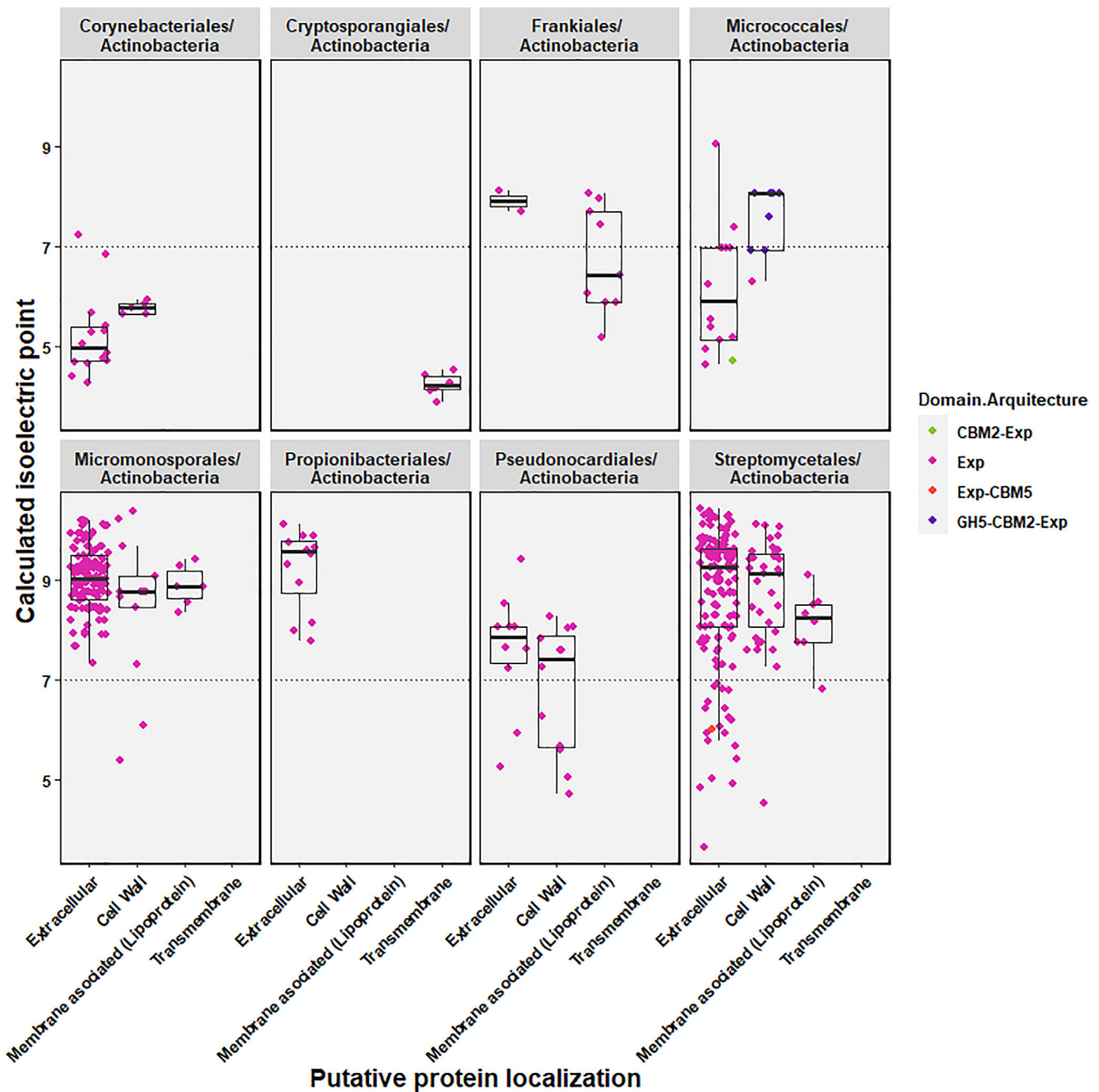


FIGURE 4 Expansins from Actinobacteria. Distribution of expansins as defined by their isoelectric point of the expansin domain (~220 amino acids) and the putative protein localization predicted with sequence information. Each dot represents a unique sequence and is colored by the entire protein domain architecture represented N- to C-terminal. CBM, carbohydrate-binding module (families 2 and 5); Exp, expansin domain; GH5, family 5 glycoside hydrolase

Actinobacteria-a) and the other group comprising Micromonosporales, Micrococcales, Corynebacteriales, Cryptosporangiales, and Frankiales orders (Figure 1b, Actinobacteria-b). The majority of Actinobacterial expansins (75%) are basic, monomodular, and lack sequence-derived evidence of cell surface attachment; however, we found that a number of expansins in Corynebacteriales (26%), Micromonosporales (10%),

Micrococcales (39%), Pseudonocardiales (54%), and Streptomycetales (18%) contain putative sortase motifs; also, the orders of Frankiales (84%), Micromonosporales (4.4%), and Streptomycetales (4%) contain expansins with signal peptides for lipid attachment and membrane localization (in the case of Micromonosporales and Streptomycetales the lipidation signal peptide is not conserved within closely related species and may be false positives).

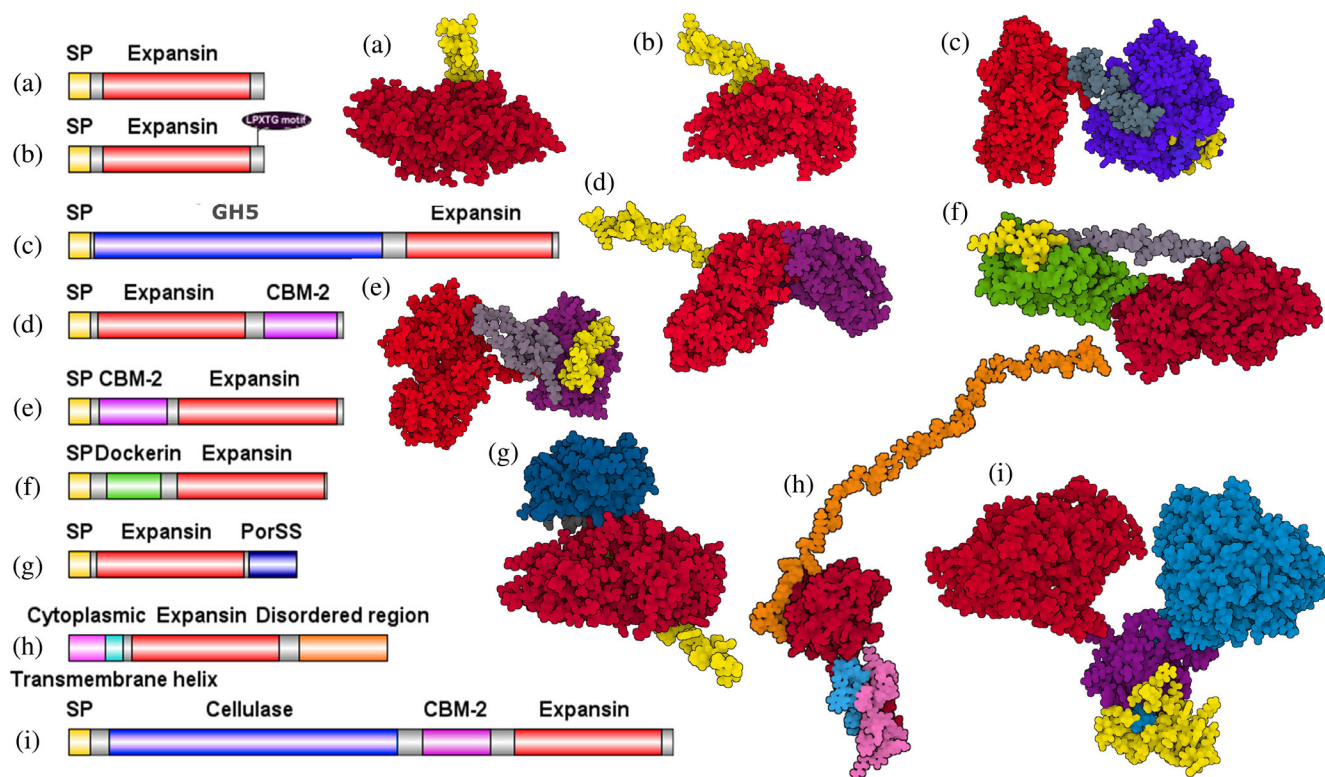


FIGURE 5 Schematic examples of primary and tertiary structures of the most abundant modular expansin domain architectures (N- to C-terminal). (a) Monomodular expansin (UniProt Id: C6DHV9). (b) Monomodular expansin with a putative LPXTG-like sorting motif (UniProt Id: D3JXR9; PDB: 3D30). (c) Modular Expansin with an N-terminal cellulase domain (UniProt Id: Q8P513). (d) Modular expansin with a C-terminal CBM2 domain (UniProt Id: F0Q783). (e) Modular expansin with an N-terminal CBM2 domain (UniProt Id: A9FSQ7). (f) Modular expansin with an N-terminal dockerin domain (UniProt Id: G8LUU0). (g) Expansin with a C-terminal PorSS (Type-IX secretion system) domain (UniProt Id: A0A2N3TEV7). (h) Expansin with N-terminal cytoplasmic domain, a transmembrane helix, and a C-terminal highly polar disordered tail (UniProt Id: A0A010ZNG5). (i) Modular expansin with both N-terminal GH5 domain and a CBM2 domain (UniProt Id: A5CLK3). Tertiary structures were predicted with trRosetta and visualized in the Mol* web server. The signal peptide (in yellow) is cleaved upon membrane translocation and was not excised for structure prediction. CBM2, family 2 carbohydrate-binding module; GH5, family 5 glycoside hydrolase. No structures have been experimentally determined for multimodular expansins

Interestingly, expansins from Corynebacteriales, which resemble diderm bacteria due to a second cell membrane composed of mycolic acids, are acidic expansins (Figure 4). This finding is consistent with the correlation of diderm bacteria containing negatively charged expansins, whereas the monoderm cell surface structure favors positively charged expansins, but there are some species that contain an acidic expansin, like in *Streptomyces acidiscabies* and *Kutzneria sp. 744*.³³ Similarly, expansins from the order Cryptosporangiales are also acidic proteins but predicted to be part of transmembrane proteins, as they are fused with an N-terminal cytoplasmic domain, which is a transmembrane domain. Also, they present a C-terminal disordered region predicted to be extracellular (Figure 5h), which could bind another protein, as a binding region is predicted (Figure S3). Another possibility is that the disordered domain exists to stabilize the protein, as it has been reported that

disordered regions in transmembrane proteins provide structural flexibility to stabilize interactions with negatively charged head groups of the lipid molecules,³⁴ which seems plausible considering these are negatively charged expansins. *Clavibacter* species from the Order Micrococcales contain modular GH5-CBM2-expansins (Figures 4 and 5i) and have been the focus of research on their involvement in the pathogenesis of *Clavibacter michiganensis*, and because of a possible synergic enhancement of cellulose degradation by commercial cellulases, where they show a synergistic effect with *Trichoderma reesei* cellulase on polysaccharide hydrolysis.^{7,10,35}

Streptomycetales and Micromonosporales are the orders with the most expansins (198 and 134 expansins, respectively), with 99.5% being monomodular expansins and 86% predicted to be basic expansins. The relative abundance of expansins in these orders could be due to

strain overrepresentation in sequence databases, as these are important secondary-metabolite producers, but it is interesting that plant pathogenic species such as *Streptomyces scabiei*, *Streptomyces stelliscabiei*, *Streptomyces turgidiscabies*, and *Streptomyces ipomoeae* contain expansin genes in their genomes (Table SS1).

