



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

**Genómica estructural de *Bizonia argentinensis*:
búsqueda y caracterización de proteínas activas a bajas
temperaturas**

Tesis presentada para optar por el título de Doctor de la Universidad de Buenos Aires en
el área de Química Biológica

Lic. Leonardo Agustín Pellizza Pena

Director de Tesis: Dr. Martín Arán

Directora Asistente: Dra. Clara Smal

Consejero de Estudios: Dr. Julio Caramelo

Lugar de Trabajo: Fundación Instituto Leloir – IIBBA - CONICET

Buenos Aires, 2017

Fecha de defensa: 17/3/2017

Resumen

Genómica estructural de *Bizionia argentinensis*: Búsqueda y caracterización de proteínas activas a bajas temperaturas

La capacidad de los organismos psicrófilos para sobrevivir y proliferar a bajas temperaturas implica la superación de grandes obstáculos inherentes a los ambientes permanentemente fríos. Desafíos impuestos por las bajas temperaturas tales como la reducción de la actividad enzimática, alteración de la fluidez de la membrana, bajo transporte de nutrientes y productos de desecho, disminución de las tasas de transcripción, traducción y división celular, muestran que uno de los determinantes claves de la adaptación de la vida al frío yace indudablemente en la función de las proteínas que conducen el metabolismo y el ciclo celular de los organismos.

En este escenario, la bacteria marina *Bizionia argentinensis*, aislada en el Territorio Antártico Nacional, constituye un modelo interesante para el estudio de nuevas proteínas que posean actividad biológica en condiciones de bajas temperaturas. La caracterización bioquímica y estructural de estas biomoléculas es un punto de partida para la comprensión general de los mecanismos que permiten a las bacterias marinas adaptarse a estas condiciones de vida extremas.

En esta tesis se presenta un estudio de genómica estructural de proteínas de función desconocida de la bacteria *B. argentinensis*. En este contexto, se describe la caracterización bioquímica y estructural de dos blancos seleccionados: las proteínas BA42 y BA40. En particular, la estructura tridimensional de la proteína BA42 fue resuelta mediante RMN y Cristalografía de Rayos X. Estudios bioquímicos y biofísicos determinaron que BA42 es una fosfatasa la cual depende estrictamente de metales para su función y que la flexibilidad de la región comprendida por el sitio de unión a metales está asociada a la actividad de la proteína. Un análisis del contexto genómico de BA42 mostró que la proteína integra un operón constituido por algunas proteínas de función desconocida y otras asociadas a diversos roles celulares. Una de ellas, denominada BA40, pertenece a la superfamilia de factores de transcripción MerR. En este contexto, se comenzó el estudio de esta proteína con el fin de indagar sobre el rol biológico del operón y en particular de BA42. Los resultados confirmaron la actividad de unión al ADN de BA40 y se logró elucidar una secuencia consenso de ADN a la cual BA40 se une de forma específica. Así mismo, se observó que la región C-terminal de la proteína podría participar en la estabilización de la unión con el ADN, revelándose que la probable actividad transcripcional de BA40 podría encontrarse modulada mediante una interacción proteína-proteína, como ya ha sido descrito para proteínas con la misma arquitectura. Por último, el estudio *in vivo* de BA40 mostró que este factor de transcripción podría estar asociado a la respuesta a condiciones oligotróficas. De esta forma, los resultados y análisis que se detallan en esta tesis contribuyen al entendimiento de la relación entre la estructura y la función de proteínas de organismos psicotolerantes, aportando valiosa información para la discusión de los mecanismos moleculares y estrategias relevantes de adaptación a bajas temperaturas de la vida en los océanos.

Palabras claves: Genómica estructural, Bacteria antártica, *Bizionia argentinensis*, TPM domain, Actividad fosfatasa, MerR, Resonancia Magnética Nuclear.

Abstract

Structural genomics of *Bizionia argentinensis*: Search and characterization of proteins active at low temperatures

The ability of psychrophilic organisms to survive and proliferate at low temperatures implies the overcoming of major obstacles inherent to permanently cold environments. Challenges imposed by low temperatures such as reduced enzyme activity, decreased membrane fluidity, low transport of nutrients, decreased transcription rates, translation and cell division, show that one of the key determinants to cold adaptation lies undoubtedly in the function of the proteins that drive the metabolism and the cell cycle of organisms.

In this scenario, the marine bacterium *Bizionia argentinensis*, isolated in the National Antarctic Territory, constitutes an interesting model for the study of new proteins that possess biological activity in conditions of low temperature. The biochemical and structural characterization of these biomolecules is a starting point for a general understanding of the mechanisms that allow marine bacteria to adapt to these extreme life conditions.

This thesis presents a structural genomics study of proteins of unknown function from the bacterium *B. argentinensis*. Here, the biochemical and structural characterization of two selected candidates is described: the proteins BA42 and BA40. In particular, the three-dimensional structure of the BA42 protein was resolved by NMR and X-ray crystallography. Biochemical and biophysical studies determined that BA42 has metal-dependent phosphatase activity and that the flexibility of the region comprised by the metal binding site is associated with the protein activity. An analysis of the genomic context of BA42 showed that the gene integrates an operon that contains genes of unknown function and others associated with diverse cellular roles. One of them, called BA40, belongs to the superfamily MerR of transcription factors. In this sense, we started the study of BA40 in order to investigate the biological role of the operon and to better assess the function of BA42. The results were in line with the predicted DNA binding activity of BA40 and we found a consensus DNA sequence which binds specifically to BA40. It was also observed that the C-terminal region of BA40 could participate in the stabilization of the DNA-protein complex, revealing that the putative transcriptional activity of BA40 could be modulated by a protein-protein interaction, as it was already described for proteins bearing the same architecture. In summary, the results shown in this thesis contribute to the understanding of the relationship between structure and function of proteins from psychrotolerant organisms, providing valuable information for novel strategies and relevant molecular mechanisms associated to cold adaptation of bacteria to the life in the ocean.

Keywords: Structural genomics, Antarctic bacteria, *Bizionia argentinensis*, TPM domain, Phosphatase activity, MerR, Nuclear Magnetic Resonance

Agradecimientos

En primer lugar quiero agradecerles a mis dos directores: Clara y Martín. Gracias por todo lo que aprendí a su lado; por la dedicación, paciencia y libertad que me otorgaron, esta tesis ha sido posible en gran parte por ustedes.

También quiero acordarme de las personas que me formaron anteriormente y que son de gran importancia para mí. Gracias Ceci y Martín.

A todo el personal del Instituto Leloir: en particular biblioteca, administración, mantenimiento, a Diego (Droguería) quienes nos facilitan mucho el trabajo de cada día.

A la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) y al Concejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) por haberme otorgado el apoyo económico para la realización de esta tesis.

Al jurado, quién acepto evaluar esta tesis: los Doctores Patricio Craig, Rodolfo Rasia y Javier Santos. Muchas gracias.

A cada una de las personas que me ayudaron y tendieron una mano, a ellos mi profundo agradecimiento.

A mis amigos, por las risas, los consejos, la música, el fútbol. En especial a Andrés (Ruso), gracias por estar siempre.

A mi abuela Maruja, gracias por tanto cariño.

A mi hermana Gabi y mi cuñado Gastón, por acompañarme incondicionalmente en todo este camino.

A mi madre, casi todo se lo debo a ella.

A Romi y Fede por iluminarme la vida con tanto amor.

Abreviaturas

1D, 2D y 3D: Una, dos y tres dimensiones

Abs: Absorbancia

ATP: Adenosín trifosfato

BSA: Seroalbúmina bovina

CD: Dicroísmo circular

ChIP-seq: Inmunoprecipitación de Cromatina-secuenciación

CS: Desplazamiento químico

CSP: Perturbación del desplazamiento químico

dNTP: Desoxirribonucleósido trifosfato

DUF: Dominios de función desconocida

DTT: Ditioneitol

EDTA: Ácido etilendiamino tetra acético

EMSA: Ensayo de retardo en la movilidad electroforética

FAM: Amidita de fluoresceína

FPLC: Cromatografía líquida rápida de proteínas

HEPES: Ácido N-(2-hidroietil) piperazina- N'-2-etanosulfónico

HPLC: Cromatografía líquida de alta presión

HSQC: *Heteronuclear single quantum coherence spectroscopy*

HTH: Hélice vuelta hélice

IP: Co-inmunoprecipitación

IPTG: Isopropil-L-D-tiogalactósido

k_{cat}: Constante catalítica

K_D: Constante de disociación

K_M: Constante Michaelis-Menten

MD: Dinámica molecular

MALDI: Desorción/deionización laser asistida por matriz

NOE: Efecto Nuclear Overhauser

NTA: Ácido nitrilotriacético

ON: Toda la noche

ORF: Marcos abiertos de lectura

PBS: Buffer fosfato salino

PDB: Protein Data Bank

PFG: *Pulsed Field Gradient*

PNP: p-Nitrofenol

pNPP: p-Nitrofenil fosfato

pSer: Fosfoserina

RMN: Resonancia Magnética Nuclear

RMSD: Desviación de la media cuadrática

RMSF: Fluctuación cuadrática media

SDS: Duodecilsulfato de sodio

SDS-PAGE: Electroforesis en geles de poliacrilamida con SDS

SEC: Cromatografía de exclusión molecular

SELEX: *Systematic Evolution of Ligands by Exponential Enrichment*

SLS: Dispersión estática de luz

T₁: Tiempo de relajación longitudinal

T₂: Tiempo de relajación transversal

TBS: Tris buffer salino

Tris: Tris (hidroximetil) amino metano

V_{max}: Velocidad máxima

Wt: Salvaje

θ : Elipticidad

θ_{mrw} : Elipticidad molar residual

Índice

Resumen	I
Abstract	III
Abreviaturas	VII
Índice	XI
Introducción	1
Genómica estructural	1
El inicio y el presente de la genómica estructural	1
La genómica estructural y el desarrollo de las estrategias de alto rendimiento	2
La cobertura estructural del universo proteico	5
Los microorganismos psicrófilos en los océanos y sus adaptaciones	7
Bizionia argentinensis y la familia Flavobacteriaceae	10
El filo Bacteroidetes	10
El clado marino de la familia Flavobacteriaceae	11
<i>Bizionia argentinensis</i>	12
Capítulo I	15
Genómica estructural de <i>Bizionia argentinensis</i>	15
Selección bioinformática de los ORFs candidatos	19
Clonado, expresión y solubilidad de los ORFs candidatos	20
Evaluación de los candidatos por RMN (1D ¹ H y 2D HSQC ¹ H- ¹⁵ N)	23
Screening de cristalogénesis	27
Estructuras resueltas	29
Discusión	31
Capítulo II	35
Caracterización estructural y funcional de la proteína BA42	35
BA42 muestra una arquitectura $\alpha\beta\alpha$ de tres capas y muestra una conformación rígida en solución	39
BA42 muestra una característica estructural infrecuente: un par de “crossing loops”	41

La unión de los metales induce un cambio conformacional significativo en BA42	43
BA42 presenta una constante de disociación micromolar para la unión de Ca^{2+}	48
BA42 muestra una actividad hidrolasa de esteres monofosfóricos.....	50
La unión de Mg^{2+} a BA42 se asocia con un aumento en la actividad enzimática	58
BA42- Mg^{2+} presenta cambios conformacionales locales con respecto a BA42- Ca^{2+}	60
BA42- Mg^{2+} presenta un incremento de movimientos internos lentos en el “ <i>crossing loop</i> ” inferior y la región C-terminal con respecto a BA42- Ca^{2+}	63
BA42- Mg^{2+} presenta cambios en la estabilidad térmica con respecto a BA42- Ca^{2+} y BA42-Apo	64
Las mutantes de BA42 que son incapaces de unirse a Mg^{2+} revelan la localización del sitio activo	66
Estudio in vivo de BA42	68
BA42 integraría un operón.....	70
Discusión	71
Capítulo III	77
<i>Caracterización estructural y funcional de la proteína BA40</i>	77
BA40 muestra actividad de unión a ADN	84
BA40 es un monómero en solución	85
Análisis de BA40 y BA40- ΔH5 mediante RMN	89
Elucidación del sitio de unión específico y caracterización del tipo de regulación implicada ..	90
Evaluación de la secuencia consenso obtenida por SELEX.....	93
La región C-terminal de BA40 incrementa la afinidad por el ADN.....	95
Estudio in vivo de BA40.....	96
Discusión	98
Discusión general	103
Materiales y métodos	107
Clonado Molecular	109
Construcción de mutantes	110
Reacciones de PCR	110
Extracción de ARN y síntesis de cDNA.....	110
Screening de solubilidad	111
Expresión de proteínas recombinantes solubles y purificación.....	111

Ensayos de actividad fosfatasa.....	112
Análisis por Dicroísmo Circular (DC).....	113
Dispersión estática de luz (SLS)	113
Marcación selectiva con ^{15}N y ^{13}C	114
Experimentos de RMN	114
HSQC y perturbación de desplazamientos químicos	114
Cálculo del “Chemical Shift Index”	115
Espectros 1D ^1H	115
Relajación del ^{15}N y NOE ^1H ^{15}N	115
Mediciones de radio hidrodinámico	115
Cristalogénesis.....	117
Ensayos de retardo en la movilidad electroforética (EMSA).....	117
SELEX	118
Ensayos de anisotropía de fluorescencia	119
Dinámica Molecular	119
Co-inmunoprecipitación.....	120
Cultivos de <i>Bizonia argentinensis</i>	120
Cromatografía de exclusión molecular (SEC)	121
Experimentos de entrecruzamiento	121
Anexo	123
Bibliografía	127

Introducción

Genómica estructural

El inicio y el presente de la genómica estructural

La genómica estructural o proteómica estructural como también se conoce, es un campo de la biología de rápido desarrollo en la era post-genómica. El principal objetivo de esta área de investigación es tratar de determinar la estructura de cada proteína codificada por un genoma dado, con el fin de obtener información valiosa sobre el plegamiento y función de las proteínas.

Con el advenimiento de los avances en la tecnología de secuenciación a mediados de los años 90, se marcó el comienzo de los proyectos de gran escala genómica, obteniéndose desde entonces los genomas completos de muchos organismos, dando lugar en la última década, a un número enorme (y en continuo crecimiento) de secuencias depositadas en las diferentes bases de datos [1, 2]. Fue tan drástico este crecimiento que por ejemplo el número de secuencias de proteínas en la base de datos no redundante RefSeq, alcanzó casi los 17 millones a principios de 2012, mientras que en 2005, la misma contenía 2 millones de secuencias. Más allá de este avasallante crecimiento en el universo de la genómica, estos esfuerzos globales de secuenciación nos han proporcionado modelos genéticos de organismos en todos los reinos de la vida, y como tal, la infraestructura de secuencias necesaria para el análisis en profundidad de las proteínas. Una de las mayores revelaciones de esta revolución, fue el descubrimiento de un gran número de genes conservados y cuya función era esencialmente desconocida. Alrededor del 40% del total de las secuencias génicas se clasifican como proteínas “hipotéticas conservadas” (*“conserved hypothetical” proteins*) indicando con esta denominación que no hay más información disponible que la traducción de la secuencia nucleotídica.