2.4 | Variety of domain architectures in bacterial modular expansins

To understand the influence of extra modules in the evolution and selection of the isoelectric point of the expansin domain, we also searched for relationships of bacterial order, electric charge, and the diversity of modular expansins (Figure 5). The general trend of acidic expansin-diderm organisms differs in some cases, such as in (a) the modular basic expansins of the diderm order Xanthomonadales (Figure 2), (b) the acidic expansins of the monoderm Clostridial cellulosome dockerin-expansins (Figure 3), (c) the basic monomodular expansins in *Ralstonia* species (Figure 2), and (d) acidic expansins from Corynebacteriales. Although, experimental data are inconclusive for significant enzymatic enhancement between expansins and cellulose degrading enzymes,^{8,10,12,16,26,35,36} fusions between expansins and domains related to polysaccharide metabolism are common in bacteria, and are conserved within the groups where they appear (Figure 1). In fact, the most abundant modular expansins are expansin-family 5-glycosylhydrolase domain belonging to Xanthomonadales (GH5-Exp, Figure 5c); expansin-carbohydrate-binding domain at either N- or C-end, in several orders (Figure 5d,e); expansin-dockerin domain in Clostridiales (Figure 5f); and expansins with both GH5 and CMB2 domains from *Clavibacter* genus (Figure 5I), which directly relate to the metabolism of cellulose, as previously reported.²⁵ Complementing Chase et al. findings, our results show other combinations of polysaccharide-metabolism-related domains (GH9, CBM5, CBM6, and Galactose-binding-like domains) fused with expansins that are found less frequently (Figure SS1), possibly having took place quite recently, or that the evolutionary advantage they confer on each species is minimal compared with the most conserved modular expansins. It is interesting that despite finding just a few sequences in Cyanobacteria and Verrucomicrobia most were modular, with rare or even unique domain architectures: the expansins in Verrucomicrobia that are fused with a Galactose-binding-like domain (Figure SS1); the GH9-CBM2-expansin in Nostocales; a double CBM5 domain fusion (CBM5-CBM5-expansin) in Pleurocapsales (Figure 2); a unique domain architecture with three

CBM5 repeats and two flanking expansin domains (expansin-CBM5[x3]-expansin) in a *Flavobacteria* species (Figure 2); and a unique domain architecture composed of an expansin fused to cell wall/choline-binding repeats in a *Ruminococcus* species (Figure 3). Altogether, our results suggest specific roles for expansins in different species, which in turn would explain the large diversity of existing expansins.

3 | DISCUSSION

Our updated analysis of prokaryotic expansin sequences suggests that expansin evolution may be driven by the characteristics of the bacterial cell envelope type and by domain fusion, which may have different functions or targets depending on the ecological niche and physiology of each species. Information derived from expansin signal peptide sequences and localization motifs, which has been overlooked in other studies, were most informative as they shed light on the export pathways of expansins among bacteria. Expansin export was expected as their substrate in plants is the cell wall, the existence of secretion signal peptides and conserved residues for plant cell wall binding. Thus, it was assumed that expansins are exported and solubilized into the surrounding media, and indeed, experimental data show microbial expansin binding of plant tissues and cellulose. However, these experiments used soluble expansins heterologously produced in *E. coli*. Our data indicate that endogenous bacterial expansins might not solubilize completely after secretion and instead remain associated with the bacteria through different mechanisms (Figure 6): (a) the secretion type through which expansins are translocated across the membranes in diderms, as 143 expansin sequences contain signal peptides for secretion through the Sec system-peptidase II, which processes expansins to become lipoproteins; (b) by the presence of sorting domains in Fibrobacteres and Bacteroidetes predicting expansin attachment to the cell exterior via the T9SS by covalently binding an A-LPS to the protein²⁹; (c) the presence of C-terminal LPXTG-like motifs and a hydrophobic C-terminal tail that are potential substrates for sortases, linking the protein to peptidoglycan in 168 monoderm expansins; and (d) expansin fusion to domains that anchor them to scaffolding as part of larger multiprotein complex, such as the cellulosome. All mechanisms combined predict that up to 39% of expansins localize to the cell surface after membrane translocation, suggesting that their function is required within the surface or at the cell proximity. As ever, it is important to stress that these predictions and interpretations must be investigated experimentally, as sequence divergence between related

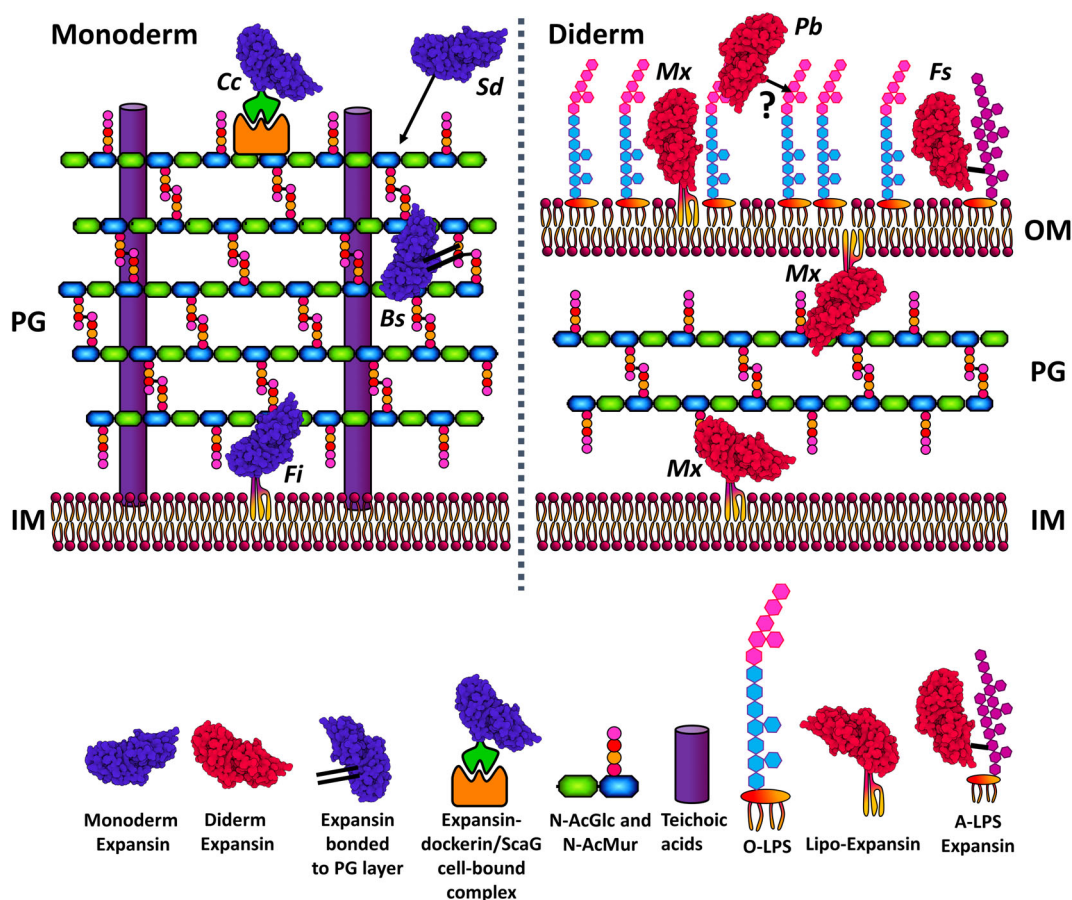


FIGURE 6 Schematic representation of diverse strategies to associate expansins to the bacterial cell surface. Representative expansins from: *Bs*, *B. subtilis*; *Pb*, *Pectobacterium brasiliense* (for which binding properties and its ligand are yet unknown⁴); *Fs*, *Fibrobacter succinogenes* (a A-LPS is attached to the expansin by the T9SS); *Mx*, *Myxococcus xanthus* (lipo-expansins could attach to either membrane in diderms); *Cc*, *Clostridium clariflavum* (associated with the cell-bound scaffoldin protein ScaG); *Sd*, *Streptomyces davaonensis* (expansins have high affinity for peptidoglycan); *Fi*, *Frankia inefficax*. OM, outer membrane. PG, peptidoglycan layer. IM, inner membrane

bacteria, and the lack of a better understanding of all signal peptides and sorting signals (as all programs currently available are trained with a subset of model organisms) could obscure the accurate prediction of sequences or give false positives. For sequences without information for surface localization, our previous observations show that Exl1 is retained in the pellet of bacterial cultures in LB media.⁴ Whether Exl1 binds to an exopolysaccharide, a lipopolysaccharide, or another ligand via its polysaccharide binding surface or through electrostatic interactions remain to be confirmed, but also opens the possibility that expansin localization on the cell exterior is a more frequent phenomenon.

What function expansins play on the cell envelope remains to be answered, but we propose two possibilities: a first in which expansin activity is required for remodeling or stabilizing the bacterial cell wall of some species in a process similar or equivalent to plant expansins, or a second in which expansin density contribute to the modulation of the electrostatic or hydrodynamic properties of

the cell given their highly charged and hydrophilic nature. The electrical properties of the cell exterior might affect bacterial processes or have implications for bacteria interacting with plants in the case of pathogens or plant colonizing species such as *B. subtilis*, and also for free living bacteria such as the Myxococcales that exhibit a multicellular behavior dependent on exopolysaccharide production.^{37–39} Depending on the taxonomic group, expansins could play different or combined roles: act on their own cell wall, act on the plant cell wall, be part of the cellulosome for plant biomass digestion, or contribute to the modulation of the electrical properties of the bacterial cell envelope. Studying different expansins from closely related bacteria (the Burkholderiales, for instance) could help identify the underlying causes for bacteria preferences over specific expansin types; or the high sequence similarity between expansins from Proteobacteria and Firmicutes (attributed to horizontal gene transfer events^{25,40}), which could provide an advantage to the recipient species.⁶ Whether more than one function

affects bacterial physiology must be determined, as they are not mutually exclusive. For instance, we have previously observed that soluble acidic Exl1 from *P. brasiliense* and soluble basic EXLX1 from *B. subtilis* have distinctive binding patterns in celery cell walls,⁵ whereby Exl1 bound to the cell wall corners surrounding xylem vessels, opening the possibility that localization of expansin-bound bacteria would be driven by expansin expression, that is, expansin affinity for a cell corner ligand determines bacterial localization, which until now remains an open question. This would also explain the recurrence of expansins in xylem-invading phytopathogens. Additionally, we have previously reported an *exl1* null mutant strain of *Pectobacterium atrosepticum* that is defective in swarming motility,⁴ a plant-independent process with requirement for hydration and therefore susceptible to the density of hydrophilic molecules, such as the expansins, on the cell surface. In support of this, Valdivia et al. reported that among different phenotypes of an expansin *expb1* mutant of maize pollen, a dehydration defect occurs after anther shedding.⁴¹ Modification of the cell surface properties alone could influence other phenotypes in bacteria that express variable expansin concentrations. The only other reported expansin mutant that shows a plant-independent phenotype is EXLX1 from *B. subtilis*, which renders bacteria more susceptible to autolysis, possibly by reducing stabilization to the peptidoglycan layer to which it might be covalently bound by sortases.¹⁸ We also found that the majority (74%) of expansins belonging to Myxococcales are predicted to be monomodular acidic lipo-expansins, which is the highest proportion of lipo-expansins in all orders (Figure 2), suggesting that expansins in this order need to be embedded in the membrane. It would be interesting to study the biological role of lipo-expansins in a bacterial social context, such as that of free-living species *Myxococcus xanthus* (acidic lipo-expansin) and *Sorangium cellulosum* (acidic lipo-CBM2-expansin), which exhibit a multicellular behavior, to determine whether expansins influence colony hydrodynamics or cell-to-cell or cell-to-surface interactions, having a role during the swarming motility of *P. brasiliense*⁴ and in the stalk formation of *Dictyostelium*.⁴² The lipo-expansins from Frankiales, which are nitrogen-fixating and plant symbiotic Actinobacteria,⁴³ could be playing an important part in plant colonization, similar to other plant-growth promoting bacteria like *B. subtilis*,¹⁸ and could hold relevant information on how lipo-expansins function in Gram-positive bacteria.

By analyzing a large number of expansin sequences, we found a weaker correlation between expansin electric charge and bacteria type compared with a previous report with fewer sequences.¹³ Interestingly, basic expansins

belonging to Xanthomonadales and *Clavibacter* species are fused to a glycosyl hydrolase (GH5) domain (Figure 5c,i), thus, an expansin basic nature may be required for the functionality of the GH5 domains, explaining their presence in Gram-negative bacteria such as *Xanthomonas*. These expansins have been studied for synergistically enhancing the enzymatic efficiency of cellulose hydrolysis,^{10,16,35,36,44,45} but their mechanism in vivo must be analyzed under different conditions as it is unclear whether, in native conditions, modular expansins enhance the activity of adjacent enzymes or whether they provide other unknown functionality, such as substrate selectivity, substrate alignment, or cell adhesion to the substrate (this could be the case for the dockerin-expansins of *Clostridium* species¹⁵). Analysis of monomodular and modular expansin activity would provide additional information if studied under natural conditions and at the time it might be required in each species, as it is possible that expansin requirement occurs at specific developmental stages, and therefore the phenotypes of expansin null mutants might be cryptic and species specific. Thus, it is necessary to develop new experimental procedures for bacterial expansin research considering that they could be associated with the bacteria exterior. For example, analyzing how *Xanthomonas* or *Clavibacter* utilize different polysaccharides as their only carbon source, in a comparison of wild-type versus expansin null-mutants, could tell if modular expansins containing a cellulase domain help bacteria metabolize complex carbohydrates more efficiently. At the same time, determination of the mechanism in phenotypes nonrelated to plant infection of null-mutants and expansin overexpressing strains, like the aberrant stalk formation in *Dictyostelium* spp.,⁴² the autolysis susceptibility in *B. subtilis*¹⁸ and the swarming motility of *P. atrosepticum* and *P. carotovorum*⁴ could give light on the expansin function in these microbes. A comprehensive experimental study that corroborates our findings is needed, taking into consideration all types of bacterial expansins and different cell envelope composition in bacteria.

4 | MATERIALS AND METHODS

4.1 | Expansin search and dataset generation

We constructed an HMM profile using 44 excised expansin domains using the *hmmbuild* algorithm from the HMMER v3.3.2 software package⁴⁶ and conducted a search on nonredundant bacterial proteomes from UniProt database⁴⁷ (accessed on November 2020) with a

score cutoff of 70 and a sequence length threshold of 190, to filter any false-positive sequences with high similarity to either expansin subdomains. The resulting sequences were manually curated by checking subdomain organization (D1–D2) and conserved sites. For each protein sequence, taxonomic and related information was retrieved from the UniProt database (available at <https://www.uniprot.org/uploadlists/>) and additional domain annotation was retrieved from the InterProScan web service⁴⁸ (available at <https://www.ebi.ac.uk/interpro/search/sequence/>). The resulting expansin database is provided in Table S1.

4.2 | Cladogram generation and bioinformatic analysis

Excised expansin domains (portion of each protein that aligned to the expansin HMM profile) were aligned with ClustalO, and a maximum likelihood tree was constructed with IQ-TREE using automatic substitution model selection and 1000 ultrafast bootstrap iterations.⁴⁹ The resulting cladogram was visualized and edited with iTol web server.⁵⁰ To determine the major sequence groups, we constructed a heat map with the all-vs-all pairwise alignment matrix using ClustalO with the flags --distmat-out, --full, and ----percent-id. Isoelectric point calculation was estimated for the excised expansin domains (~220 aa) using the IPC v2.0 algorithm, taking the average isoelectric point between all models.⁵¹ Signal peptides were predicted using SignalP v5.0 with default parameters.⁵² C-terminal regions of proteins were scanned for reported LPXTG motifs²³ with the FIMO software from the MEME Suite toolkit using default parameters and an *e*-value cutoff of $1e-5$.⁵³ For representative expansin proteins, tertiary structure prediction was performed with trRosetta webserver and visualized with Mol* program (using Spacefill representation and plastic material style preset). Each domain color represents the domain information available in Interpro.^{54,55} Prediction of disordered domains was done with IUPred2 (long) and ANCHOR2.⁵⁶

ACKNOWLEDGEMENTS

This project received financial support from DGAPA-PAPIIT (IN211019 and IN203622) and CONACYT Ciencia Básica (252551) to C.M.-A. CONACYT provided scholarships to A.S., J.J.S.-C., and I.T.-V. We are grateful to Chris D. Wood for critical reading of this manuscript.

AUTHOR CONTRIBUTIONS

Andrés De Sandozequi: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation

(equal); writing – original draft (lead); writing – review and editing (supporting). **Juan José Salazar-Cortés:** Conceptualization (supporting); writing – review and editing (supporting). **Irán Tapia-Vázquez:** Investigation (equal); writing – review and editing (supporting). **Claudia Martínez-Anaya:** Conceptualization (equal); funding acquisition (lead); project administration (lead); writing – original draft (equal); writing – review and editing (lead).

ORCID

Andrés de Sandozequi  <https://orcid.org/0000-0002-7387-9883>

Irán Tapia-Vázquez  <https://orcid.org/0000-0003-0399-0751>

Claudia Martínez-Anaya  <https://orcid.org/0000-0002-2483-6115>

REFERENCES

1. Cosgrove DJ. Plant expansins: Diversity and interactions with plant cell walls. *Curr Opin Plant Biol.* 2015;25:162–172.
2. Cosgrove DJ. Catalysts of plant cell wall loosening. *F1000Res.* 2016;5:119.
3. Tanco MA, Lowe-Power TM, Peritore-Galve FC, Tran TM, Allen C, Smart CD. Plant-like bacterial expansins play contrasting roles in two tomato vascular pathogens. *Mol Plant Pathol.* 2018;19:1210–1221.
4. Narváez-Barragán DA, Tovar-Herrera OE, Torres M, et al. Expansin-like Ex11 from *Pectobacterium* is a virulence factor required for host infection, and induces a defence plant response involving ROS, and jasmonate, ethylene and salicylic acid signalling pathways in *Arabidopsis thaliana*. *Sci Rep.* 2020;10:1–14.
5. Tovar-Herrera OE, Rodríguez M, Olarte-Lozano M, et al. Analysis of the binding of Expansin Ex11, from *Pectobacterium carotovorum*, to plant xylem and comparison to EXLX1 from *Bacillus subtilis*. *ACS Omega.* 2018;3:7008–7018.
6. Rocha J, Shapiro LR, Kolter R. A horizontally acquired expansin gene increases virulence of the emerging plant pathogen *Erwinia tracheiphila*. *Sci Rep.* 2020;10:1–15.
7. Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R (2000). The Endo- β -1,4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* Is a Pathogenicity Determinant Required for Induction of Bacterial Wilt of Tomato. *Mol Plant Microbe Interact.*, 13, , 703. –714.
8. Georgelis N, Nikolaidis N, Cosgrove DJ. Bacterial expansins and related proteins from the world of microbes. *Appl Microbiol Biotechnol.* 2015;99:3807–3823.
9. Cosgrove DJ. Microbial expansins. *Annu Rev Microbiol.* 2017; 71:479–497.
10. Bunterngsook B, Mhuantong W, Champreda V, Thamchaiphenet A, Eurwilaichitr L. Identification of novel bacterial expansins and their synergistic actions on cellulose degradation. *Bioresour Technol.* 2014;159:64–71.
11. Georgelis N, Nikolaidis N, Cosgrove DJ. Biochemical analysis of expansin-like proteins from microbes. *Carbohydr Polym.* 2014;100:17–23.