De esta forma, la genómica estructural nació con la idea de que la función de las proteínas que estos genes codifican podría ser elucidada mediante el estudio de sus estructuras 3D. Es así que a principios del siglo XXI se establecieron varios proyectos de genómica estructural a gran escala, con el propósito de lograr una cobertura estructural integral del universo proteico, y con la ambiciosa apuesta de poder asignar la función de este gran número de proteínas [3-6].

El objetivo original de esta nueva área pretendía que la estructura tridimensional de la mayoría de las proteínas estuviera fácilmente disponible, ya sea directamente de la base de datos “*Protein Data Bank*” (PDB) o fuera obtenida a través de métodos de modelado computacional. Sin embargo, esto se volvió extremadamente difícil de lograr debido al continuo crecimiento del número de secuencias [1].

De todas formas, muchos proyectos de gran y pequeña escala se han llevado a cabo desde entonces [6-12]. Estos se han centrado en la creación y puesta a punto de estrategias de alto rendimiento para la determinación de estructuras de alta resolución, utilizando técnicas como la cristalografía de rayos X [13-15] y la espectroscopia de Resonancia Magnética Nuclear (RMN) [16-18], focalizando sus esfuerzos en la selección de proteínas “hipotéticas conservadas” y dominios de función desconocida (“*Domain of Unknown Function*” - DUF) [5, 19, 20]. En este sentido, los DUF se han convertido en los blancos naturales de los principales centros de genómica estructural, ya que tales dominios cubren una fracción significativa del universo inexplorado de proteínas [21]. Es importante considerar, que esta selección de los blancos se opone a los esfuerzos "clásicos" en los cuales el estudio de proteínas se centraba principalmente en sistemas bien caracterizados, dejando a muchas de las familias de proteínas fuera de su esfera de interés, incluyendo, por defecto, a todas las familias DUF.

Por lo tanto, diferentes ejemplos han demostrado la viabilidad de la obtención de información funcional a través de la estructura [12, 18, 22], siendo muchos los DUF estructuralmente resueltos y depositados en la base de datos PDB. Sin embargo, estos representan un gran recurso no explotado para comprender su función y su mecanismo, inferir el plegamiento de otros miembros de la familia, crear modelos comparativos detallados de otras proteínas similares [23, 24], o abordar aspectos evolutivos.

En conclusión, hoy en día esta nueva área de la biología sigue en continuo desarrollo, promoviendo la creación de nuevas metodologías, tecnologías, y conocimientos que hacen posible la determinación de estructuras difíciles y novedosas, lográndose una evolución significativa en la biología estructural.

La genómica estructural y el desarrollo de las estrategias de alto rendimiento

El panorama de la biología estructural ha cambiado drásticamente durante la última década generando avances tecnológicos que posibilitaron aumentar el número y la complejidad de los problemas biológicos específicos que podrían abordarse. Como se dijo anteriormente, la genómica estructural busca describir la estructura 3D de cada proteína codificada por un genoma dado. Este enfoque basado en una visión total del genoma ha generado como consecuencia de su desarrollo, métodos de alto rendimiento específicos para la determinación de las estructuras 3D en proyectos de cientos de candidatos [23, 25].

Una de las primeras áreas en desarrollarse fue la bioinformática. En ella, se han descrito

métodos computacionales para optimizar la selección de proteínas dianas a través de algoritmos basados en los grandes conjuntos de datos obtenidos por estos proyectos, siendo un ejemplo los programas generados para predecir el comportamiento de proteínas en ensayos de cristalización [26]. Así mismo, aunque la expresión recombinante de las proteínas sigue siendo en gran medida empírica, el uso de todos los datos bioinformáticos disponibles en la elección de las secuencias de partida puede contribuir significativamente al éxito global de un proyecto de genómica estructural [26].

Otra área de vital trascendencia como la biología molecular, ha desarrollado métodos de alto rendimiento de clonación, expresión y purificación de los marcos abiertos de lectura (ORFs) [27]. Estas importantes mejoras en la clonación y expresión de proteínas en sistemas heterólogos [28], se desarrollaron en paralelo con las nuevas estrategias de purificación por cromatografías de afinidad que permitieron un aumento significativo de la capacidad de obtener cantidades del orden de miligramos de proteína, hecho de suma importancia para la determinación de las estructuras [13]. Esto último se acompaña de nuevas estrategias para mejorar la solubilidad de las proteínas heterólogas. Algunas de ellas como el replegado de proteínas, proteínas de fusión, uso de variantes mutacionales, y la elección de proteínas ortólogas, forman parte de las estrategias de alto rendimiento para la expresión de proteínas.

En cuanto a las técnicas experimentales utilizadas para la determinación de las estructuras de alta resolución de proteínas, la cristalografía de rayos X y la espectroscopia de RMN son los dos métodos más frecuentemente utilizados, y en estos también se han logrado importantes desarrollos en los procedimientos automatizados para la determinación de las estructuras 3D [13, 18, 29]. En la cristalografía de rayos X, los principales avances se han generado al aumentar significativamente la resolución, velocidad y calidad de la recopilación de datos. Estos desarrollos incluyeron nuevos métodos de cristalogénesis, la miniaturización de los experimentos de cristalización (que no sólo reduce la demanda de la cantidad de proteína purificada requerida, sino que también permite la exploración extensiva del "espacio" químico para identificar las condiciones óptimas de cristalización), nuevas técnicas criogénicas para obtener mejoras en la calidad y estabilidad de la recolección de datos, y una mayor disponibilidad y accesibilidad a los sincrotrones [30]. En lo que respecta a la espectroscopia de RMN, se han logrado importantes progresos en las metodologías, incluyendo también mejoras en el hardware, técnicas para la preparación de muestras y métodos para la adquisición de datos con su posterior análisis para el cálculo de las estructuras [16]. Paralelamente, el continuo desarrollo de nuevos paquetes de

software, ha aumentado la velocidad, fiabilidad y calidad en la determinación de las estructuras.

Todas estas áreas se integran en la genómica estructural generándose estrategias de una gran eficiencia en la determinación de estructuras de alta resolución (Figura I.1).

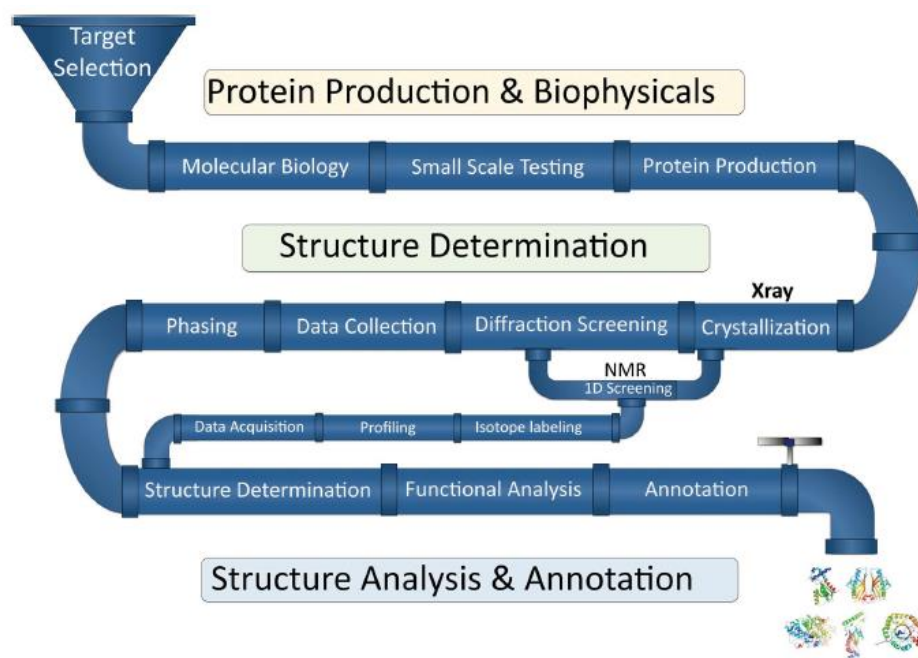


Figura I.1. Estrategia de alto rendimiento para la determinación de estructuras del consorcio Joint Center for Structural Genomics (JSCG). Los candidatos fueron priorizados para la determinación de su estructura por cristalografía de rayos X, mientras que la determinación de la estructura por RMN incluye principalmente proteínas que no cristalizan y tienen espectros aptos para su estudio estructural. Figura tomada de Elsiger et al, 2010 [14].

Sin embargo, a pesar de los múltiples avances técnicos conseguidos, la pérdida de proteínas diana a través de las diferentes etapas de la estrategia diseñada es un hecho que resulta casi inevitable. Por lo tanto, un reto importante de la genómica estructural es seleccionar las proteínas diana de forma realista, eficiente y a un bajo costo [31]. Al igual que con otros proyectos, es importante establecer los objetivos y las medidas cuantitativas de éxito. Se vuelve fundamental medir la importancia biológica y la dificultad de resolver las estructuras macromoleculares. Para esto, cada día las estrategias se basan más en servidores de predicción de diferentes parámetros para estimar la factibilidad de la resolución de las estructuras y, aunque cada nueva estructura experimental se suma a la base de datos, la mayoría de los biólogos estructurales están de acuerdo en que las estructuras novedosas son especialmente las que deben ser determinadas a

priori. En este sentido, la primera estructura de una proteína en una familia de función desconocida puede ser utilizada para entender la función y el mecanismo, inferir el plegamiento de otros miembros de la familia, crear modelos comparativos detallados de las proteínas más similares, o identificar las relaciones evolutivas.

La cobertura estructural del universo proteico

Con el fin de investigar el progreso hacia una cobertura completa del espacio estructural de secuencias proteicas, resulta útil analizar el número de estructuras resueltas en el contexto de las familias de proteínas, a modo de medir el aumento del número de familias que están estructuralmente caracterizadas, es decir, que tienen al menos un miembro con su estructura conocida.

En este sentido, la base de datos Pfam es una colección de dominios de proteínas y familias, y cada familia se encuentra representada por alineamientos de secuencias múltiples y modelos ocultos de Markov [32]. De esta forma, analizando los datos registrados en el año 2005, de las 7677 familias Pfam, el 36% contenía un miembro de estructura conocida [32]. Trascorridos solamente 7 años, el número de familias registradas pasó a 14831 familias y el 44% de estas tenía uno o más miembros caracterizados estructuralmente [21]. Finalmente el número de familias Pfam hoy en día es de 16306 [33]. Estos datos evidencian el continuo crecimiento de secuencias que se sucede año tras año, y aunque la cobertura estructural de las familias ha aumentado, dista de ser la que se pensaba. Más aun, de acuerdo con estos análisis, las estructuras 3D resueltas frecuentemente caen en familias que ya tienen un representante estructural (Figura I.2). Una razón que justifica este hecho es que las familias Pfam estructuralmente no caracterizadas son en promedio mucho más pequeñas en tamaño con respecto a las familias estructuralmente caracterizadas (Figura I.2.B). Otro dato interesante, es que estas familias pequeñas se encuentran enriquecidas en DUFs (36%) en comparación con las familias con representantes estructurales (10%) (Figura I.2.D). Al mismo tiempo, estas familias se ven enriquecidas en regiones *coiled-coil*, desordenadas y transmembrana, que a menudo hacen que la caracterización estructural de la proteína sea aún más complicada experimentalmente.

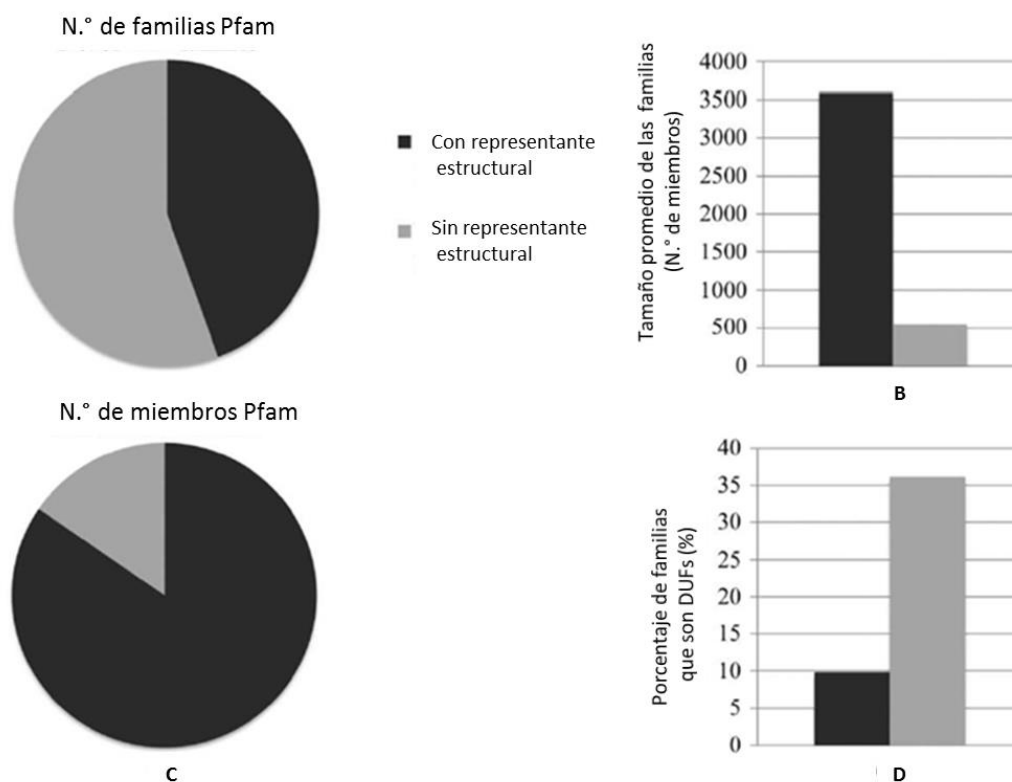


Figura 1.2. Evaluación de las Familias Pfam. Se señalan en negro y gris, las familias con y sin representantes estructurales, respectivamente. **(A)**. Proporción de familias Pfam con al menos un representante estructural. **(B)**. Tamaño medio de familias Pfam (número de integrantes) con y sin representantes estructurales. **(C)**. Proporción de miembros de Pfam que tienen estructura resuelta. **(D)** Porcentaje de familias Pfam DUFs con y sin un representante estructural. Figura tomada y adaptada de Mistry et al, 2013 [21].