12. Chen C, Cui Z, Song X, Liu YJ, Cui Q, Feng Y. Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation. *Appl Microbiol Biotechnol*. 2016;100:2203–2212.
13. Pastor N, Dávila S, Pérez-Rueda E, Segovia L, Martínez-Anaya C. Electrostatic analysis of bacterial expansins. *Proteins Struct Funct Bioinform*. 2015;83:215–223.
14. Olarte-Lozano M, Mendoza-Núñez MA, Pastor N, Segovia L, Folch-Mallol J, Martínez-Anaya C. PcEx11 a novel acid expansin-like protein from the plant pathogen *Pectobacterium carotovorum*, binds cell walls differently to BsEXLX1. *PLoS One*. 2014;9:e95638.
15. Artzi L, Morag E, Shamshoum M, Bayer EA. Cellulosomal expansin: Functionality and incorporation into the complex. *Biotechnol Biofuels*. 2016;9:61.
16. Lee HJ, Kim IJ, Kim JF, Choi I-G, Kim KH. An expansin from the marine bacterium *Hahella chejuensis* acts synergistically with xylanase and enhances xylan hydrolysis. *Bioresour Technol*. 2013;149:516–519.
17. Georgelis N, Tabuchi A, Nikolaidis N, Cosgrove DJ. Structure-function analysis of the bacterial expansin EXLX1. *J Biol Chem*. 2011;286:16814–16823.
18. Kerff F, Amoroso A, Herman R, et al. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proc Natl Acad Sci USA*. 2008;105:16876–16881.
19. Green ER, Meccas J. Bacterial secretion systems: An overview. *Virulence Mech Bact Pathog*. 2016;4:213–239.
20. Costa TRD, Felisberto-Rodrigues C, Meir A, et al. Secretion systems in gram-negative bacteria: Structural and mechanistic insights. *Nat Rev Microbiol*. 2015;13:343–359.
21. Tseng TT, Tyler BM, Setubal JC. Protein secretion systems in bacterial-host associations, and their description in the gene ontology. *BMC Microbiol*. 2009;9:1–9.
22. De Diego I, Ksiazek M, Mizgalska D, et al. The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal β -sandwich domain. *Sci Rep*. 2016;6:1–17.
23. Malik A, Kim SB. A comprehensive in silico analysis of sortase superfamily. *J Microbiol*. 2019;57:431–443.
24. Lohoff C, Buchholz PCF, Le Roes-Hill M, Pleiss J. Expansin engineering database: A navigation and classification tool for expansins and homologues. *Proteins*. 2020;89:149–162.
25. Chase WR, Zhaxybayeva O, Rocha J, Cosgrove DJ, Shapiro LR. Global cellulose biomass, horizontal gene transfers and domain fusions drive microbial expansin evolution. *New Phytol*. 2020;226:921–938.
26. Martínez-Anaya C. Understanding the structure and function of bacterial expansins: A prerequisite towards practical applications for the bioenergy and agricultural industries. *J Microbiol Biotechnol*. 2016;9:727–736.
27. Sato K, Naito M, Yuki take H, et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci USA*. 2010;107:276–281.
28. Gorasia DG, Veith PD, Chen D, et al. *Porphyromonas gingivalis* type IX secretion substrates are cleaved and modified by a Sortase-like mechanism. *PLoS Pathog*. 2015;11:1–31.
29. Veith PD, Glew MD, Gorasia DG, Reynolds EC. Type IX secretion: The generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Mol Microbiol*. 2017;106:35–53.
30. Douillard FP, Rasinkangas P, Von Ossowski I, Reunanen J, Palva A, De Vos WM. Functional identification of conserved residues involved in *Lactobacillus rhamnosus* strain GG sortase specificity and pilus biogenesis. *J Biol Chem*. 2014;289:15764–15775.
31. Dai X, Böker A, Glebe U. Broadening the scope of sortagging. *RSC Adv*. 2019;9:4700–4721.
32. Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. *Clostridium clariflavum*: Key cellulosome players are revealed by proteomic analysis. *MBio*. 2015;6:1–12.
33. Armijos-Jaramillo V, Santander-Gordón D, Tejera E, Perez-Castillo Y. The dilemma of bacterial expansins evolution. The unusual case of *Streptomyces acidiscabies* and *Kutzneria* sp. 744. *Commun Integr Biol*. 2018;11:1–10.
34. Tusnády GE, Dobson L, Tompa P. Disordered regions in transmembrane proteins. *Biochim Biophys Acta Biomembr*. 2015;1848:2839–2848.
35. Bunterngsook B, Eurwilaichitr L, Thamchaipenet A, Champreda V. Binding characteristics and synergistic effects of bacterial expansins on cellulosic and hemicellulosic substrates. *Bioresour Technol*. 2015;176:129–135.
36. Junior AT, Dolce LG, de Oliveira NM, Polikarpov I. *Xanthomonas campestris* expansin-like X domain is a structurally disordered beta-sheet macromolecule capable of synergistically enhancing enzymatic efficiency of cellulose hydrolysis. *Biotechnol Lett*. 2015;37:2419–2426.
37. Wu Y, Jiang Y, Kaiser D, Alber M. Social interactions in myxobacterial swarming. *PLoS Comput Biol*. 2007;3:2546–2558.
38. Schneiker S, Perlova O, Kaiser O, et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat Biotechnol*. 2007;25:1281–1289.
39. Keane R, Berleman J. The predatory life cycle of *Myxococcus xanthus*. *Microbiol (United Kingdom)*. 2016;162:1–11.
40. Nikolaidis N, Doran N, Cosgrove DJ. Plant expansins in bacteria and fungi: Evolution by horizontal gene transfer and independent domain fusion. *Mol Biol Evol*. 2014;31:376–386.
41. Valdivia ER, Stephenson AG, Durachko DM, Cosgrove D. Class B β -expansins are needed for pollen separation and stigma penetration. *Sex Plant Reprod*. 2009;22:141–152.
42. Ogasawara S, Shimada N, Kawata T. Role of an expansin-like molecule in *Dictyostelium* morphogenesis and regulation of its gene expression by the signal transducer and activator of transcription protein Dd-STATA: Original article. *Dev Growth Differ*. 2009;51:109–122.
43. Gtari M, Ghodhbane-Gtari F, Nouioui I. *Frankia soli* sp. nov., an actinobacterium isolated from soil beneath *Ceanothus jepsonii*. *Int J Syst Evol Microbiol*. 2020;70:1203–1209.
44. Pech-Cervantes AA, Ogunade IM, Jiang Y, et al. An expansin-like protein expands forage cell walls and synergistically increases hydrolysis, digestibility and fermentation of livestock feeds by fibrolytic enzymes. *PLoS One*. 2019;14:1–23.
45. Kim ES, Lee HJ, Bang WG, Choi IG, Kim KH. Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. *Biotechnol Bioeng*. 2009;102:1342–1353.
46. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 2009;23:205–211.

47. Bateman A, Martin MJ, Orchard S, et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49: D480–D489.
48. Jones P, Binns D, Chang HY, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–1240.
49. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32: 268–274.
50. Letunic I, Bork P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44:W242–W245.
51. Kozłowski LP. IPC—Isoelectric point calculator. *Biol Direct.* 2016;11:1–16.
52. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37:420–423.
53. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. *Bioinformatics.* 2011;27:1017–1018.
54. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci USA.* 2020;117: 1496–1503.
55. Sehnal D, Bittrich S, Deshpande M, et al. Mol * Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* 2021;49:W431–W437.
56. Mészáros B, Erdős G, Dosztányi Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 2018;46:W329–W337.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: de Sandozequi A, Salazar-Cortés JJ, Tapia-Vázquez I, Martínez-Anaya C. Prevalent association with the bacterial cell envelope of prokaryotic expansins revealed by bioinformatics analysis. *Protein Science.* 2022;31(5):e4315. <https://doi.org/10.1002/pro.4315>

1 Artículo de revisión enviado a la revista MicrobiologyOpen y en proceso de revisión

2 **Bacterial surface-associated proteins in plant infection**

3 Andrés de Sandozequi and Claudia Martínez-Anaya*

4

5 Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, UNAM.

6 Av. Universidad 2001, Chamilpa. 62210, Cuernavaca. México

7

8

9

10 *Corresponding author. E-mail: claudia.martinez@ibt.unam.mx

11 Andrés de Sandozequi <https://orcid.org/0000-0002-7387-9883>

12 Claudia Martínez-Anaya <https://orcid.org/0000-0002-2483-6115>

13

14

15

16

17 Keywords: Surface proteins, Surfaceome, Phytopathogenic bacteria, Bacterial cell surface,
18 Lipoproteins, Sortase.

19

20

21

22 **Abstract**

23 The bacterial cell envelope is involved in all stages of infection and the study of its components and
24 structures are an important step to understand how bacteria interact with the extracellular milieu.
25 Thanks to new techniques that focus on identifying bacterial surface proteins, we now better
26 understand the specific components involved in host-pathogen interactions. In the fight against
27 deleterious effects of pathogenic bacteria, bacterial surface proteins (at the cell envelope) are
28 important targets as they play crucial roles in colonization and infection of host tissues. These surface
29 proteins serve functions such as protection, secretion, biofilm formation, nutrient intake, metabolism,
30 and virulence. Bacteria use different mechanisms to associate proteins to their cell envelope including
31 post-translational modifications, extracellular trafficking by sortases, post-translational attachment of
32 lipopolysaccharide A, attachment to the extracellular matrix, or the presence of protein domains that
33 allow binding to other components or embedding into membranes. In the case of phytopathogenic
34 bacteria, surface-associated proteins are often involved in several processes; some include plant
35 surface attachment and colonization, signal transduction, cell to cell communication, migration, plant
36 cell wall degradation, nutrient intake, protection from the host defense response and general
37 housekeeping functions of the cell. In this review, we focus on proteins that are associated to the
38 bacterial cell envelope by post-translational modifications and their roles in plant infection.

39 **Introduction**

40 Bacterial infectious disease progression involves a stage of surface colonization during which cells
41 adapt and produce proteins involved in pathogenesis. Technological advances for the analysis of the
42 bacterial surfaceome and proteosurfaceome (Desvaux et al., 2018; Rodríguez-Ortega, 2018), have
43 facilitated identifying proteins relevant to virulence and infection, yet their importance extends
44 beyond characterizing these processes as they are potential drug targets and candidates for protein-

45 based vaccines. In the context of bacterial phytopathogens, surface proteins have been somewhat
46 overlooked, as most of them are not considered effectors, but understanding them could lead to
47 breakthroughs in the fight against phytopathogenic bacteria. The bacterial cell envelope is the first
48 structure to interact with plant tissues, and it often serves as the furthestmost line of attack to colonize
49 and infect plant hosts. The cell envelope is composed of a variety of molecules, including
50 polysaccharides, lipids, and proteins (Bogino et al., 2013; Silhavy et al., 2010; Viljoen et al., 2020),
51 and of these, the proteins have numerous functions in protection, secretion, biofilm formation,
52 nutrient intake, metabolism, and virulence. Bacteria use different mechanisms to associate proteins to
53 their cell envelope, such as: i) extracellular trafficking by sortases ([Figure 1](#)) (Dai et al., 2019); ii)
54 post-translational modification, for example adding a lipid moiety to create membrane-bound
55 lipoproteins ([Figure 2](#)) (Cole et al., 2021; Kovacs-Simon et al., 2011; Narita and Tokuda, 2017); iii)
56 post-translational attachment of a lipopolysaccharide A (A-LPS) by the type XI secretion system
57 present in the superphylum Fibrobacteres-Chlorobi-Bacteroidetes (De Diego et al., 2016; Veith et al.,
58 2017); iv) by attachment to the extracellular matrix (such as the S-layer in monoderm bacteria, [Figure](#)
59 [1](#)) (Desvaux et al., 2006; Flemming and Wingender, 2010); or v) presence of protein domains that
60 allow binding to other components or embedding into membranes (for instance: pore-forming
61 domains, cell wall binding domains and transmembrane helices, [Figure 2](#)) (Bateman et al., 2021;
62 Desvaux et al., 2018; Rodríguez-Ortega, 2018). In phytopathogenic bacteria, surface-associated
63 proteins accomplish physiological and invasive roles (Cole et al., 2021). Different studies show
64 involvement of the bacterial surface in pathogenesis (Bogno et al., 2013; Desvaux et al., 2018;
65 Pfeilmeier et al., 2016; Rodríguez-Ortega, 2018) mainly centered on human pathogens, but there is
66 an increased interest in how the bacterial surface is involved in the process of infection by plant
67 pathogenic bacteria.

68 **Surface-exposed proteins anchored to the peptidoglycan layer.**

69 Monoderm bacteria comprehend only one lipid bilayer and a thick peptidoglycan layer. In contrast to
70 diderm bacteria that have two membranes and a peptidoglycan layer in between, in monoderms the
71 peptidoglycan layer is the outermost structure interacting with the exterior, and there, surface proteins
72 play important roles. Monoderms attach proteins to the peptidoglycan layer in a unique manner via
73 the sortase system. Sortases are peptidoglycan-bound proteins that recognize a motif in the C-terminal
74 region of proteins. The most studied is the LPXTG motif, which is recognized by sortases conserved
75 across multiple bacterial clades. Sortases recognize and cleave this motif between the Thr and Gly
76 residues and anchor the LPXTG-containing protein to the Lipid II molecule, which is the building
77 block of the growing peptidoglycan layer (Comfort and Clubb, 2004; Dai et al., 2019; Desvaux et al.,
78 2006; Dramsi et al., 2008; Hendrickx et al., 2011; Navarre and Schneewind, 1994; Novick, 2000;
79 Pishesha et al., 2018). Indeed, sortase-mediated surface anchoring of proteins are important for some
80 bacterial pathogens such as staphylococci, streptococci, enterococci and *Listeria monocytogenes* (Dai
81 et al., 2019; Egan et al., 2010; Hendrickx et al., 2009; Hendrickx et al., 2011). In *Clavibacter*
82 *michiganensis* ssp. *michiganensis*, an important pathogen of tomatoes and potatoes, sortase SrtA is
83 essential for blister formation which is a hallmark of the infectious process, but not for wilting
84 symptoms, suggesting that it is important for the attachment of bacteria to the leaf surface
85 (Chalupowicz et al., 2017). Furthermore, an important virulence factor of this pathogen, the cell wall-
86 bound protease Pat-1 -a substrate of SrtA- is absent in some non-virulent isolates of *Clavibacter*
87 species (Gartemann et al., 2003). In *Bacillus* (many of which are plant growth-promoting bacteria
88 and in some cases opportunistic pathogens) sortases are vital for plant colonization and are involved
89 in pili and endospore biosynthesis (Budzik et al., 2007). Inhibition of sortases as a therapeutic target
90 in multi-drug resistant pathogens has been a hot topic (Cascioferro et al., 2014; Cossart and
91 Jonquieres, 2000; Kudryavtsev et al., 2021; Ouyang et al., 2018; Si et al., 2016), and could prove
92 useful in dealing with phytopathogenic bacteria, but more research is needed to fully understand all
93 the implications that sortases and their substrates have for plant disease.