Por lo tanto, estos análisis muestran la necesidad de seleccionar los candidatos específicamente tratando de direccionar los esfuerzos hacia los DUFs, a fin de evitar que los proyectos proporcionen cada vez más estructuras a las familias ya caracterizadas estructuralmente.

Los microorganismos psicrófilos en los océanos y sus adaptaciones

Aproximadamente el 80% de la biosfera de nuestro planeta se encuentra a temperaturas inferiores a los 5 °C. Esto incluye la mayor parte de los océanos de todo el mundo (los cuales cubren tres cuartas partes de la superficie terrestre), las regiones polares (la Antártida y las zonas dentro del círculo ártico), regiones elevadas como cadenas montañosas, entre otras [34]. Esto es un claro indicio de que la baja temperatura es la condición extrema más extendida, y como tal, los psicrófilos representan los extremófilos más abundantes, diversos, y ampliamente distribuidos en la Tierra [35]. La mayoría de los organismos psicrófilos se encuentran en los océanos y han sido recientemente el foco de exhaustivas revisiones [34].

En un sentido estricto, los organismos adaptados a los ambientes fríos son clasificados en dos clases, psicrófilos y psicrotolerantes, agrupándolos de esta forma según el rango de temperaturas de crecimiento. Es así que, los psicrófilos propiamente dichos, proliferan en un rango de temperaturas de 0 a 20 °C, con una temperatura óptima de 15 °C, mientras que los psicrotolerantes crecen óptimamente en un rango de 20 a 25 °C. Estos organismos adaptados a las bajas temperaturas no sólo sobreviven o soportan estas condiciones extremadamente inhóspitas, sino que están condicionados irreversiblemente a estos ambientes, ya que la mayoría de ellos no pueden crecer a las temperaturas óptimas de crecimiento de los organismos mesófilos. En estos ambientes, las temperaturas que tienden a 0 °C imponen severas restricciones fisicoquímicas a la función celular influyendo negativamente en la integridad celular, la viscosidad del agua, la permeabilidad de la membrana citoplasmática, las tasas de difusión de los solutos y la cinética enzimática, entre muchas otras [36]. Sin embargo, la presión selectiva más importante de las bajas temperaturas se ejerce sobre la cinética de las reacciones químicas, disminuyéndose también los movimientos moleculares asociados a la función de cada proteína. Un parámetro clave a la hora de analizar la actividad enzimática es la constante catalítica de reacción k_{cat} , la cual corresponde al número máximo de moléculas de sustrato convertidas en producto por sitio activo por unidad de tiempo. En este sentido, la constante catalítica de las reacciones tiene una dependencia exponencial con la temperatura según la siguiente ecuación (I.1):

$$k_{cat} = \kappa \frac{k_B T}{h} \exp(\Delta G^\# / RT) \quad (\text{Ecuación I.1})$$

dónde k_{cat} es la constante catalítica de reacción, κ es el coeficiente de transmisión (su valor depende de la viscosidad del solvente), k_B es la constante de Boltzmann, T es la temperatura, h es la constante de Planck, R es la constante universal de los gases ideales y ΔG^\ddagger es la energía libre de activación del estado de transición [37].

En este contexto, se ha visto que las enzimas de los organismos psicrófilos han evolucionado hacia estructuras más flexibles disminuyendo los valores de ΔG^\ddagger (Figura I.3), logrando un aumento en el valor de k_{cat} a temperaturas bajas [38]. Otro efecto adicional adjudicado a la mayor flexibilidad de estas enzimas es la observación de una unión más laxa del sustrato a los sitios activos. Esto se ve reflejado en el alto valor de la constante de Michaelis-Menten (K_M) que presentan muchas enzimas psicrófilas [38, 39].

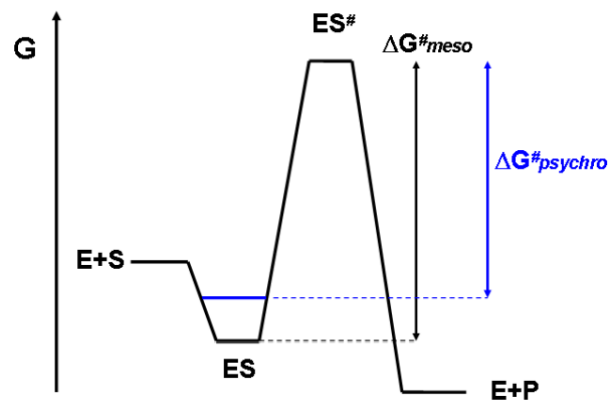


Figura I.3. Optimización de la actividad a través de la disminución de la afinidad al sustrato en las enzimas psicrófilas. Se muestra el perfil de reacción siguiendo los cambios de energía libre de Gibbs bajo una condición de sustrato saturante. La unión de sustrato débil (en azul) disminuye la energía libre de activación (ΔG^\ddagger psicrófilos), aumentando la velocidad de reacción. Figura tomada de Struvay y Fellers, 2012 [39].

De esta manera, la hipótesis comúnmente aceptada para esta adaptación al frío radica en la relación actividad-estabilidad-flexibilidad. Esta mayor flexibilidad observada en las enzimas de psicrófilos, se exhibe en algunos casos en la totalidad de la proteína. Sin embargo, muchos estudios han mostrado que tiende a estar restringida a regiones particulares de la proteína, especialmente al sitio activo de estas y su entorno [40]. Más aun, también se ha visto que regiones proteicas lejos de los sitios catalíticos pueden ser aún más rígidas que las de sus contrapartes mesófilas [41]. En este sentido, también se ha observado que las estructuras cristalográficas de proteínas psicrófilas muestran que éstas no tienen conformaciones inusuales, sino que comparten

una alta similitud con sus homólogos mesófilos y termófilos.

La mayoría de las enzimas psicrófilas muestran un mismo comportamiento: una inactivación por el aumento de temperatura, la cual es independiente de la estabilidad total de la proteína, inhibiendo su actividad antes de que la proteína llegue a perder su estructura. Contrariamente, las enzimas de organismos mesófilos se inactivan cuando estas comienzan a perder su conformación 3D. Esto muestra que el sitio activo en enzimas psicrófilas es mucho más susceptible al aumento de temperatura que la estructura proteica en su totalidad (Figura I.4).

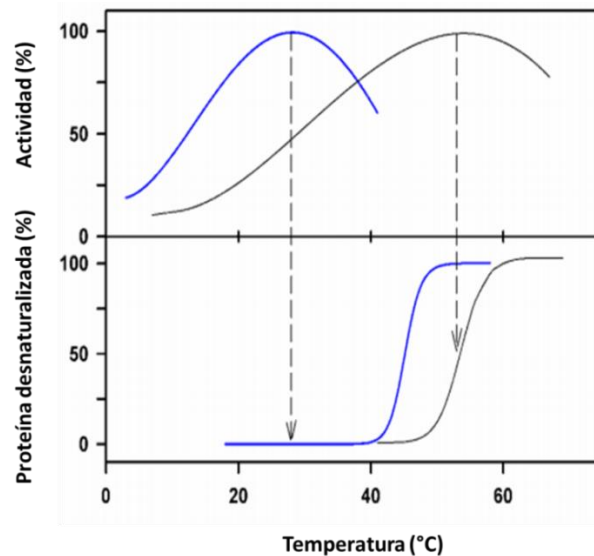


Figura I.4. Inactivación y desnaturalización de enzimas psicrófilas y mesófilas. La actividad de las enzimas psicrófilas (panel superior, línea azul) es inhibida por la temperatura antes de la desnaturalización total de la proteína (panel inferior, línea azul) mostrando la alta labilidad de los sitio activos de las enzimas psicrófilas. Por el contrario, la inactivación de las enzimas mesófilas (curvas negras) se corresponde estrechamente con la pérdida de la conformación de la proteína. Figura adaptada de Struvay y Fellers, 2012 [39].

Dentro de las formas para aumentar la flexibilidad se han observado diferentes modificaciones estructurales que conducen a la atenuación en la fuerza y el número de diferentes factores estabilizadores, logrando una modificación de los valores entálpicos y entrópicos. Algunos de ellos son la reducción de puentes salinos, enlaces de hidrógeno e interacciones hidrofóbicas; disminución de las interacciones entre diferentes subunidades de proteína; aumento de la Interacción con el solvente; reducción de la hidrofobicidad del núcleo proteico; una mayor accesibilidad al sitio activo; un aumento de la exposición de los residuos apolares al solvente; disminución de la unión de cofactores; un menor contenido de prolina y arginina entre otros [35].

***Bizionia argentinensis* y la familia Flavobacteriaceae**

El filo Bacteroidetes

Los miembros del filo bacteriano Bacteroidetes (conocido también como “Cytophaga-Flavobacterium-Bacteroides” o “CFB”) han podido colonizar gran diversidad de nichos ecológicos, desde ambientes acuáticos y terrestres hasta las microfloras intestinales. Particularmente, este grupo de bacterias heterotróficas es uno de los más abundantes en los ecosistemas marinos, siendo sólo superado por el filo Proteobacteria [42-44]. Este filo compuesto por bacterias gram-negativas, en mayor parte aerobias, se presentan en aguas marinas abiertas, costas, sedimentos marinos, regiones polares, etc [45].

En base al porcentaje de identidad de secuencia del ARN ribosomal 16S, este filo se puede dividir en cuatro clases: Bacteroidia, Sphingobacteria, Cytophagia y Flavobacteria. Dentro de esta última clase se encuentra la familia Flavobacteriaceae, la cual contiene más de 90 géneros, representando la rama principal de filo Bacteroidetes (Figura I.5).

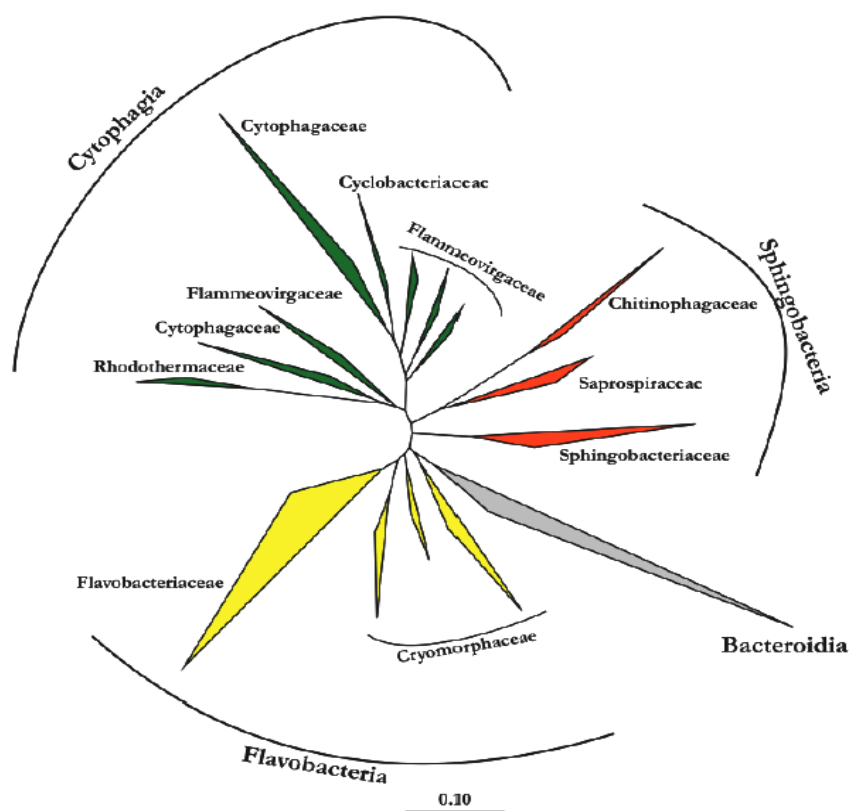


Figura I.5. Descripción del árbol filogenético del filo Bacteroidetes basado en la identidad de secuencia del rRNA 16S. El árbol incluye clases y algunas familias. Este fue adaptado de "All-Species Living Tree" (YARZA et al., 2008) [46]. La figura fue tomada de la Tesis Doctoral de Christin M. Bennke.

El clado marino de la familia Flavobacteriaceae

La mayor parte de las especies de la familia Flavobacteriaceae (algunas veces nombradas de manera informal como “flavobacterias”) forman un clado marino. En la última década, éste ha experimentado una rápida expansión, incorporándosele un gran número de nuevas especies, entre ellas la bacteria *Bizionia argentinensis*. Los miembros de la familia Flavobacteriaceae son bacilos gram-negativos (de lados paralelos o ligeramente irregulares y extremos redondeados a ligeramente cónicos), no formadores de esporas, quimioorganotrofos con un metabolismo primario respiratorio. La mayor parte de las especies son aerobias aunque se pueden encontrar algunas aerobias facultativas. Uno de los rasgos fisiológicos más llamativos es su nivel de pigmentación amarillo-naranja. Los dos tipos de pigmentos que se encuentran en casi todas las “flavobacterias” son carotenoides y flexirubinas, siendo estos últimos muy poco comunes en otros organismos (Figura 1.6).



Figura 1.6. Colonias en agar marino de la Flavobacteria *Polaribacter* sp. MED152. Se observa la pigmentación naranja característica. Foto tomada de González et al, 2008 [47].

La mayor parte de las especies que forman el clado marino de la familia Flavobacteriaceae son organismos psicrotolerantes o psicrófilos, y constituyen un componente mayoritario en la biomasa marina de la zona pelágica de los océanos [42, 44]. La característica más conocida de los miembros de esta familia es el procesamiento de materia orgánica, mostrando una marcada preferencia por los compuestos de alto peso molecular como proteínas y polisacáridos, rasgo que comparten con otros miembros del filo Bacteroidetes [48]. Esta característica sobresaliente, se confirma en los análisis de muchos genomas recientemente secuenciados, mostrando que estos organismos poseen un gran número de peptidasas; hidrolasas y liasas de polisacáridos; como también una gran variedad de transportadores de compuestos de alto peso molecular [49].

Debido a este rol degradador, el estilo de vida de estas bacterias en los océanos es

preferentemente unido a partículas, fitoplancton o macroalgas [50], observándose aumentos drásticos en su población durante y luego de eventos de floraciones de algas [51]. Sin embargo, en ausencia de estas superficies de unión las “flavobacterias” pueden adoptar un estilo de vida libre. En este sentido, estos microorganismos tienen un gran número de proteínas de adhesión y proteínas para deslizarse sobre superficies en un proceso conocido como “*gliding motility*”, en el cual la bacteria se adhiere a la superficie y la recorre sin el uso de flagelos. Este tipo de desplazamiento necesita de un gasto activo y muy importante de energía, interrumpiéndose automáticamente si el metabolismo energético se paraliza [52]. Un estudio reciente muestra que el número de “flavobacterias” clasificadas como aptas para realizar “*gliding motility*” esta subestimado en la familia [53], observándose en análisis genéticos que casi la totalidad de los organismos secuenciados pertenecientes a la familia Flavobacteriaceae poseen el núcleo de los genes para realizarlo.

En resumen, estas características diferencian claramente a las “flavobacterias” de otras bacterias marinas fotoautótrofas (Cyanobacteria) o bacterias heterótrofas que prefieren el uso de compuestos monoméricos y viven suspendidas en el agua de los océanos.

Hoy en día es ampliamente reconocida la extrema importancia de estos microorganismos en los procesos biogeoquímicos en los océanos, donde son partícipes fundamentales en el ciclo de carbono degradando el material orgánico disuelto (MOD) y en suspensión, el cual compone una de las mayores reservas de carbono orgánico de la biosfera [54].

Bizionia argentinensis

La cepa de *Bizionia argentinensis* JUB59^T fue aislada del agua de mar en la Antártida [55]. Concretamente, a partir de muestras de agua de la superficie del mar en la caleta Potter, cerca de la Estación Científica Jubany en la Antártida Argentina. Hasta este momento, el género *Bizionia* está compuesto por 10 especies aisladas de invertebrados marinos y ambientes marinos fríos como el agua del océano antártico y sus zonas costeras.

Consistente con las características del filo Flavobacteriacea, esta especie está compuesta por bacterias gram-negativas, no móviles, preferentemente aeróbicas, las cuales crecen sobre agar marino mostrando colonias muy pigmentadas (color naranja amarillento). Estas bacterias no presentan formación de endosporas y pueden crecer en un rango de temperatura que va de 2 a 28 °C (óptimo, 20-25 °C), pero no superior a los 30 °C.

El genoma de *B. argentinensis* se encuentra secuenciado parcialmente y los genes se

agruparon en 107 contigs [56], presentando una densidad de codificación de 85.09% y un contenido de GC de 33.77%. Se identificaron 3195 marcos abiertos de lectura (ORFs) de los cuales 979 se asignaron a proteínas “hipotéticas conservadas”. Un análisis de la composición de los genes de la cepa JUB59^T mostró diferentes características destacadas. En la figura 1.7 se muestra como se distribuyen los genes en las diferentes clases funcionales.

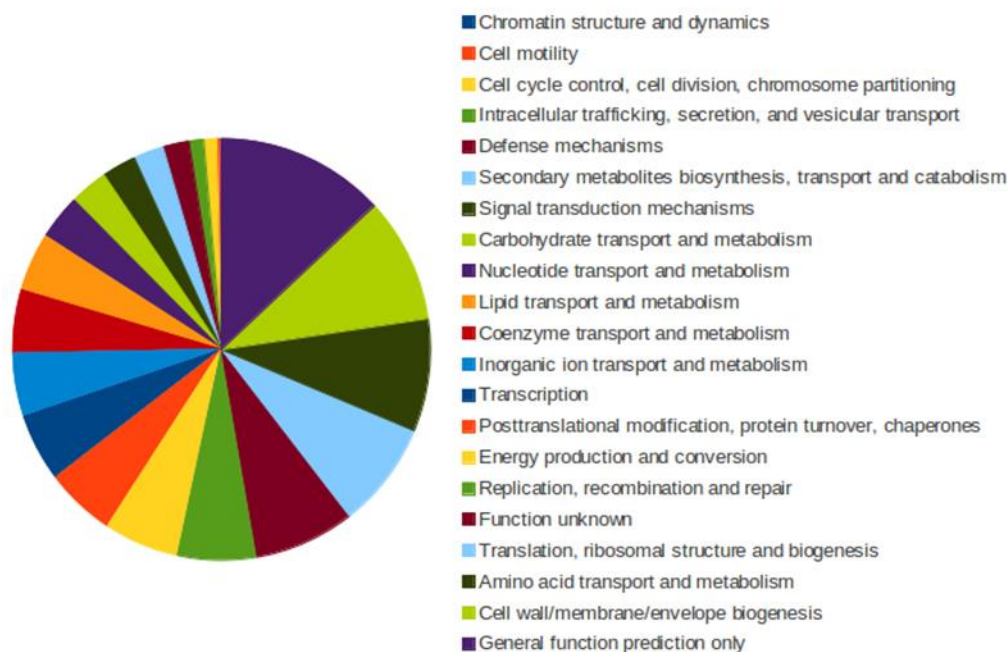


Figura 1.7. Asignación de genes a clases funcionales según la base de datos COG (*Clusters of Orthologous Groups of proteins*) [57].

La bacteria contiene todos los genes que involucran las vías de los ácidos tricarboxílicos, glucólisis y la vía anabólica de las pentosas fosfato. Así mismo, todos los genes relacionados al proceso de desnitrificación, con la excepción de la enzima nitrato reductasa. En particular, la reducción de NO^{2-} a NO parece estar catalizada por una nitrito reductasa dependiente de cobre. Por la vía asimilatoria, el NO^{2-} sería reducido por una nitrito reductasa formando amoníaco, siendo éste usado luego como fuente de nitrógeno para la síntesis de aminoácidos. Consistentemente con los altos niveles de radiación ultravioleta (UV), dadas tanto por las latitudes donde fue aislada la bacteria como por la presencia del agujero de ozono en primavera, el genoma contiene 3 genes codificantes para fotoliasas y sistemas completos de reparación de ADN. Además, la bacteria

presenta genes involucrados en la síntesis de carotenoides y xantófilas. En particular, tiene una copia de una β -caroteno hidroxilasa que cataliza la producción de β -criptoxantina y zeaxantina, pigmentos detectados previamente en otras Flavobacterias de la Antártida pero no en el género Bizionia. De acuerdo con las bajas temperaturas del ambiente, este análisis mostró un gran número de genes que codifican proteínas para la biosíntesis de ácidos grasos no saturados. Otra particularidad interesante de *B. argentinensis* es su habilidad para exportar peptidasas al medio y degradar compuestos circundantes. Se encontraron 24 genes pertenecientes a diferentes familias de peptidasas, conteniendo todos ellos péptidos señales. Esto podría permitir extender, a este miembro del género Bizionia, la relevancia del rol ecológico en el ciclo natural del carbono atribuido a otros Bacteroidetes marinos.

Capítulo I

Genómica estructural de Bizionia argentinensis

Como ya se mencionó en la introducción, los organismos adaptados al frío han evolucionado para contrarrestar los efectos potencialmente negativos de las bajas temperaturas. La vida en estos ambientes de baja energía cinética, con temperaturas cercanas a la congelación del agua requiere una amplia gama de adaptaciones en casi todos los niveles celulares. Estas características se pueden detectar con frecuencia a nivel genómico. Sin embargo, un factor determinante para el desarrollo de la vida en el frío subyace en las proteínas y sus funciones, y por lo tanto, su comprensión a nivel molecular se torna un requisito inevitable para entender esta increíble adaptación ambiental. Una gran proporción de los hábitats terrestres contienen microorganismos psicrófilos y la mayor parte de ellos se encuentra en los océanos, los cuales cubren alrededor del 70% de la superficie de la Tierra, encontrándose a una temperatura promedio por debajo de 5 °C [40]. De esta forma los océanos representan el más grande reservorio de psicrófilos. En este contexto, la bacteria *B. argentinensis* aislada en el Territorio Antártico Nacional [55], constituye un modelo muy interesante para el estudio de este tipo de microorganismos adaptados al frío y a la vida en los océanos. Además del interés como organismo psicotolerante, el género *Bizionia* es uno de los miembros marinos de la familia Flavobacteriaceae los cuales son en gran medida responsables de la remineralización de la materia orgánica en los océanos [45]. Una creciente evidencia demuestra que este grupo homogéneamente quimioheterótrofo tiene gran importancia ambiental en los ecosistemas acuáticos siendo un componente mayoritario de la biomasa microbiana de la zona pelágica del océano y estando fuertemente unido a la producción primaria (fitoplancton). Sin embargo, aún existe escasa información sobre sus ciclos biológicos, las interacciones con el ecosistema y los mecanismos moleculares que los regulan. Resulta entonces muy importante el descubrimiento de nuevas enzimas activas en frío, siendo su caracterización bioquímica y estructural un punto de partida para la comprensión de los mecanismos que permiten a las bacterias marinas adaptarse a estas condiciones de vida extremas.

En este contexto, la secuenciación del genoma de *B. argentinensis* [56] y sus características adecuadas para su manipulación en el laboratorio nos proporcionaron una oportunidad única para investigar la función de genes anotados con función desconocida que se encuentran conservados en la familia Flavobacteriaceae. De esta forma se dio inicio a un proyecto de genómica estructural de *B. argentinensis*, focalizado en la determinación estructural de proteínas de función desconocida conservadas dentro del filo Bacteriodetes. El objetivo último de esta estrategia consistió entonces en determinar la estructura tridimensional del mayor número de proteínas diana para poder entender, a través de un análisis comparativo posterior, la función de estas

biomoléculas. Una búsqueda bibliográfica mostró la casi inexistencia de información en la descripción de estudios de genómica estructural en organismos psicrófilos y psicrotolerantes. Esto generó un objetivo adicional en el presente trabajo: poner a punto y evaluar las diferentes técnicas utilizadas en los proyectos de genómica estructural de alto rendimiento utilizando como modelo a *B. argentinensis*.

Este primer capítulo de la tesis describe la estrategia de alto rendimiento utilizada para la rápida selección de candidatos con características que posibilitaran la determinación de su estructura de alta resolución por RMN o cristalografía de rayos X. Particularmente, el trabajo llevado a cabo en la presente tesis no abarca la selección bioinformática de los candidatos, realizada por el laboratorio con anterioridad.

La estrategia utilizada priorizó el estudio de los candidatos mediante RMN. Esta decisión estuvo basada en la posibilidad estratégica que tenemos como laboratorio de poder estudiar la estructura y dinámica de las proteínas dianas en solución; condiciones indispensables para comprender los mecanismos subyacentes de adaptación al frío que involucran regiones flexibles del sitio activo.

Por último, es muy importante notar, como se describió en la introducción, que muchos proyectos de esta índole han generado valiosa información sobre los organismos y sus proteínas, proporcionando datos estructurales que fueron de gran relevancia para comprender la función de estas biomoléculas y así avanzar hacia un mayor conocimiento de proteomas enteros [12, 18, 58]. Estos esfuerzos realizados con el fin de descubrir nuevas estructuras proteicas indagando en sus funciones asociadas, desarrollarán nuevos conocimientos que sin duda darán lugar a puntos de vista funcionales y biológicos inesperados.

A continuación se describen los resultados y conclusiones obtenidas de esta primera etapa del proyecto.

Selección bioinformática de los ORFs candidatos

Con el objetivo de seleccionar los ORFs candidatos buscando maximizar la elección de blancos aptos para estudios estructurales se aplicó un filtro bioinformático tomando los siguientes criterios:

1. la falta de datos estructurales y funcionales de los ORFs seleccionados.
2. la existencia de homólogos de los ORFs seleccionados en diferentes organismos.
3. el tamaño de los ORFs considerando aquellos menores a 350 aminoácidos.
4. la predicción de localización celular citosólica o extracelular.
5. el contenido de metionina, mayor al 1,1 % (para la futura incorporación de selenometionina).

Por lo tanto, en la primera fase del filtro bioinformático se seleccionaron aquellos ORFs que pertenecieran a una familia de Pfam de función desconocida. Para esto se utilizó el programa HMMER (hmmscan) que implementa modelos ocultos de Markov (HMMs) y los compara con las secuencias proporcionadas. Como resultado de este análisis se seleccionaron 979 ORFs del genoma de *B. argentinensis* que codifican proteínas (16) de baja identidad de secuencia con proteínas de función conocida. Posteriormente, a partir de estos 979 ORFs se seleccionaron aquellos que no tuvieran homólogos con estructuras depositadas en la base de datos del Protein Data Bank (PDB). Para esto se realizaron búsquedas con el programa Blastp sobre la base de datos del PDB, descartando los ORFs con hits significativos ($E\text{-value} \leq 0.001$).

Posteriormente, se verificó la existencia de secuencias ortólogas en KEGG GENES (base de datos de genomas completos) de los ORFs candidatos seleccionados. Para esto se utilizó el método bioinformático *Bidirectional Best Hits* (BBH) [59]. Este consiste en realizar una búsqueda con una secuencia proteica de un genoma de partida (en este caso, *B. argentinensis*) sobre un genoma en el cual se quiere determinar la existencia de ortólogos. Si existen secuencias similares se procede a seleccionar aquella correspondiente al mejor hit. Con esta secuencia seleccionada se realiza una nueva búsqueda, pero esta vez contra el genoma de partida. Si la secuencia correspondiente al mejor hit de esta segunda búsqueda resulta ser la secuencia de partida, se considera que los genes pueden ser ortólogos. En nuestra búsqueda solo fueron seleccionadas aquellas secuencias que dieron hits con $E\text{-value} \leq 0.00001$. Finalmente, se realizó un último paso en el cual se seleccionaron aquellas secuencias que tuvieran un tamaño menor a 350 aminoácidos, que

carecieran de péptidos señales (se utilizó el programa SignalP), y que su contenido de metionina fuera mayor al 1,1 % a fin de sustituir la metionina en las proteínas recombinantes por selenometionina para una futura resolución de estructura por cristalografía de rayos X. El esquema del filtro bioinformático generado se puede observar en la Figura 1.1. Como resultado, mediante su aplicación se seleccionaron 30 ORFs candidatos de los 3195 identificados en la bacteria.

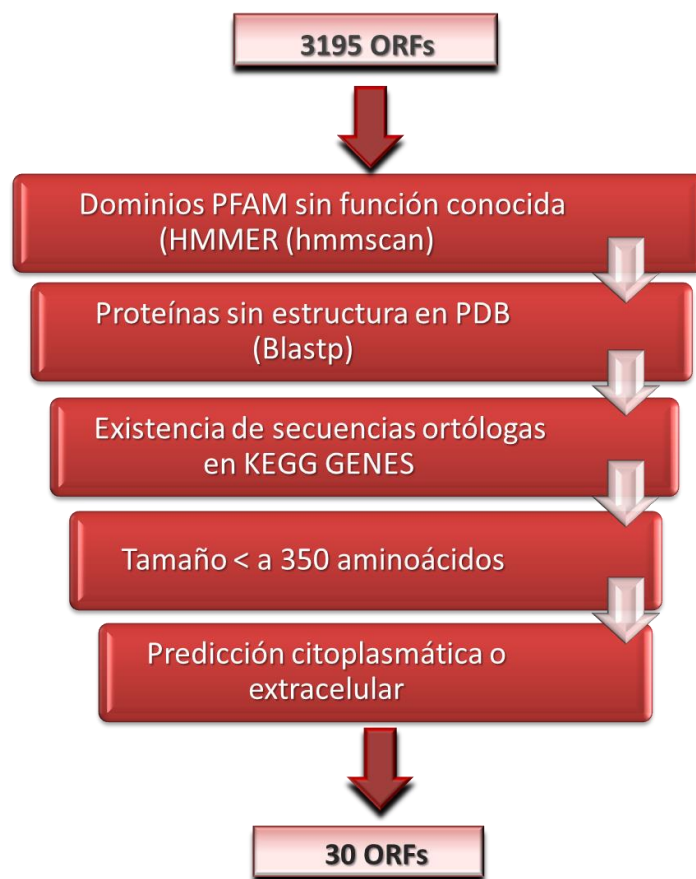


Figura 1.1. Filtro bioinformático de selección de proteínas de función y estructura desconocida para estudios estructurales por RMN y cristalografía de rayos X.