94 **Lipoproteins in plant pathogenic bacteria.**

95 Diderm and monoderm bacteria produce lipoproteins, being the periplasm the predominant
96 localization site for most lipoproteins in diderms ([Figure 1](#) and 2). Although in recent years, there has
97 been an increased discovery of surface-exposed lipoproteins in diderms after the identification of
98 surface lipoprotein traslocons such as SLAM and Bam (Huynh et al., 2022; Kovacs-Simon et al.,
99 2011; El Rayes et al., 2021; Remans et al., 2010; Wilson and Bernstein, 2016), which leads us to
100 question the implications of these proteins in cell function. For instance, from the 175 predicted
101 lipoproteins in the genome of *Pseudomonas aeruginosa*, more than half are classified as hypothetical
102 proteins for which a function still needs to be determined, as thus, their involvement in pathogenesis
103 (Nguyen et al., 2020; Remans et al., 2010). Similar numbers were predicted, at around 2% lipoprotein
104 genes per genome (Babu et al., 2006; Teufel et al., 2022) for the most important plant pathogens
105 (Mansfield et al., 2012). As our knowledge of lipoprotein signal peptides grows and more accurate
106 algorithms for their prediction are developed, we will continue finding more lipoproteins that are
107 essential to the bacterial physiology, but potentially, also for the interaction with their hosts (Table
108 1). In comparison, the DOLOP database, which was the first comprehensive database for lipoproteins,
109 had certain limitations in recognizing lipoprotein signal peptides across various bacterial species
110 (Babu et al., 2006; Babu and Sankaran, 2002). More available genetic information and knowledge of
111 the properties of signal peptides has improved the development of better software for their analysis.
112 The most advanced algorithm so far that predicts all known types of signal peptides (including those
113 not yet known when the DOLOP database was created), is SignalP which is currently in its 6.0
114 version. SignalP v6.0 has introduced a novel machine learning method that recognizes and classifies
115 five different types of signal peptides, and it is able to predict signal peptides in distantly related
116 sequences or metagenomic sequences from unknown origin (Almagro Armenteros et al., 2019; Teufel
117 et al., 2022).

118 Lipoproteins are involved in processes such as maintaining bacteria cell envelope integrity
119 (Dubuisson et al., 2005; Hellman et al., 2002; Lazzaroni and Portalier, 1992), cellular morphology
120 (Paulsson et al., 2021), drug efflux (Li et al., 1995), iron uptake and protein secretion (Alvarez-
121 Martinez et al., 2021; Green and Meccas, 2016; El Rayes et al., 2021). Some lipoproteins are part of
122 cellular appendages such as flagella and pili (Jin et al., 2001), while others are integral components
123 of the peptidoglycan biosynthesis machinery and transport systems for effectors, proteins,
124 lipopolysaccharides, and antibiotics (Lorenz et al., 2019; Okuda and Tokuda, 2011). In some cases
125 lipoproteins are enzymes that have proteolytic and plant cell wall-degrading activities (Liao et al.,
126 2022). Interestingly, the first lipoprotein to be discovered is the most numerically abundant protein
127 in *E. coli*, Lpp, also called murein lipoprotein or Braun's lipoprotein (Braun and Rehn, 1969). There
128 can be up to a million Lpp molecules in a single cell and it is also one of the most studied over the
129 years (Asmar and Collet, 2018). Lpp seems to exist in two forms ([Figure 1](#)): a 'bound' form, wherein
130 about a third of the Lpp molecules are covalently bonded to the PG and control the width of the
131 periplasmic space; and a surface exposed 'free' form, whose role is still unknown (Asmar et al., 2017;
132 Choi et al., 1986; Cowles et al., 2011; Dramsi et al., 2008).

133 Besides the roles described for lipoproteins in model bacteria and human pathogens, phytopathogenic
134 bacteria often rely on lipoproteins to attach to their preferred biological substrates, which is an
135 important step in host colonization. The LolABCDE system plays a critical step in the correct sorting
136 of periplasmic and surface exposed lipoproteins in phytopathogens (Konovalova et al., 2017; Narita
137 and Tokuda, 2017; Okuda and Tokuda, 2011; Wilson and Bernstein, 2016). A particular component
138 of this system is the LolB lipoprotein, which tethers the target lipoprotein to the outer lipid bilayer
139 (Hooda and Moraes, 2018; Wilson and Bernstein, 2016). A recent study showed that a *Xanthomonas*
140 *campestris* pv. *campestris* mutant strain unable to produce LolB has pleiotropic phenotypes, ranging
141 from decreased biofilm production to decreased virulence. LolB is required for the correct attachment
142 of lipoproteins to the outer layer, so these symptoms could be explained by the general defect in cell

143 membrane integrity and the perturbation of lipoprotein homeostasis (Liao et al., 2022). LolB is also
144 required for *in planta* survival in the xylem-invading phytopathogen *Ralstonia solanacearum* (Su et
145 al., 2021).

146 There are lipoproteins with unusual functions in phytopathogenic bacteria. For example, the two
147 surface-exposed small lipoproteins EcnA and EcnB (originally described as components of the
148 entericidin toxin-antitoxin system in an *Enterobacter* strain (Bishop et al., 1998; Schubiger et al.,
149 2015)) also modulate cell aggregation, biofilm formation, motility, outer membrane vesicle release,
150 and resistance to reactive oxygen species (ROS) in *Xanthomonas citri* subsp. *citri* and *Agrobacterium*
151 *tumefaciens*. (Granato et al., 2019; Knoke et al., 2020; Sidhu et al., 2008). Lipoproteins also
152 participate in host-pathogen interactions. For example, the lipoprotein VacJ is a component of the
153 lipid asymmetry maintenance machinery, and when absent in *X. citri* results in decreased biofilm
154 formation in glass tubes and on leaf surfaces, decreased swarming motility, reduced bacterial growth
155 in planta, and reduced virulence (Li and Yu, 2020). Another lipoprotein of *X. citri*, OmlA, is important
156 for multidrug resistance, and it may be implicated in protein-protein interactions and maintenance of
157 the outer membrane integrity (Fuangthong et al., 2008; Vanini et al., 2008). Lipoproteins are also
158 important for the production and export of exopolysaccharides in some bacteria, for example the
159 lipoprotein GumB is involved in the biosynthesis of xanthan gum by *X. campestris* (Jacobs et al.,
160 2012). In *Rhizobium leguminosarum*, the lipoprotein PssN is part of the exopolysaccharide
161 polymerization and export complex PssTNOP, which are needed for the successful invasion of its
162 host (Marczak et al., 2006; Wielbo et al., 2004). In a previous report by our laboratory we found that
163 in diderm bacteria a third of the expansin precursors have a lipoprotein signal peptide, principally in
164 the phytopathogenic members of Burkholderiales (excluding *Ralstonia* species) and
165 Xanthomonadales, in the soil-dwelling Myxococcales, and in the Gram-positive Frankiales order (de
166 Sandozequi et al., 2022). In addition, bacteria of other Phyla appear to associate expansins to their
167 cell envelopes in diverse manners: in Firmicutes, expansin precursors contain a putative LPXTG-like

168 motif; in the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) superphylum, expansins contain a C-
169 terminal domain recognized by the recently discovered Type 9 secretion system (T9SS) that attaches
170 A-LPS to the protein (Veith et al., 2017); and Cryptosporangiales expansins contain a transmembrane
171 domain (de Sandozequi et al., 2022). The biological function and mechanism of action of bacterial
172 expansins has yet to be determined, but it is likely associated to plant cell wall degradation and
173 virulence. Several reports of expansin mutants of plant growth-promoting and phytopathogenic
174 bacteria evidence a role for expansins during colonization and infection of the plant host (Georgelis
175 et al., 2011; Junior et al., 2015; Narváez-Barragán et al., 2020; Pastor et al., 2015; Tancos et al., 2018).
176 Although in monoderms, most lipoproteins attach to the outer layer of the cytoplasmic membrane and
177 are covered by the peptidoglycan layer ([Figure 1](#)), some lipoproteins could be big enough or have
178 specialized domains that allow them to be exposed to the surface (Fischetti, 2019; Nguyen et al.,
179 2020; Saleh et al., 2013). Nevertheless, lipoproteins in monoderm bacteria are needed for general
180 housekeeping and cell wall homeostasis, therefore important for the survival of bacteria in any
181 environment (Desvaux et al., 2018; Nguyen et al., 2020).

182 **Perspectives**

183 Surface-associated proteins participate in plant colonization and infection by bacteria, and some are
184 important in phytopathogenic bacterial fitness. For example, gene RSc2007, encoding an unknown
185 lipoprotein (UniProt Id: Q8XXV6), appears to be a xylem sap fitness factor in *R. solanacearum*
186 (Georgoulis et al., 2021). Importantly, lipoproteins could be the target of potential biocontrol targets,
187 such as the coumarins that inhibit the expression of the outer membrane lipoprotein OmlA in *X.*
188 *campestris* (Fuangthong et al., 2008). Yet, a large number of predicted lipoproteins and cellwall-
189 anchored proteins with unknown functions still needs characterization bringing about new
190 possibilities for identifying bacteria-plant interactors, some of which may be key players for
191 virulence. Finally, we must leverage artificial intelligence (AI) and machine learning (ML) tools (such
192 as SignalP v6.0 (Teufel et al., 2022), CW-PRED (Fimereli et al., 2012) and AlphaFold (Jumper et al.,

193 2021) for faster and more precise identification of surface-associated proteins in newly sequenced
194 genomes of emerging and established phytopathogens as agriculture crises requires diverse strategies
195 to combat the challenges at hand.

196

197 **Declaration of Competing Interest**

198 We declare no conflict of interest.

199

200 **Data sharing statement**

201 Data sharing is not applicable to this article as no new data were created or analyzed in this study.

202

203 **Acknowledgements**

204 We are grateful for financial support from DGAPA (PAPIIT)-UNAM grant # IN203622 to C.M.-A;
205 and to CONACYT for a scholarship to A. de S. We are also grateful with Dr Christopher D. Wood,
206 for critical reading of our work.

207

208 **References**

- 209 Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S.,
210 et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks.
211 *Nature Biotechnology*, 37, 420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
- 212 Alvarez-Martinez, C.E., Sgro, G.G., Araujo, G.G., Paiva, M.R.N., Matsuyama, B.Y., Guzzo, C.R.,
213 et al. (2021) Secrete or perish: The role of secretion systems in *Xanthomonas* biology.

214 *Computational and Structural Biotechnology Journal*, 19, 279–302.
215 <https://doi.org/10.1016/j.csbj.2020.12.020>.

216 Asmar, A.T. & Collet, J.F. (2018) Lpp, the Braun lipoprotein, turns 50—major achievements and
217 remaining issues. *FEMS Microbiology Letters*, 365, 1–8.
218 <https://doi.org/10.1093/femsle/fny199>.

219 Asmar, A.T., Ferreira, J.L., Cohen, E.J., Cho, S.H., Beeby, M., Hughes, K.T., et al. (2017)
220 Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS*
221 *Biology*, 15, 1–16. <https://doi.org/10.1371/journal.pbio.2004303>.