Clonado, expresión y solubilidad de los ORFs candidatos

Los 30 ORFs seleccionados se amplificaron por PCR empleando como molde el ADN genómico de *B. argentinensis* y sus respectivos productos fueron posteriormente clonados mediante el uso del sistema GATEWAY (Invitrogen, Carlsbad, CA), para lo cual se utilizaron los vectores pDONR-201 y pDEST-527 (vector de expresión). Todos los candidatos contenían en su extremo N-terminal un

sitio de clivaje para la proteína TEV seguido de un “tag” de 6X histidinas para su purificación mediante cromatografía de afinidad.

Las proteínas codificadas por cada ORF fueron producidas en células de *E. coli* en medio rico LB empleando la cepa BL21 DE3 (Stratagene, La Jolla, CA). Los niveles de expresión y solubilidad se siguieron a través de SDS-PAGE (Figura 1.2) observando las intensidades de las bandas de las proteínas recombinantes expresadas en las fracciones solubles e insolubles del lisado bacteriano.

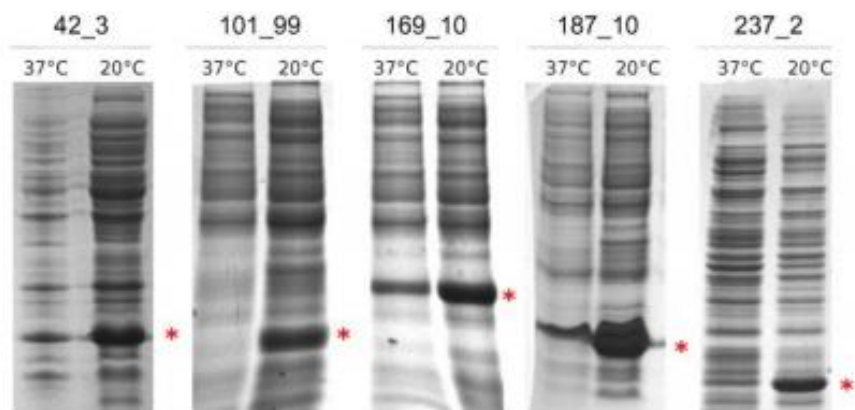


Figura 1.2. SDS-PAGE de cinco de las proteínas diana seleccionadas que mostraron expresión soluble en medio rico. La expresión se ensayó a diferentes temperaturas de inducción según se indica. Luego de la lisis celular y centrifugación, la fracción soluble se sometió a una separación por SDS-PAGE y las proteínas fueron visualizadas por tinción con Coomassie blue. *: Indica la posición de la banda correspondiente a cada proteína expresada.

Como resultado, de los 30 ORFs clonados y expresados en ensayos de pequeña escala, 17 de ellos se lograron obtener en la fracción soluble del lisado bacteriano (Tabla 1.1). De esta forma, el porcentaje de ORFs solubles logrado fue del 56%. Este porcentaje se alcanzó cuando la expresión de las proteínas fue realizada a 20 °C. Cuando la inducción se ensayó a 28 °C o 37 °C, temperaturas utilizadas clásicamente para la expresión de proteínas recombinantes (provenientes en su mayoría de organismos mesófilos), los porcentajes de proteínas solubles cayeron al 10% y 0%, respectivamente. En este caso, al tratarse de un organismo psicrófilo, si bien fue esperable este comportamiento, en la bibliografía no hay reportes hasta el momento de este tipo de procedimientos de expresión a gran escala. Por último, se evaluó la expresión y solubilidad de los candidatos en un medio mínimo con el fin de llevar a cabo el protocolo de marcado isotópico con ^{15}N y ^{13}C , condición necesaria para el estudio de los candidatos por RMN (ver más adelante). Los

resultados mostraron que la mayoría de ellos se expresó en buen nivel en medio mínimo, observándose solo en algunos candidatos una gran disminución de su expresión.

ORF ID	Soluble
BA_3_15	Si
BA_5_34	Si
BA_6_46	Si
BA_8_56	Si
BA_13_18	Si
BA_27_18	No
BA_30_17	Si
BA_31_44	Si
BA_47_29	No
BA_53_45	Si
BA_212_10	Si
BA_300_2	No
BA_2_7	No
BA_5_69	No
BA_6_27	No
BA_7_45	Si
BA_40_17	Si
BA_43_2	No
BA_58_6	No
BA_64_66	No
BA_80_14	Si
BA_123_14	No
BA_119_21	No
BA_121_2	No
BA_42_3	Si
BA_101_99	Si
BA_75_21	No
BA_150_2	Si
BA_169_10	Si
BA_237_2	Si

Tabla 1.1. ORFs seleccionados como blancos potenciales para futuros estudios estructurales. Los 30 ORFs fueron clonados y sometidos a pruebas de expresión y solubilidad. La expresión de las proteínas heterólogas se llevó a cabo en células de *E. coli* de la cepa BL21 (DE3). La numeración de cada ORF refiere a la posición y el contig que integra el mismo (BA_posición en el contig_número de contig).

Evaluación de los candidatos por RMN (1D ^1H y 2D HSQC ^1H - ^{15}N)

El siguiente paso fue evaluar cuáles de los candidatos solubles cumplían con los requisitos para su resolución estructural por RMN o cristlografía de RX. Para esto, los candidatos que mostraron buenos niveles de expresión en medio mínimo fueron sometidos a una marcación selectiva con ^{15}N .

En esta estrategia se utilizó el método descrito por Almeida et al. [60], el cual consiste en el uso del antibiótico rifampicina para lograr únicamente la marcación de las proteínas heterólogas. Esto se logra gracias a que la rifampicina inhibe a la ARN polimerasa bacteriana pero no a la ARN polimerasa T7 usada para la transcripción de las proteínas recombinantes en sistemas de expresión basados en el promotor T7 para el control de la transcripción. Con esta metodología las bacterias ya crecidas son expuestas a rifampicina para posteriormente resuspenderlas en un medio mínimo isotópicamente marcado para la inducción de la expresión de las proteínas recombinantes. Finalmente, se obtiene mayoritariamente la proteína recombinante marcada selectivamente con el isótopo deseado.

Luego de expresar los candidatos marcados selectivamente con ^{15}N , la fracción soluble del lisado bacteriano se sometió a experimentos de RMN 1D ^1H y 2D HSQC ^1H - ^{15}N , a modo de seleccionar cuales de las 17 proteínas eran buenos candidatos para llevar a cabo estudios estructurales.

A través del análisis de la dispersión del desplazamiento químico se puede obtener información estructural [16, 61]. También, el ancho de las resonancias observadas aporta información inherente a la estructura proteica. Así, se puede tener una idea del estado conformacional de cada proteína en solución, observando cuan estables y estructurados se pueden encontrar los diferentes candidatos.

De esta forma, las proteínas solubles y con un alto grado de expresión en medio mínimo fueron clasificadas según la característica de sus espectros de RMN HSQC ^1H - ^{15}N siguiendo los siguientes criterios:

- i)* Las proteínas que presentaron espectros de HSQC conteniendo un número de picos similar al número de aminoácidos de su cadena polipeptídica, que resultaron ser homogéneamente intensos y con buena dispersión, fueron clasificadas como buenos candidatos (Figura 1.3.A). Este tipo de espectros es característico de proteínas plegadas con estructura definida y estable.
- ii)* Las proteínas que presentaron espectros de HSQC conteniendo un número de picos mayor al 60% del número de aminoácidos de su cadena polipeptídica, que no resultaron ser

homogéneamente intensos, aunque algunos mostraron buena dispersión, fueron clasificadas como candidatos prometedores (Figura 1.3.B). Este tipo de espectros es característico de proteínas con plegamiento definido, pero que, ya sea por su tamaño, grado de oligomerización, grado de exposición al solvente o flexibilidad de alguna de sus regiones, muestra un espectro heterogéneo.

iii) Las proteínas que presentaron espectros de HSQC conteniendo un número de picos menor al 30% del número de aminoácidos de su cadena polipeptídica, cuyos picos no resultaron ser homogéneamente intensos y tenían una relación señal/ruido muy baja comparada con la cantidad de proteína en la muestra, fueron clasificadas como pobres candidatos (Figura 1.3.C). Esta categoría probablemente incluye proteínas que forman grandes entidades solubles tales como oligómeros de orden superior (tetrámeros o mayores) o agregados no específicos, también puede incluir proteínas que están experimentando fluctuaciones en su conformación en el rango intermedio de la escala de tiempo de RMN.

En el caso de las proteínas solubles que mostraron bajos niveles de expresión en medio mínimo, no se les realizó en primera instancia el marcado isotópico y fueron clasificadas según la característica de sus espectros unidimensionales de RMN ^1H (1D) siguiendo los siguientes criterios:

i) Las proteínas que presentaron espectros 1D con buena dispersión y resolución de picos en la región de las amidas (aproximadamente 6,5-9 ppm) y se observaron picos definidos en la región de los metilos (-1-0 ppm) fueron clasificadas como buenos candidatos (Figura 1.4.A). Este tipo de espectros son característicos de proteínas plegadas con estructura definida y estable.

ii) Las proteínas que presentaron espectros 1D con relativa buena dispersión y resolución de picos en la región de las amidas (6,5-9 ppm), aunque no se observaron picos definidos en la región de los metilos (-1-0 ppm) fueron clasificadas como candidatos prometedores (Figura 1.4.B).

iii) Las proteínas que presentaron espectros 1D con mala dispersión y baja resolución de picos en la región de las amidas (6,5-9 ppm) y no se observaron picos definidos en la región de los metilos (-1-0 ppm) fueron clasificadas como pobres candidatos. Estos espectros son característicos de proteínas sin plegamiento definido, muy flexibles o agregadas, lo que dificulta la resolución de su estructura (Figura 1.4.C).

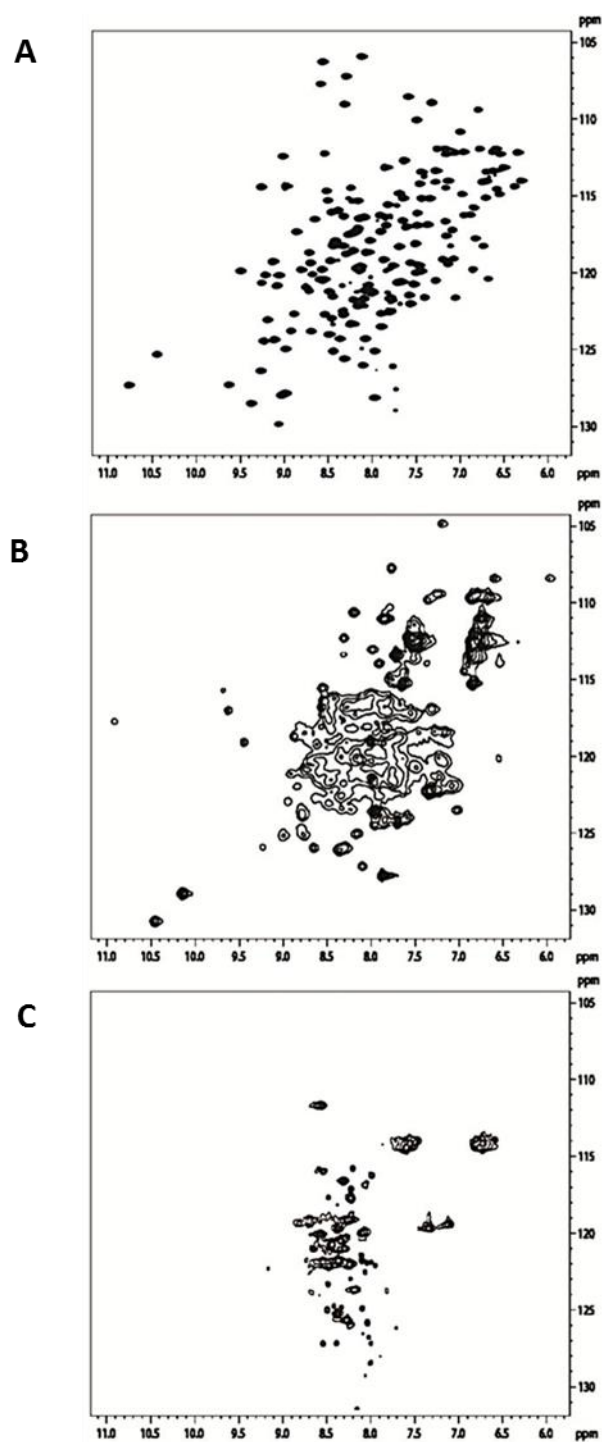


Figura 1.3. Experimentos de RMN HSQC ^1H - ^{15}N de las proteínas solubles con un alto grado de expresión en medio mínimo. Después del marcado selectivo con ^{15}N , las proteínas fueron analizadas por HSQC ^1H - ^{15}N . Dependiendo de la calidad de los espectros, las proteínas se clasificaron en función de la viabilidad de los futuros estudios estructurales: (A) candidatos buenos, (B) candidatos prometedores y (C) candidatos pobres.

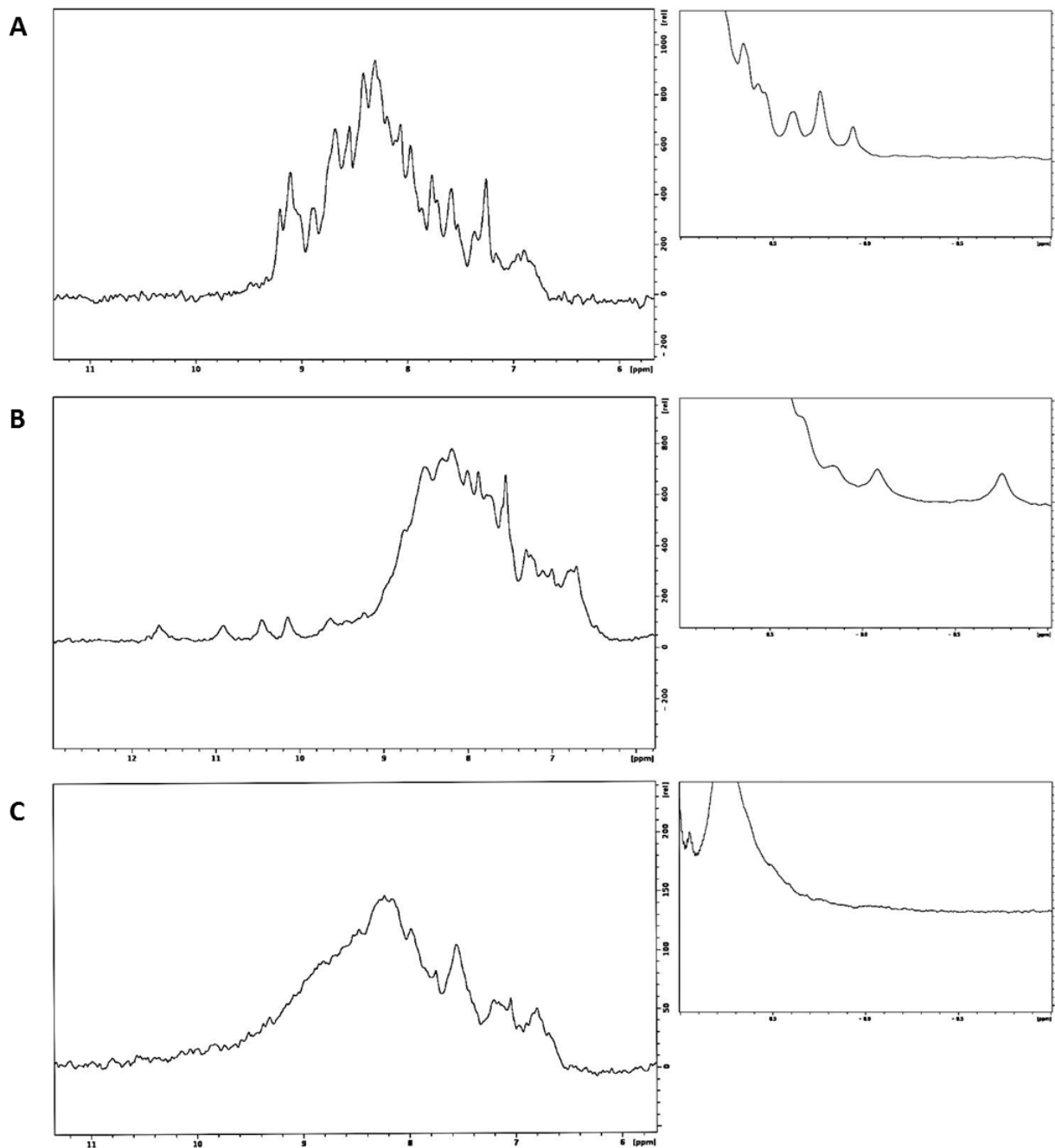


Figura 1.4. Experimentos de RMN 1D ^1H de las proteínas solubles que mostraron bajos niveles de expresión en medio mínimo. Las proteínas que no fueron marcadas selectivamente con ^{15}N (principalmente debido a sus bajos niveles de expresión dentro de la célula) fueron purificadas en pequeña escala por cromatografía de afinidad, concentradas y examinadas mediante experimentos unidimensionales de RMN ^1H . En los espectros sobre la izquierda se muestra la región de los protones amídicos, mientras que en los espectros de la derecha se presenta la región de los protones metílicos. Se clasificaron como: **(A)** candidatos buenos, **(B)** candidatos prometedores y **(C)** candidatos pobres.

Mediante este filtro estructural se lograron clasificar: 3 candidatos buenos, 5 candidatos prometedores y 7 candidatos pobres (Figura 1.5). Las 2 proteínas restantes se precipitaron en los pasos posteriores de purificación indicando su baja estabilidad y, por lo tanto, fueron descartadas.

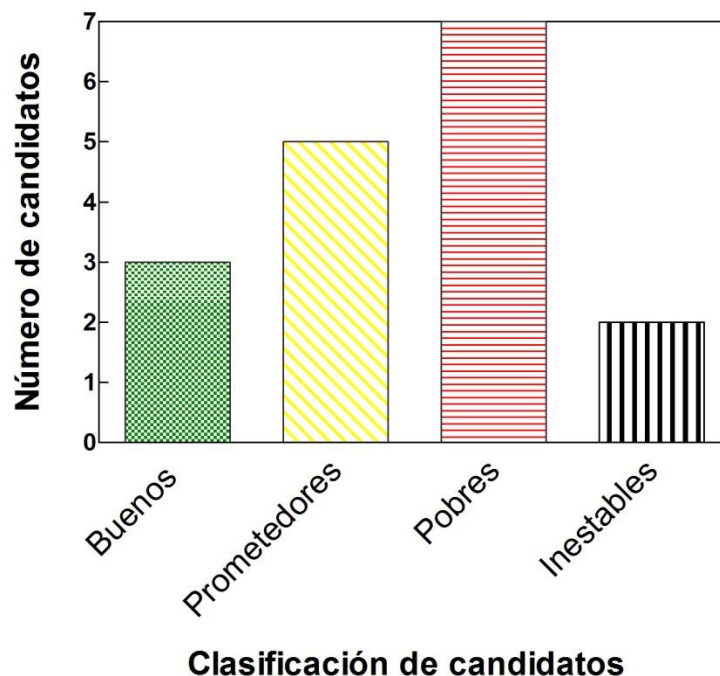


Figura 1.5. Gráfico de barras mostrando la clasificación obtenida mediante los experimentos de RMN 1D ^1H y 2D HSQC ^1H - ^{15}N de los 17 candidatos obtenidos de forma soluble.

Screening de cristalogenésis

En una primera instancia sólo aquellas proteínas que se clasificaron como candidatos buenos o prometedores fueron sometidas a un *screening* cristalográfico. Posteriormente se tomó la decisión de incluir en éste a las proteínas clasificadas como pobres. La razón de esta incorporación consistió en que la información estructural obtenida de los experimentos de RMN 1D ^1H y 2D HSQC ^1H - ^{15}N para los candidatos clasificados como pobres no es concluyente al momento de asignar una probabilidad de cristalización de la proteína en cuestión [29, 61]. Las proteínas clasificadas como pobres pueden mostrar estos espectros debido a oligomerización o estados agregados, condiciones en las cuales las proteínas son factibles para la resolución cristalográfica de su estructura 3D. Así mismo, diferentes trabajos han mostrado una complementariedad prácticamente sin redundancias en la resolución de estructuras a través de la cristalografía de rayos X y RMN en proyectos de genómica estructural [61].

De esta forma se inició la producción de los candidatos buscando obtener concentraciones de proteína recombinante del orden de 15 mg/ml. Para obtener cristales adecuados para los estudios de cristalografía, las proteínas deben tener un alto grado de pureza, siendo la homogeneidad de la preparación un factor clave en la obtención de cristales que se difractan a alta resolución [62]. Por lo tanto, la purificación de las proteínas se llevó a cabo utilizando cromatografía de afinidad, exclusión molecular e intercambio iónico. Se buscó lograr un porcentaje de pureza del 98% o superior.

El método utilizado para la cristalización fue el de difusión del vapor, utilizando la técnica de *"Sitting y Hanging drop"*. La cristalización requiere llevar la macromolécula a la sobresaturación. En este popular método una gota de la solución que contiene una mezcla del precipitante y la proteína se deposita en una cámara la cual contiene precipitante puro en su interior, dejándola finalmente sellada. Con el tiempo ocurre la difusión del vapor de agua de la gota hasta que la osmolaridad en ésta es igual a la de la solución pura de precipitante. Esta deshidratación de la gota causa un lento cambio en las concentraciones de proteína y precipitante hasta que se alcanza un equilibrio, que idealmente ocurre en la zona de nucleación del cristal en el diagrama de fases. En estos ensayos iniciales se utilizó un sistema robotizado para placas de 96 wells Honeybee 963 (Digilab, Marlborough, EEUU) y kits comerciales de cristalización, trabajando a escalas del orden de los 300 nl. A la fecha se realizó el screening cristalográfico de 9 candidatos ensayando 327 condiciones de cristalogénesis para cada uno. Dos de las proteínas ensayadas mostraron formación de cristales en diferentes condiciones de solución precipitante.



Figura 1.6. Se muestran los cristales generados de la proteína BA_212_10. Se analizaron 327 condiciones de cristalogénesis a una concentración de proteína de 15 y 7,5 mg/ml. La solución cristalogénica final contenía 12% (p/v) PEG 8000, 0,1 M citrato de sodio y 0,1 M sulfato de amonio.

Las 2 proteínas que presentaron “Hits” positivos de cristalización fueron sometidas posteriormente a la optimización manual de las condiciones de cristalogénesis (Figura 1.6). Se ajustaron las variables iniciales de las condiciones de los kits comerciales que dieron cristales (concentración y naturaleza de reactivos, pH, temperatura de incubación, etc) y se aumentó el volumen de la solución proteica en la escala de los μl .

Estructuras resueltas

Como resultado de la utilización del método de alto rendimiento descrito para la selección de proteínas de función desconocida adecuadas para estudios estructurales, se logró determinar hasta el momento la estructura tridimensional de dos de los 30 candidatos seleccionados, las proteínas BA_42_3 (clasificada como buena) y BA_212_10 (clasificada como pobre), nombradas BA42 y C24, respectivamente. La estructura de alta resolución de la proteína BA42 fue determinada mediante procedimientos estándares de RMN y cristalografía de rayos X [63] (Figura 1.7.A); su descripción estructural y bioquímica son parte fundamental de esta tesis y se detallarán en el siguiente capítulo.

La estructura de alta resolución de la proteína C24 se pudo resolver mediante cristalografía de rayos X (Figura 1.7.B) en colaboración con la Dra Jimena Rinaldi y el Dr. Sebastián Klinke de la Plataforma de Biología Estructural y Metabólica (Plabem). En este caso, los cristales difractaron con una resolución de 1.82 Å en el grupo espacial $P2_12_12_1$. La estructura cristalina se resolvió mediante el método de dispersión anómala de longitud de onda única. Aunque se realizaron estudios bioinformáticos y experimentales sobre esta proteína, estos son aún preliminares y no están dentro de los objetivos de esta tesis. A continuación se resume la información obtenida con el fin de brindar conocimiento sobre el estado de estos estudios hasta el momento.

Comparando la estructura 3D de C24 con las bases de datos existentes, encontramos que la proteína C24 presenta una estructura similar al extremo de las fibras de cola del bacteriófago T4, cuyo rol es el reconocimiento y la unión de los receptores en la membrana de la célula hospedadora. Un análisis bioinformático de las regiones flanqueantes del gen C24 en el genoma de *B. argentinensis* reveló la presencia de un gran número de genes de función desconocida y un menor número de genes relacionados con proteínas virales. Hasta el momento se desconoce la función de C24, si bien estos resultados sugieren que la proteína podría pertenecer a un bacteriófago integrado en el genoma de *B. argentinensis*, también existe la posibilidad de que haya evolucionado a partir de un evento de integración viral adquiriendo una nueva función en la

bacteria. Más estudios deben realizarse para indagar en el rol de C24.

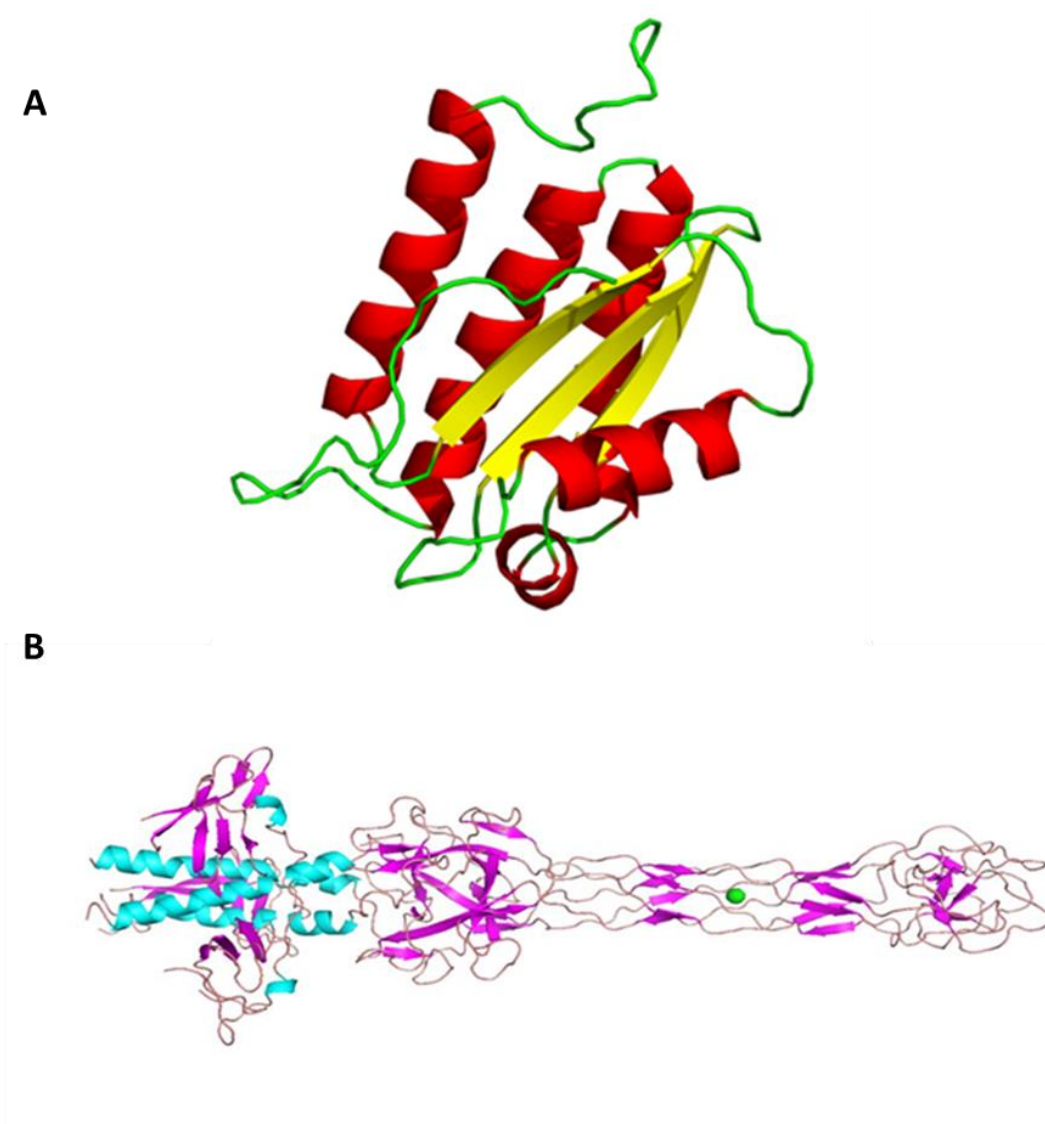


Figura. 1.7. Estructuras resueltas en el proyecto de genómica estructural de la bacteria *B. argentinensis*. (A). Estructura resuelta por RMN de BA42 (códigos PDB 2MPB). En rojo se indican las α -hélices, en amarillo las hebras β y en verde segmentos sin estructura secundaria (B). Estructura cristalina de C24 (código PDB “on hold”). La proteína es un homo trímero que posee un átomo de Mn^{2+} estructural que presenta una coordinación octaédrica con seis residuos de histidina (dos de cada monómero) mediante el motivo His-X-His.