222 Babu, M.M., Priya, M.L., Selvan, A.T., Madera, M., Gough, J., Aravind, L., et al. (2006) A
223 database of bacterial lipoproteins (DOLOP) with functional assignments to predicted
224 lipoproteins. *Journal of Bacteriology*, 188, 2761–2773. [https://doi.org/10.1128/JB.188.8.2761-](https://doi.org/10.1128/JB.188.8.2761-2773.2006)
225 [2773.2006](https://doi.org/10.1128/JB.188.8.2761-2773.2006).

226 Babu, M.M. & Sankaran, K. (2002) DOLOP - Database of bacterial lipoproteins. *Bioinformatics*,
227 18, 641–643. <https://doi.org/10.1093/bioinformatics/18.4.641>.

228 Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021)
229 UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49, D480–
230 D489. <https://doi.org/10.1093/nar/gkaa1100>.

231 Bishop, R.E., Leskiw, B.K., Hodges, R.S., Kay, C.M. & Weiner, J.H. (1998) The entericidin locus
232 of *Escherichia coli* and its implications for programmed bacterial cell death. *Journal of*
233 *Molecular Biology*, 280, 583–596. <https://doi.org/10.1006/jmbi.1998.1894>.

234 Bogino, P.C., Oliva, M. de las M., Sorroche, F.G. & Giordano, W. (2013) The role of bacterial
235 biofilms and surface components in plant-bacterial associations. *International Journal of*
236 *Molecular Sciences*, 14, 15838–15859. <https://doi.org/10.3390/ijms140815838>.

237 Braun, V. & Rehn, K. (1969) Chemical Characterization, Spatial Distribution and Function of a
238 Lipoprotein (Murein-Lipoprotein) of the E. coli Cell Wall: The Specific Effect of Trypsin on
239 the Membrane Structure. *European Journal of Biochemistry*, 10, 426–438.
240 <https://doi.org/10.1111/j.1432-1033.1969.tb00707.x>.

241 Budzik, J.M., Marraffini, L.A. & Schneewind, O. (2007) Assembly of pili on the surface of *Bacillus*
242 *cereus* vegetative cells. *Molecular Microbiology*, 66, 495–510. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2958.2007.05939.x)
243 [2958.2007.05939.x](https://doi.org/10.1111/j.1365-2958.2007.05939.x).

244 Cascioferro, S., Totsika, M. & Schillaci, D. (2014) Sortase A: An ideal target for anti-virulence
245 drug development. *Microbial Pathogenesis*, 77, 105–112.
246 <https://doi.org/10.1016/j.micpath.2014.10.007>.

247 Chagnot, C., Zorgani, M.A., Astruc, T. & Desvaux, M. (2013) Proteinaceous determinants of
248 surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein
249 secretion perspective. *Frontiers in Microbiology*, 4, 1–26.
250 <https://doi.org/10.3389/fmicb.2013.00303>.

251 Chalupowicz, L., Barash, I., Reuven, M., Dror, O., Sharabani, G., Gartemann, K.H., et al. (2017)
252 Differential contribution of *Clavibacter michiganensis* ssp. *michiganensis* virulence factors to
253 systemic and local infection in tomato. *Molecular Plant Pathology*, 18, 336–346.
254 <https://doi.org/10.1111/mpp.12400>.

255 Choi, D.S., Yamada, H., Mizuno, T. & Mizushima, S. (1986) Trimeric structure and localization of
256 the major lipoprotein in the cell surface of *Escherichia coli*. *Journal of Biological Chemistry*,
257 261, 8953–8957. [https://doi.org/10.1016/s0021-9258\(19\)84474-5](https://doi.org/10.1016/s0021-9258(19)84474-5).

258 Cole, G.B., Bateman, T.J. & Moraes, T.F. (2021) The surface lipoproteins of gram-negative
259 bacteria: Protectors and foragers in harsh environments. *Journal of Biological Chemistry*, 296,
260 100147. <https://doi.org/10.1074/jbc.REV120.008745>.

261 Comfort, D. & Clubb, R.T. (2004) A Comparative Genome Analysis Identifies Distinct Sorting
262 Pathways in Gram-Positive Bacteria. *Infection and Immunity*, 72, 2710–2722.
263 <https://doi.org/10.1128/IAI.72.5.2710-2722.2004>.

264 Cossart, P. & Jonquieres, R. (2000) Sortase, a universal target for therapeutic agents against Gram-
265 positive bacteria? *Proceedings of the National Academy of Sciences of the United States of*
266 *America*, 97, 5013–5015. <https://doi.org/10.1073/pnas.97.10.5013>.

267 Cowles, C.E., Li, Y., Semmelhack, M.F., Cristea, I.M. & Silhavy, T.J. (2011) The free and bound
268 forms of Lpp occupy distinct subcellular locations in Escherichia coli. *Molecular*
269 *Microbiology*, 79, 1168–1181. <https://doi.org/10.1111/j.1365-2958.2011.07539.x>.

270 Dai, X., Böker, A. & Glebe, U. (2019) Broadening the scope of sortagging. *RSC Advances*, 9,
271 4700–4721. <https://doi.org/10.1039/c8ra06705h>.

272 Desvaux, M., Candela, T. & Serror, P. (2018) Surfaceome and proteosurfaceome in parietal
273 monoderm bacteria: Focus on protein cell-surface display. *Frontiers in Microbiology*, 9, 1–28.
274 <https://doi.org/10.3389/fmicb.2018.00100>.

275 Desvaux, M., Dumas, E., Chafsey, I. & Hébraud, M. (2006) Protein cell surface display in Gram-
276 positive bacteria: From single protein to macromolecular protein structure. *FEMS*
277 *Microbiology Letters*, 256, 1–15. <https://doi.org/10.1111/j.1574-6968.2006.00122.x>.

278 Diego, I. De, Ksiazek, M., Mizgalska, D., Koneru, L., Golik, P., Szmigielski, B., et al. (2016) The
279 outer-membrane export signal of Porphyromonas gingivalis type IX secretion system (T9SS)
280 is a conserved C-terminal β -sandwich domain. *Scientific Reports*, 6, 1–17.
281 <https://doi.org/10.1038/srep23123>.

282 Dramsi, S., Magnet, S., Davison, S. & Arthur, M. (2008) Covalent attachment of proteins to
283 peptidoglycan. *FEMS Microbiology Reviews*, 32, 307–320. [https://doi.org/10.1111/j.1574-](https://doi.org/10.1111/j.1574-6976.2008.00102.x)
284 [6976.2008.00102.x](https://doi.org/10.1111/j.1574-6976.2008.00102.x).

285 Dubuisson, J.F., Vianney, A., Hugouvieux-Cotte-Pattat, N. & Lazzaroni, J.C. (2005) Tol-Pal
286 proteins are critical cell envelope components of *Erwinia chrysanthemi* affecting cell
287 morphology and virulence. *Microbiology*, 151, 3337–3347.
288 <https://doi.org/10.1099/mic.0.28237-0>.

289 Egan, S.A., Kurian, D., Ward, P.N., Hunt, L. & Leigh, J.A. (2010) Identification of sortase A (SrtA)
290 substrates in *Streptococcus uberis*: Evidence for an additional hexapeptide (LPXXXD) sorting
291 motif. *Journal of Proteome Research*, 9, 1088–1095. <https://doi.org/10.1021/pr901025w>.

292 Fimereli, D.K., Tsirigos, K.D., Litou, Z.I., Liakopoulos, T.D., Bagos, P.G. & Hamodrakas, S.J.
293 (2012) Artificial Intelligence: Theories and Applications. CW-PRED: A HMM-Based Method
294 for the Classification of Cell Wall-Anchored Proteins of Gram-Positive Bacteria. *Hellenic*
295 *Conference on Artificial Intelligence*, 7297, 285–290.

296 Fischetti, V.A. (2019) Surface Proteins on Gram-Positive Bacteria. *Gram-Positive Pathogens*, 7,
297 19–31. <https://doi.org/10.1128/9781683670131.ch2>.

298 Flemming, H.C. & Wingender, J. (2010) The biofilm matrix. *Nature Reviews Microbiology*, 8, 623–
299 633. <https://doi.org/10.1038/nrmicro2415>.

300 Fuangthong, M., Sallabhan, R., Atichartpongkul, S., Rangkadilok, N., Sriprang, R., Satayavivad, J.,
301 et al. (2008) The omlA gene is involved in multidrug resistance and its expression is inhibited
302 by coumarins in *Xanthomonas campestris* pv. *phaseoli*. *Archives of Microbiology*, 189, 211–
303 218. <https://doi.org/10.1007/s00203-007-0310-1>.

304 Gartemann, K.H., Kirchner, O., Engemann, J., Gräfen, I., Eichenlaub, R. & Burger, A. (2003)
305 *Clavibacter michiganensis* subsp. *michiganensis*: First steps in the understanding of virulence
306 of a Gram-positive phytopathogenic bacterium. *Journal of Biotechnology*, 106, 179–191.
307 <https://doi.org/10.1016/j.jbiotec.2003.07.011>.

308 Georgelis, N., Tabuchi, A., Nikolaidis, N. & Cosgrove, D.J. (2011) Structure-function analysis of
309 the bacterial expansin EXLX1. *Journal of Biological Chemistry*, 286, 16814–16823.
310 <https://doi.org/10.1074/jbc.M111.225037>.

311 Georgoulis, S.J., Shalvarjian, K.E., Helmann, T.C., Hamilton, C.D., Carlson, H.K., Deutschbauer,
312 A.M., et al. (2021) Genome-Wide Identification of Tomato Xylem Sap Fitness Factors for
313 Three Plant-Pathogenic *Ralstonia* Species Gilbert, J.A. (Ed.). *mSystems*, 6.
314 <https://doi.org/10.1128/mSystems.01229-21>.

315 Granato, L.M., Picchi, S.C., Oliveira Andrade, M. De, Martins, P.M.M., Takita, M.A., Machado,
316 M.A., et al. (2019) The ECNA antitoxin is important not only for human pathogens: Evidence
317 of Its Role in the Plant Pathogen *Xanthomonas citri* subsp. *Citri*. *Journal of Bacteriology*, 201.
318 <https://doi.org/10.1128/JB.00796-18>.

319 Green, E.R. & Mecsas, J. (2016) Bacterial secretion systems: An overview. *Virulence Mechanisms*
320 *of Bacterial Pathogens*, 4, 213–239. <https://doi.org/10.1128/9781555819286.ch8>.

321 Hellman, J., Roberts, J.D., Tehan, M.M., Allaire, J.E. & Shaw Warren, H. (2002) Bacterial
322 peptidoglycan-associated lipoprotein is released into the bloodstream in Gram-negative sepsis
323 and causes inflammation and death in mice. *Journal of Biological Chemistry*, 277, 14274–
324 14280. <https://doi.org/10.1074/jbc.M109696200>.

325 Hendrickx, A.P.A., Budzik, J.M., Oh, S.Y. & Schneewind, O. (2011) Architects at the bacterial
326 surface-sortases and the assembly of pili with isopeptide bonds. *Nature Reviews Microbiology*,
327 9, 166–176. <https://doi.org/10.1038/nrmicro2520>.

328 Hendrickx, A.P.A., Willems, R.J.L., Bonten, M.J.M. & Schaik, W. van (2009) LPxTG surface
329 proteins of enterococci. *Trends in Microbiology*, 17, 423–430.
330 <https://doi.org/10.1016/j.tim.2009.06.004>.

331 Hooda, Y. & Moraes, T.F. (2018) Translocation of lipoproteins to the surface of gram negative
332 bacteria. *Current Opinion in Structural Biology*, 51, 73–79.
333 <https://doi.org/10.1016/j.sbi.2018.03.006>.

334 Huynh, M.S., Hooda, Y., Li, Y.R., Jagielnicki, M., Chieh-, C., Lai, L., et al. (2022) Reconstitution
335 of surface lipoprotein translocation through the Slam translocon. 1–24.

336 Jacobs, M., Salinas, S.R., Bianco, M.I. & Ielpi, L. (2012) Expression, purification and
337 crystallization of the outer membrane lipoprotein GumB from *Xanthomonas campestris*. *Acta*
338 *Crystallographica Section F: Structural Biology and Crystallization Communications*, 68,
339 1255–1258. <https://doi.org/10.1107/S1744309112036597>.

340 Jin, S., Joe, A., Lynett, J., Hani, E.K., Sherman, P. & Chan, V.L. (2001) JlpA, a novel surface-
341 exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial
342 cells. *Molecular Microbiology*, 39, 1225–1236. [https://doi.org/10.1046/j.1365-](https://doi.org/10.1046/j.1365-2958.2001.02294.x)
343 [2958.2001.02294.x](https://doi.org/10.1046/j.1365-2958.2001.02294.x).

344 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021) Highly
345 accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.
346 <https://doi.org/10.1038/s41586-021-03819-2>.

347 Junior, A.T., Dolce, L.G., Oliveira Neto, M. de & Polikarpov, I. (2015) *Xanthomonas campestris*
348 expansin-like X domain is a structurally disordered beta-sheet macromolecule capable of
349 synergistically enhancing enzymatic efficiency of cellulose hydrolysis. *Biotechnology Letters*,
350 37, 2419–2426. <https://doi.org/10.1007/s10529-015-1927-9>.

351 Knoke, L.R., Abad Herrera, S., Götz, K., Justesen, B.H., Günther Pomorski, T., Fritz, C., et al.
352 (2020) Agrobacterium tumefaciens Small Lipoprotein Atu8019 Is Involved in Selective Outer
353 Membrane Vesicle (OMV) Docking to Bacterial Cells. *Frontiers in Microbiology*, 11, 1–20.
354 <https://doi.org/10.3389/fmicb.2020.01228>.

355 Konovalova, A., Kahne, D.E. & Silhavy, T.J. (2017) Outer Membrane Biogenesis. *Annual Review*
356 *of Microbiology*, 71, 539–556. <https://doi.org/10.1146/annurev-micro-090816-093754>.

357 Kovacs-Simon, A., Titball, R.W. & Michell, S.L. (2011) Lipoproteins of bacterial pathogens.
358 *Infection and Immunity*, 79, 548–561. <https://doi.org/10.1128/IAI.00682-10>.

359 Kudryavtsev, K. V., Fedotcheva, T.A. & Shimanovsky, N.L. (2021) Inhibitors of Sortases of Gram-
360 Positive Bacteria and their Role in the Treatment of Infectious Diseases (Review).
361 *Pharmaceutical Chemistry Journal*, 55, 751–756. [https://doi.org/10.1007/s11094-021-02488-](https://doi.org/10.1007/s11094-021-02488-9)
362 9.

363 Lazzaroni, J. -C & Portalier, R. (1992) The excC gene of Escherichia coli K-12 required for cell
364 envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL). *Molecular*
365 *Microbiology*, 6, 735–742. <https://doi.org/10.1111/j.1365-2958.1992.tb01523.x>.

366 Li, X.Z., Nikaido, H. & Poole, K. (1995) Role of MexA-MexB-OprM in antibiotic efflux in
367 Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 39, 1948–1953.
368 <https://doi.org/10.1128/AAC.39.9.1948>.

369 Li, Y. & Yu, Q. (2020) Role of the lipoprotein VacJ/MlaA in the physiology and virulence of the
370 plant pathogen Xanthomonas citri subsp. citri. *Physiological and Molecular Plant Pathology*,
371 112, 1–6. <https://doi.org/10.1016/j.pmpp.2020.101546>.

372 Liao, C.T., Li, C.E., Chang, H.C., Hsu, C.H., Chiang, Y.C. & Hsiao, Y.M. (2022) The lolB gene in
373 *Xanthomonas campestris* pv. *campestris* is required for bacterial attachment, stress tolerance,
374 and virulence. *BMC Microbiology*, 22, 1–13. <https://doi.org/10.1186/s12866-021-02416-7>.

375 Lorenz, C., Dougherty, T.J. & Lory, S. (2019) Correct sorting of lipoproteins into the inner and
376 outer membranes of *Pseudomonas aeruginosa* by the *Escherichia coli* lolCde transport system.
377 *mBio*, 10, 1–13. <https://doi.org/10.1128/MBIO.00194-19>.

378 Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012) Top 10
379 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13, 614–
380 629. <https://doi.org/10.1111/j.1364-3703.2012.00804.x>.

381 Marczak, M., Mazur, A., Król, J.E., Gruszecki, W.I. & Skorupska, A. (2006) Lipoprotein PssN of
382 *Rhizobium leguminosarum* bv. *trifolii*: Subcellular localization and possible involvement in
383 exopolysaccharide export. *Journal of Bacteriology*, 188, 6943–6952.
384 <https://doi.org/10.1128/JB.00651-06>.

385 Narita, S. Ichiro & Tokuda, H. (2017) Bacterial lipoproteins; biogenesis, sorting and quality control.
386 *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1862, 1414–1423.
387 <https://doi.org/10.1016/j.bbalip.2016.11.009>.

388 Narváez-Barragán, D.A., Tovar-Herrera, O.E., Torres, M., Rodríguez, M., Humphris, S., Toth, I.K.,
389 et al. (2020) Expansin-like Ex11 from *Pectobacterium* is a virulence factor required for host
390 infection, and induces a defence plant response involving ROS, and jasmonate, ethylene and
391 salicylic acid signalling pathways in *Arabidopsis thaliana*. *Scientific Reports*, 10, 1–14.
392 <https://doi.org/10.1038/s41598-020-64529-9>.

393 Navarre, W.W. & Schneewind, O. (1994) Proteolytic cleavage and cell wall anchoring at the
394 LPXTG motif of surface proteins in Gram-positive bacteria. *Molecular Microbiology*, 14,
395 115–121. <https://doi.org/10.1111/j.1365-2958.1994.tb01271.x>.

396 Nguyen, M.T., Matsuo, M., Niemann, S., Herrmann, M. & Götz, F. (2020) Lipoproteins in Gram-
397 Positive Bacteria: Abundance, Function, Fitness. *Frontiers in Microbiology*, 11, 1–15.
398 <https://doi.org/10.3389/fmicb.2020.582582>.

399 Novick, R.P. (2000) Sortase: The surface protein anchoring transpeptidase and the LPXTG motif.
400 *Trends in Microbiology*, 8, 148–151. [https://doi.org/10.1016/S0966-842X\(00\)01741-8](https://doi.org/10.1016/S0966-842X(00)01741-8).

401 Okuda, S. & Tokuda, H. (2011) Lipoprotein sorting in bacteria. *Annual Review of Microbiology*, 65,
402 239–259. <https://doi.org/10.1146/annurev-micro-090110-102859>.

403 Ouyang, P., He, X., Yuan, Z.W., Yin, Z.Q., Fu, H., Lin, J., et al. (2018) Erianin against
404 staphylococcus aureus infection via inhibiting sortase A. *Toxins*, 10, 1–14.
405 <https://doi.org/10.3390/toxins10100385>.

406 Pastor, N., Dávila, S., Pérez-Rueda, E., Segovia, L. & Martínez-Anaya, C. (2015) Electrostatic
407 analysis of bacterial expansins. *Proteins: Structure, Function and Bioinformatics*, 83, 215–
408 223. <https://doi.org/10.1002/prot.24718>.

409 Paterson, G.K. & Mitchell, T.J. (2004) The biology of Gram-positive sortase enzymes. *Trends in*
410 *Microbiology*, 12, 89–95. <https://doi.org/10.1016/j.tim.2003.12.007>.

411 Paulsson, M., Kragh, K.N., Su, Y.C., Sandblad, L., Singh, B., Bjarnsholt, T., et al. (2021)
412 Peptidoglycan-Binding Anchor Is a Pseudomonas aeruginosa OmpA Family Lipoprotein With
413 Importance for Outer Membrane Vesicles, Biofilms, and the Periplasmic Shape. *Frontiers in*
414 *Microbiology*, 12, 1–13. <https://doi.org/10.3389/fmicb.2021.639582>.

415 Pfeilmeier, S., Caly, D.L. & Malone, J.G. (2016) Bacterial pathogenesis of plants: future challenges
416 from a microbial perspective: Challenges in Bacterial Molecular Plant Pathology. *Molecular*
417 *plant pathology*, 17, 1298–1313. <https://doi.org/10.1111/mpp.12427>.

418 Pishesha, N., Ingram, J.R. & Ploegh, H.L. (2018) Sortase A: A Model for Transpeptidation and Its
419 Biological Applications. *Annual Review of Cell and Developmental Biology*, 34, 163–188.
420 <https://doi.org/10.1146/annurev-cellbio-100617-062527>.

421 Rayes, J. El, Rodríguez-Alonso, R. & Collet, J.F. (2021) Lipoproteins in Gram-negative bacteria:
422 new insights into their biogenesis, subcellular targeting and functional roles. *Current Opinion*
423 *in Microbiology*, 61, 25–34. <https://doi.org/10.1016/j.mib.2021.02.003>.

424 Remans, K., Vercammen, K., Bodilis, J. & Cornelis, P. (2010) Genome-wide analysis and
425 literature-based survey of lipoproteins in *Pseudomonas aeruginosa*. *Microbiology*, 156, 2597–
426 2607. <https://doi.org/10.1099/mic.0.040659-0>.

427 Rodríguez-Ortega, M.J. (2018) The Surfaceome.

428 Saleh, M., Bartual, S.G., Abdullah, M.R., Jensch, I., Asmat, T.M., Petruschka, L., et al. (2013)
429 Molecular architecture of streptococcus pneumoniae surface thioredoxin-fold lipoproteins
430 crucial for extracellular oxidative stress resistance and maintenance of virulence. *EMBO*
431 *Molecular Medicine*, 5, 1852–1870. <https://doi.org/10.1002/emmm.201202435>.

432 Sandozequi, A. de, Salazar-Cortés, J.J., Tapia-Vázquez, I. & Martínez-Anaya, C. (2022) Prevalent
433 association with the bacterial cell envelope of prokaryotic expansins revealed by
434 bioinformatics analysis. *Protein Science*, 31, 1–14. <https://doi.org/10.1002/pro.4315>.

435 Schubiger, C.B., Orfe, L.H., Sudheesh, P.S., Cain, K.D., Shah, D.H. & Calla, D.R. (2015)
436 Entericidin is required for a probiotic treatment (*Enterobacter* sp. Strain C6-6) to protect trout
437 from cold-water disease challenge. *Applied and Environmental Microbiology*, 81, 658–665.
438 <https://doi.org/10.1128/AEM.02965-14>.

439 Si, L., Li, P., Liu, X. & Luo, L. (2016) Chinese herb medicine against Sortase A catalyzed
440 transformations, a key role in gram-positive bacterial infection progress. *Journal of Enzyme*

441 *Inhibition and Medicinal Chemistry*, 31, 184–196.
442 <https://doi.org/10.1080/14756366.2016.1178639>.

443 Sidhu, V.K., Vorhölter, F.J., Niehaus, K. & Watt, S.A. (2008) Analysis of outer membrane vesicle
444 associated proteins isolated from the plant pathogenic bacterium *Xanthomonas campestris* pv.
445 *campestris*. *BMC Microbiology*, 8, 1–16. <https://doi.org/10.1186/1471-2180-8-87>.