Hasta este momento, el porcentaje de estructuras resueltas con este método de alto rendimiento es de 6,6 %, aunque es importante mencionar que falta proceder a la resolución de la estructura mediante RMN de dos candidatos restantes clasificados como buenos (que no cristalizaron), y realizar la exploración de las condiciones experimentales para intentar conseguir cristales de otros 6 candidatos y poder obtener la estructura 3D de alguno de ellos.

Discusión

A través del presente trabajo se logró poner punto y evaluar las diferentes técnicas utilizadas en los proyectos de genómica estructural de alto rendimiento para la bacteria psicotolerante *B. argentinensis*. En este estudio nos focalizamos en proteínas y dominios sin función conocida (proteínas “hipotéticas conservadas” y DUFs). En primera instancia se seleccionaron bioinformáticamente 30 candidatos, de los cuales 17 mostraron buenos niveles de expresión y solubilidad solamente a 20 °C. Los resultados muestran la importancia de la búsqueda de las condiciones óptimas de expresión a fin de maximizar el número de proteínas solubles. En los proyectos de genómica estructural, el primer gran cuello de botella ocurre en la etapa de expresión, disminuyendo el número de candidatos rápidamente. Si bien en esta etapa se perdieron aproximadamente la mitad de los candidatos por expresarse de forma insoluble, el rendimiento de expresión soluble coincidió con el logrado por otros proyectos de genómica estructural [29, 61, 64]. También, estos resultados nos sirvieron para realizar una evaluación de los programas de predicción de solubilidad en la expresión de proteínas en organismos procariontes. Comparando la predicción bioinformática con los datos experimentales se observó que estos programas predijeron en un 40 a 60% la solubilidad de las proteínas heterólogas (Tabla 1.2). Este hecho muestra que todavía es muy difícil predecir bioinformáticamente la expresión y solubilidad de proteínas recombinantes, quedando en evidencia que la expresión de estas moléculas es un proceso idiosincrático inherente a cada proteína.

ORF ID	Solubilidad	Predicción solubilidad		
	Experimental	Prosoll	SPpred	AGGRESKAN
BA_3_15	Si	No	No	Si
BA_5_34	Si	Si	Si	Si
BA_6_46	Si	No	Si	Si
BA_8_56	Si	No	Si	Si
BA_13_18	Si	Si	No	Si
BA_27_18	No	No	No	Si
BA_30_17	Si	No	No	Si
BA_31_44	Si	No	Si	Si
BA_47_29	No	No	No	Si
BA_53_45	Si	Si	Si	Si
BA_212_10	Si	No	No	Si
BA_300_2	No	Si	No	Si
BA_2_7	No	Si	No	Si
BA_5_69	No	Si	No	Si
BA_6_27	No	Si	No	Si
BA_7_45	Si	Si	No	Si
BA_40_17	Si	Si	Si	Si
BA_43_2	No	No	No	Si
BA_58_6	No	Si	No	No
BA_64_66	No	No	Si	Si
BA_80_14	Si	Si	Si	No
BA_123_14	No	Si	No	No
BA_119_21	No	No	Si	No
BA_121_2	No	Si	No	No
BA_42_3	Si	No	No	Si
BA_101_99	Si	Si	No	No
BA_75_21	No	No	Si	Si
BA_150_2	Si	No	No	Si
BA_169_10	Si	No	Si	Si
BA_237_2	Si	No	No	Si
Porcentaje de predictibilidad		43%	60%	63%

Tabla 1.2. Solubilidad experimental y predicha bioinformáticamente de los ORFs seleccionados como blancos potenciales para los estudios estructurales. Se muestra la predicción de los programas Prosoll (<http://mips.helmholtz-muenchen.de/prosoll/prosoll.seam>), AGGRESKAN (<http://bioinf.uab.es/aggrescan>), SPpred (<http://crdd.osdd.net:8081/sppred/submit.jsp>). Se detallan los porcentajes de predictibilidad para cada programa que resultan de comparar los datos experimentales con la predicción bioinformática.

La principal diferencia entre realizar una clasificación mediante experimentos 1D ^1H o 2D HSQC ^1H - ^{15}N radica en que a través de experimentos HSQC se consigue una mejor evaluación de las características de la proteína para la determinación de la estructura 3D. Sin embargo, como se ha demostrado en algunos trabajos [29], una comparación de la clasificación de los candidatos

obtenida mediante experimentos 1D ^1H y 2D HSQC ^1H - ^{15}N mostró que la utilización de experimentos 1D ^1H es adecuada en la implementación de un método de selección de candidatos para la determinación de su estructura por RMN. Este hecho plantea la discusión de la posible eliminación del marcado isotópico con ^{15}N para el *screening* de proteínas reduciendo los costos del trabajo.

Como ya se mencionó, un espectro de HSQC con picos bien dispersos es indicativo de que la proteína adopta una conformación compacta, estable y que se encuentra rotando libre e isotrópicamente en la solución. En línea con esto, un análisis de la categorización realizada por RMN en función de la masa molecular de las proteínas muestra que aquellas clasificadas como “buenos” candidatos poseen una masa molecular promedio de 13,8 kDa, mientras que para los candidatos “prometedores” y “pobres” estas fueron de 22,3 kDa y 23,9 kDa, respectivamente. Este resultado muestra la ya conocida influencia de la masa molecular en la obtención de buenos espectros.

En cuanto al número de estructuras resueltas en este proyecto, se observó una concordancia con los resultados de otros estudios de genómica estructural, corroborándose la complementariedad de la cristalografía de rayos X y la RMN en la resolución de las estructuras [61]. En este trabajo se consiguió la puesta a punto de la metodología consiguiendo que el éxito del proyecto de genómica estructural de un organismo psicrófilo, en relación al número de estructuras determinadas (6,6 % de éxito en la resolución estructural), resultara muy similar al obtenido sobre otros organismos termófilos y mesófilos [64]. Finalmente, es importante concluir que a través de este estudio de genómica estructural, conseguimos identificar nuevos pares secuencia-estructura, accediendo en última instancia a la función de las proteínas como se describe en el capítulo siguiente sobre la proteína BA42, abriéndose un camino hacia el estudio de su rol biológico.

Capítulo II

Caracterización estructural y funcional de la proteína BA42

Como se detalló en el capítulo anterior, el primer candidato seleccionado a partir del proyecto de genómica estructural de alto rendimiento de la bacteria *B. argentinensis* fue la proteína BA42. En este contexto, nuestro grupo resolvió en primera instancia su estructura en solución mediante RMN (código PDB 2MPB) [63] (Anexo Tabla A.1). Estructuralmente, BA42 se constituye de un único dominio denominado TPM. Los dominios TPM (Pfam PF04536) representan una súper familia pobremente caracterizada de dominios de proteínas que se encuentran ampliamente distribuidos en eukarya y eubacteria. En los últimos años, estos TPM han sido asociados a diversos procesos celulares como la fotosíntesis y la transmisión sináptica [65, 66]. Esta familia de dominios, (denominada anteriormente como "DUF477" y "Repair_PSII"), fue nombrada a partir de las tres primeras proteínas estudiadas; TLP18.3, Psb32 y MOLO-1 [66, 67]. La estructura de cuatro dominios TPM presentes en varias proteínas fueron determinadas (CG2496 de *Corynebacterium glutamicum* [código PDB 2KPT], PG0361 de *Porphyromonas gingivalis* [código PDB 2KW7], BA42 de *B. argentinensis* [códigos PDB 2MPB y 4OA3] [63] y AtTLP18.3 de *Arabidopsis thaliana* [código PDB 3PTJ] [66] (Figura 2.1 A).

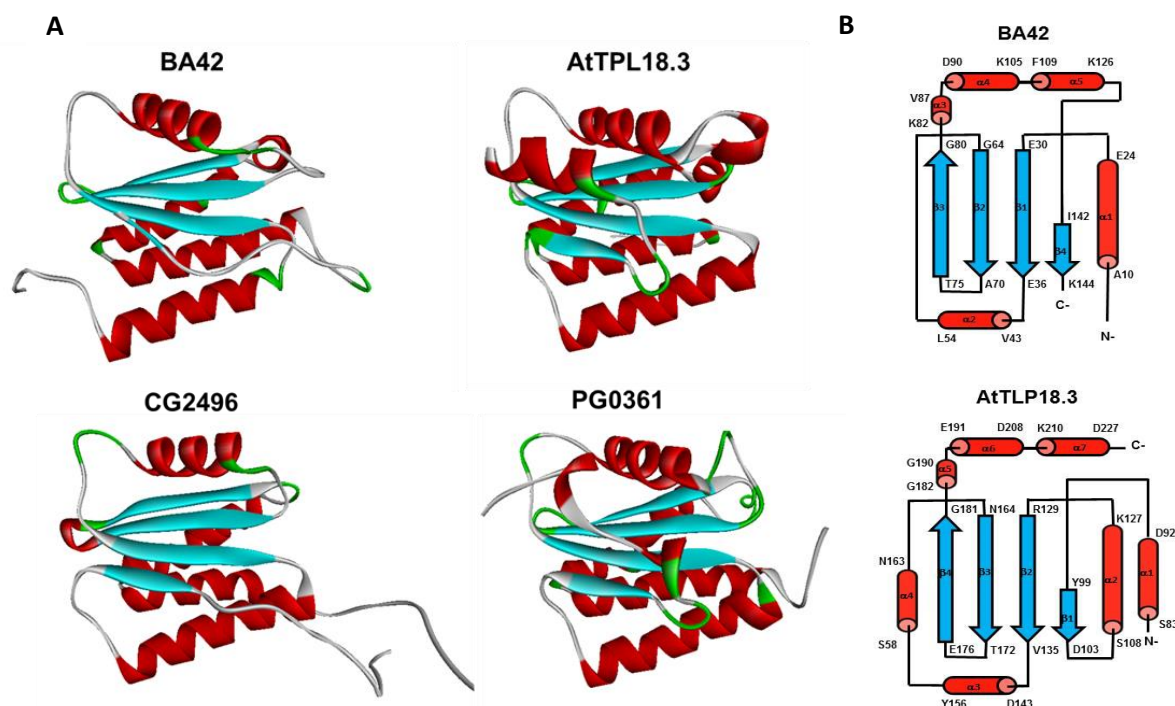


Figura 2.1. BA42 y sus homólogos estructurales. (A) Las proteínas CG2496, PG0361 y AtTLP18.3 son dominios TPM conservados. La proteína AtTLP18.3 ha sido descrita como una fosfatasa ácida de los tilacoides, mientras que CG2496 y PG0361 carecen de una función descrita. (B) Topología de plegamiento de BA42 comparada con la observada para el resto de los dominios TPM con estructura conocida hasta el momento.

Estas estructuras, a pesar de compartir una identidad de secuencia muy baja (menos del 16%), presentan una conformación $\alpha\beta\alpha$ "sandwich" y una topología de plegamiento similares (Figura 2.1), que no se encontraron previamente en otros dominios de proteínas. Interesantemente, si bien BA42 presenta una conformación $\alpha\beta\alpha$ similar, posee una topología de plegamiento diferente, no reportada hasta el momento para un miembro de esta familia de dominios (Figura 2.1.B).

La función de CG2496, PG0361 y BA42 sigue siendo desconocida, mientras que la proteína AtTLP18.3 se clasificó como una fosfatasa ácida de los tilacoides [66], vinculándose su función a la protección parcial del complejo PSII como parte de la resistencia a la fotoinhibición en condiciones fluctuantes de alta intensidad lumínica [68]. Además, de estos cuatro miembros de la familia TPM con estructura conocida, otras proteínas que contienen dominios TPM han sido caracterizadas funcionalmente hasta el momento. MOLO-1 fue identificado como un regulador positivo de los receptores de acetilcolina sensibles al levamisol en la unión neuromuscular en *Caenorhabditis elegans* [69]. La proteína Psb32 de *Synechosystis sp.*, de forma similar a AtTLP18.3, protege el fotosistema II del daño fotoquímico acelerando su reparación. También, recientemente, un análisis funcional de la proteína Rv2345 de la bacteria patógena *Mycobacterium tuberculosis* [70] mostró que esta posee actividades de fosfatasa y ATPasa. Por último, se ha demostrado que el dominio TPM de la proteína CG2496 de *Corynebacterium glutamicum* interactúa con metiotepina, un antagonista de los receptores de serotonina y un potencial compuesto para la generación de antibióticos [71]. Aunque se ha logrado cierto progreso en la comprensión del papel celular de las proteínas que contienen dominios TPM, todavía no está clara la relación entre estructura y función de esta familia de dominios. De hecho, sólo hay un miembro de la familia TPM con estructura conocida funcionalmente caracterizado hasta la fecha [66].

En este contexto, en este capítulo se describe el estudio estructural y bioquímico de la proteína BA42.

BA42 muestra una arquitectura $\alpha\beta$ de tres capas y muestra una conformación rígida en solución

La estructura de BA42 muestra una nueva variante topológica del plegamiento α/β que consta de cinco α -hélices y una hoja- β de cuatro hebras con la topología D (\uparrow) A (\uparrow) B (\uparrow) C (\downarrow). Las α -hélices I, IV y V interaccionan con uno de los lados de la hoja- β , mientras que la hélice II se sitúa en el lado opuesto, constituyendo una estructura sándwich $\alpha\beta\alpha$. La α -hélice III es perpendicular al plano de la capa media del sándwich. Las ubicaciones de los elementos de estructura secundaria son: hebra- β A (residuos 30-36), B (64-70), C (75-80), y D (142-144), y α -hélice I (10-23), II (43-53), III (82 a 85), IV (94 a 104), y V (109-127) (Figura 2.2).