446 Silhavy, T.J., Kahne, D. & Walker, S. (2010) The Bacterial Cell Envelope. *Cold Spring Harbor*
447 *Perspectives in Biology*, 2, a000414–a000414. <https://doi.org/10.1101/cshperspect.a000414>.

448 Su, Y., Xu, Y., Liang, H., Yuan, G., Wu, X. & Zheng, D. (2021) Genome-Wide Identification of
449 *Ralstonia solanacearum* Genes Required for Survival in Tomato Plants. *mSystems*, 6.
450 <https://doi.org/10.1128/msystems.00838-21>.

451 Tancos, M.A., Lowe-Power, T.M., Peritore-Galve, F.C., Tran, T.M., Allen, C. & Smart, C.D.
452 (2018) Plant-like bacterial expansins play contrasting roles in two tomato vascular pathogens.
453 *Molecular Plant Pathology*, 19, 1210–1221. <https://doi.org/10.1111/mpp.12611>.

454 Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., et
455 al. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models.
456 *Nature Biotechnology*, 40, 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3>.

457 Vanini, M.M.T., Spisni, A., Sforça, M.L., Pertinhez, T.A. & Benedetti, C.E. (2008) The solution
458 structure of the outer membrane lipoprotein OmIA from *Xanthomonas axonopodis* pv. *citri*
459 reveals a protein fold implicated in protein-protein interaction. *Proteins: Structure, Function*
460 *and Genetics*, 71, 2051–2064. <https://doi.org/10.1002/prot.21886>.

461 Veith, P.D., Glew, M.D., Gorasia, D.G. & Reynolds, E.C. (2017) Type IX secretion: the generation
462 of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of

463 complex biopolymers. *Molecular Microbiology*, 106, 35–53.
 464 <https://doi.org/10.1111/mmi.13752>.

465 Viljoen, A., Foster, S.J., Fantner, G.E., Hobbs, J.K. & Dufrêne, Y.F. (2020) Scratching the surface:
 466 bacterial cell envelopes at the nanoscale. *mBio*, 11, 1–12. [https://doi.org/10.1128/mBio.03020-](https://doi.org/10.1128/mBio.03020-19)
 467 19.

468 Wielbo, J., Mazur, A., Król, J.E., Marczak, M. & Skorupska, A. (2004) Environmental modulation
 469 of the pssTNOP gene expression in *Rhizobium leguminosarum* bv. *trifolii*. *Canadian Journal*
 470 *of Microbiology*, 50, 201–211. <https://doi.org/10.1139/w04-004>.

471 Wilson, M.M. & Bernstein, H.D. (2016) Surface-Exposed Lipoproteins: An Emerging Secretion
 472 Phenomenon in Gram-Negative Bacteria. *Trends in Microbiology*, 24, 198–208.
 473 <https://doi.org/10.1016/j.tim.2015.11.006>.

474 **Tables**

475 **Table 1.** Number of predicted lipoprotein genes in important phytopathogens.

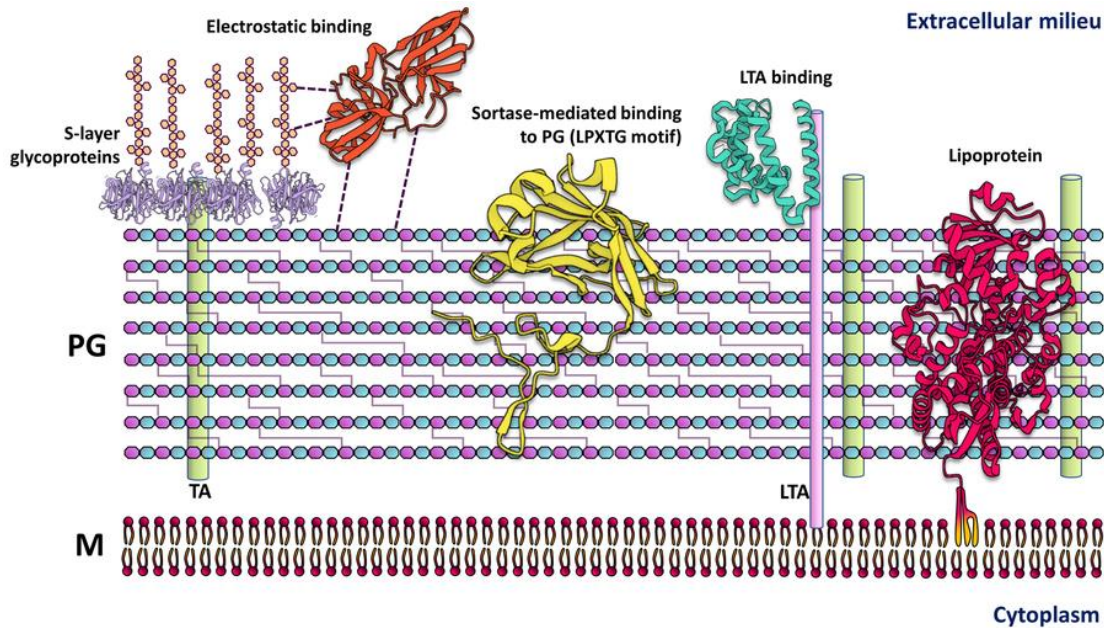
Genome	DOLOP database ^a	By using SignalP 6.0 ^b
<i>Agrobacterium fabrum</i> C58 (<i>Agrobacterium tumefaciens</i> strain C58)	29	65
<i>Burkholderia pseudomallei</i> K96243	93	209
<i>Pectobacterium atrosepticum</i> SCRI1043	112	142
<i>Pectobacterium brasiliense</i> BC1	N/P	125 ^c
<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i> PC1	N/P	159 ^c
<i>Pseudomonas aeruginosa</i> PAO1	113	204
<i>Pseudomonas putida</i> KT2440	74	166
<i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000	101	173

<i>Ralstonia solanacearum</i> GMI1000	47	191
<i>Streptomyces scabiei</i> (strain 87.22)	N/P	238
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> 306	92	171 ^c
<i>Xanthomonas albilineans</i> GPE PC73	N/P	131
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913	95	186
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	58	122
<i>Xylella fastidiosa</i> 9a5c	47	79

476 ^aDOLOP database (Babu et al., 2006). ^bLipoprotein genes predicted using SignalP 6.0 (Teufel et al.,
477 2022) in reference proteomes from UniProt release 2021_02; these include signal peptides for
478 Sec/SPII and Tat/SPII processing. ^cFor important phytopathogen without a reference proteome,
479 SignalP 6.0 online server (<https://services.healthtech.dtu.dk/service.php?SignalP-6.0>) was used to
480 predict the number of lipoproteins. N/P, species absent in the DOLOP
481 database.

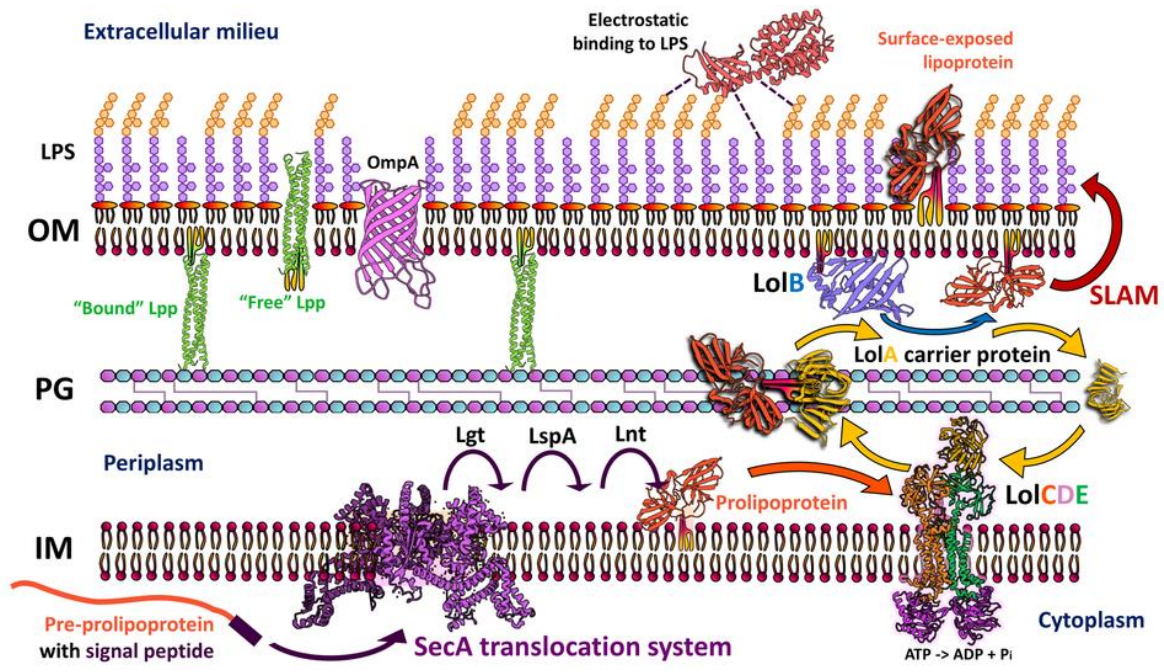
482

483 **Figures**



484

485 **Figure 1. Surface-exposed proteins in monoderm bacteria.** In monoderms, proteins attach to the
 486 surface through non-covalent direct binding or binding by specialized protein domains and
 487 electrostatic interactions to cell wall components such as the peptidoglycan layer and lipoteichoic acids
 488 (LTA), (Chagnot et al., 2013; Desvaux et al., 2006; Desvaux et al., 2018). Another mechanism
 489 involves covalent binding of proteins with an LPXTG motif in their C-terminal region by sortases.
 490 The LPXTG motif is cleaved between threonine (Thr) and glycine (Gly) residues, and sortases attach
 491 the target protein to the Lipid II molecule, which is the building block of the growing peptidoglycan
 492 layer, specifically at the Thr residue. (Dai et al., 2019; Desvaux et al., 2018; Paterson and Mitchell,
 493 2004). In the case of monoderm's lipoproteins depending on their size or other characteristics, they
 494 can span the PG layer to be surface exposed (this is the case of some adhesins and surface enzymes,
 495 for instance (Fischetti, 2019)). PG, peptidoglycan layer; M, membrane; LTA, lipoteichoic acids.



496

497 **Figure 2. Surface-exposed proteins in diderm bacteria.** In diderms, after translocation by the Sec
 498 or Tat systems, and post-translational modification by the lipidation machinery (Lgt, LspA and Lnt),
 499 the LolABCDE system correctly sorts and localizes lipoproteins. The LolCDE complex detaches the
 500 prolipoprotein from the inner membrane using ATP and transfers the LolA carrier
 501 protein. LolA then transports the lipoprotein across the periplasm and delivers it to LolB to be
 502 anchored to the inner side of the outer membrane. These lipoproteins can then be transposed to the
 503 extracellular side of the outer membrane by the SLAM or Bam complexes to become OM lipoproteins
 504 (Cole et al., 2021; Hooda and Moraes, 2018; Wilson and Bernstein, 2016). Other surface-exposed
 505 proteins can be embedded in the OM by transmembrane or pore-forming domains, such as those of
 506 the outer-membrane protein A (OmpA) family. Some proteins establish electrostatic interactions with
 507 lipopolysaccharide (LPS) chains in the outer membrane. Lpp is the most abundant protein in *E. coli*,
 508 and is found in two distinct forms: the “bound” form, which is responsible for maintaining the width
 509 of the periplasmic space; and the more abundant “free” form, with an unknown function yet
 510 discovered (Asmar et al., 2017; Asmar and Collet, 2018).

511