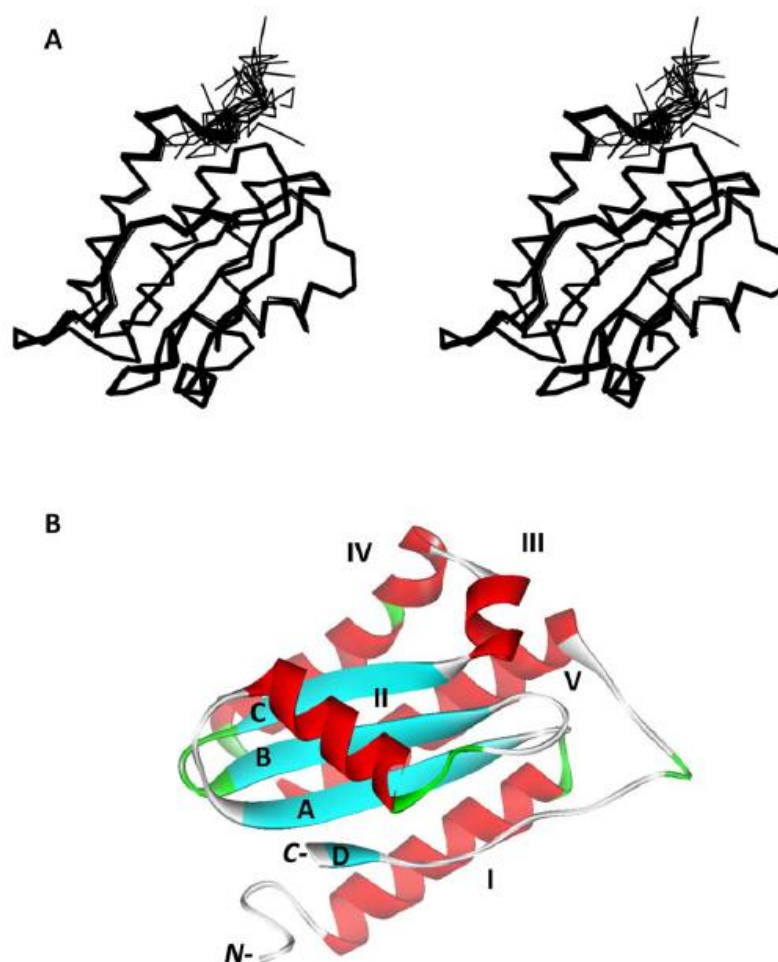


Figura 2.2. Estructura en solución de BA42. (A) Vista estero de la estructura en solución de BA42 (código PDB 2MPB). Se muestra la superposición de las 20 estructuras de mínima energía. (B) Modelo de cinta de la estructura más cercana a la media, las α -hélices se detallan en color rojo, la hoja- β en cian y los giros en verde.

El primer objetivo fue determinar el estado multimérico de BA42 en solución. Con este fin realizamos en primera instancia la evaluación del radio hidrodinámico de la proteína mediante experimentos de campo de gradiente pulsado (PFG- *Pulsed Field Gradient*) de RMN. Los resultados mostraron que BA42 se comporta como una proteína globular monomérica según el valor de radio hidrodinámico experimental obtenido ($R_h = 20,3 \pm 0,2 \text{ \AA}$).

Con el fin de poder determinar las dinámicas globales y locales de BA42 se utilizaron mediciones de relajación del ^{15}N . El tiempo de correlación de BA42 calculado mediante la relación T_1/T_2 promedio fue de $8,0 \pm 0,2 \text{ ns}$ a 28°C , de acuerdo con el valor empírico de una proteína globular de este peso molecular en un medio acuoso a 301 K [72]. Este resultado está en línea con el radio hidrodinámico discutido anteriormente, indicando que BA42 se comporta como un monómero en solución, incluso en la alta concentración utilizada para el estudio estructural (alrededor de 0,8 mM). En la figura 2.3 se observa el gráfico con los valores experimentales de T_1/T_2 ^{15}N , así como los calculados utilizando el programa HYDRONMR [73] para la estructura de más baja energía asumiéndola como un cuerpo rígido. La comparación en el gráfico muestra una muy buena concordancia con la mayoría de los residuos, a excepción de la región de 4-7, que presenta un valor inferior de T_1/T_2 , típico de regiones proteicas que exhiben movimientos rápidos. El resto de la proteína muestra una moderada anisotropía, observándose que los valores de T_1/T_2 varían en el rango de 8,5-6,3.

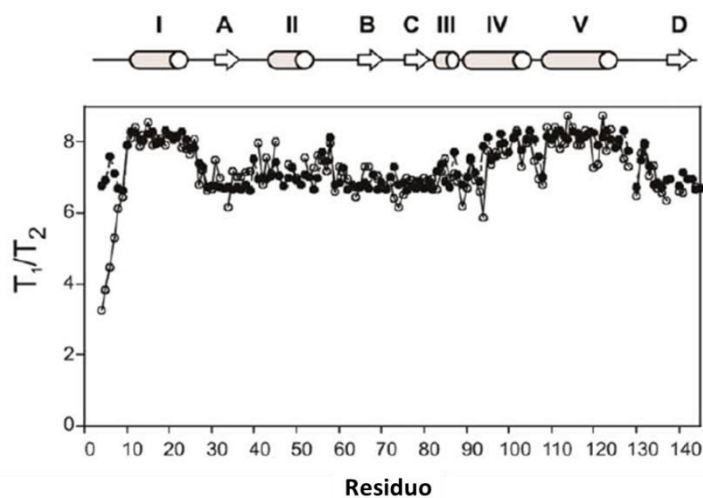


Figura 2.3. Comparación de los valores T_1/T_2 ^{15}N de BA42. Los valores experimentales T_1/T_2 ^{15}N se detallan como círculos abiertos y los valores calculados utilizando el programa HYDRONMR como círculos negros. Sobre el gráfico se muestra la estructura secundaria de BA42; las α -hélices están representadas por cilindros grises y las hebras- β flechas blancas.

Por lo tanto, estos resultados muestran que los primeros siete residuos de BA42 constituyen la única región que presenta flexibilidad local, mientras que el resto de la proteína muestra una marcada rigidez. En particular, una larga región de la proteína, que comprende los residuos 127-141, no muestra ninguna estructura secundaria asociada, pero al mismo tiempo las velocidades de relajación del ^{15}N (Figura 2.3) indican la ausencia de flexibilidad.

BA42 muestra una característica estructural infrecuente: un par de “crossing loops”

Un examen cuidadoso de la estructura en solución de BA42 reveló la presencia de una característica peculiar: el *loop* que conecta la última α -hélice V con el extremo C-terminal de la hebra- β D cruza sobre el *loop* formado entre el extremo N-terminal de la α -hélice I y la primera hebra- β A. Este par de “crossing loops” es un evento muy poco frecuente en las estructuras de proteínas y se cree que es energéticamente desfavorable [74, 75]. En la búsqueda de características que pudiesen ser responsables de la estabilización de la conformación con el extremo C-terminal pegado al núcleo estructurado de la proteína generando el par de “crossing loops”, se realizó la exploración de condiciones experimentales para poder cristalizar la proteína BA42. En dicha exploración se obtuvieron cristales de alta calidad de BA42, los cuales difractaron los rayos X a una resolución máxima de 1,39 Å en el grupo espacial $P2_12_12_1$ (código PDB 4OA3).

La estructura cristalina se resolvió por reemplazo molecular a partir de la determinada para BA42 por RMN en colaboración con el Dr. Sebastián Klinke y el Dr. Lisandro Otero de la Plabem. Las coordenadas finales muestran una excelente estereoquímica, con más de 98% de los residuos cayendo en las regiones permitidas del gráfico de Ramachandran (Anexo Tabla A.2). Se observa densidad electrónica continua en la cadena polipeptídica sin espacios vacíos. Solo unas pocas cadenas laterales correspondientes a los residuos polares expuestos no pudieron ser localizadas en la densidad electrónica, y por lo tanto, fueron solamente modelados sus C^β . Tampoco se detectó densidad electrónica para los primeros cuatro residuos. Este hecho concuerda con la flexibilidad de los primeros siete residuos en solución descrita anteriormente a través de mediciones de relajación del ^{15}N . Una comparación de la estructura cristalina y en solución de BA42 muestra que son muy similares con respecto a la conformación de la cadena principal, con una desviación cuadrática media (RMSD) para los C^α de 0,9 Å (Figura 2.4.A).

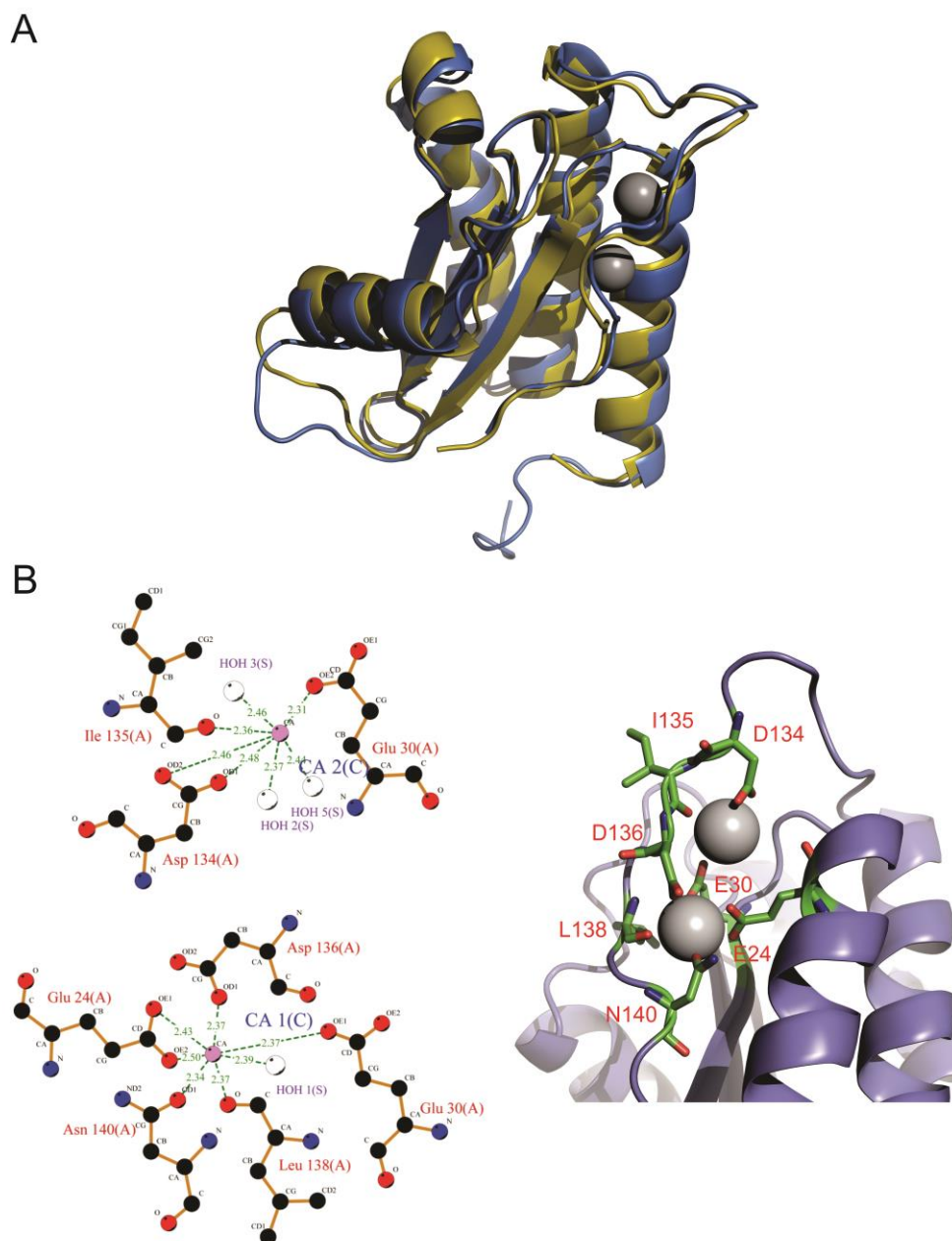


Figura 2.4. Estructura cristalina de BA42. (A) Superposición esquemática de la estructura cristalina (azul) y estructura de la solución (amarillo). Los átomos que coordinan los iones de Ca^{2+} se muestran en verde. (B) Coordinación de los dos iones de Ca^{2+} . A la izquierda, el primer ion de Ca^{2+} está coordinado por las cadenas laterales de E30 y D134, por el oxígeno del carbonilo de I135 y se encuentra interactuando con tres moléculas de agua. El segundo ion está coordinado por las cadenas laterales de E24, E30, D136, y N140, el oxígeno del carbonilo de I138 e interactúa con una molécula de agua. A la derecha, se muestran las cadenas laterales que interactúan con los dos átomos de Ca^{2+} en la estructura.

A pesar de que la solución madre usada para la cristalización no contenía cationes divalentes, una clara densidad electrónica compatible con dos cationes divalentes se pudo ver en los primeros ciclos de refinamiento en las proximidades de los residuos E24, E30, D134, I135, D136, I138, y N140. Finalmente, dos átomos de calcio fueron asignados a esta densidad electrónica después de varias observaciones: (i) los mapas de Fourier mostraron niveles de densidad electrónica compatible con átomos de calcio después del refinamiento, (ii) se observó la geometría de coordinación esperada y las distancias de unión para un ion de calcio, (iii) no se obtuvo señal en el escaneo de excitación de rayos X descartando así la presencia de metales de transición, (iv) los dos átomos de calcio y las moléculas de agua que los rodean muestran valores casi similares para sus factores B después del refinamiento.

En este contexto, la estructura cristalina confirmó que el *loop* V-D cruza sobre el *loop* I-A, pero se reveló adicionalmente una característica que no fue detectada en los experimentos de RMN en solución: la presencia de dos iones de Ca^{2+} (representados por esferas en la figura 2.4.A). El primero de los iones de Ca^{2+} está coordinado por las cadenas laterales de los residuos E30 y D134, y por el oxígeno del carbonilo del residuo I135, con tres moléculas de agua completando la esfera de coordinación. El segundo ion de Ca^{2+} interactúa con las cadenas laterales de los residuos E24, E30, D136, y N140, con el oxígeno del carbonilo de I138 y con una molécula de agua (Figura.2.4.B). Este doble sitio de unión a metales estabiliza la conformación del par de “*crossing loops*”, de forma tal que los dos Ca^{2+} unen el *loop* V-D (a través de residuos D134, I135, D136, y N140) con dos residuos localizados en los bordes del “*crossing loop*” inferior: el primer residuo del *loop* I-A (E24) y el primer residuo de la hebra- β A (E30). Además, dos puentes salinos entre R21 (α -hélice I) y D136 (*loop* V-D), y entre E13 (α -hélice I) y K144 (hebra- β D) estabilizan aún más esta particular conformación de *loops*.

La unión de los metales induce un cambio conformacional significativo en BA42

Para demostrar que la unión de los metales es crucial en la estabilización de la estructura de BA42 anclando el segmento C-terminal al núcleo $\alpha\beta\alpha$, se trató una muestra de proteína ^{15}N con un exceso de EDTA. El espectro HSQC ^1H - ^{15}N reveló un gran número de picos que cambió su posición (Figura 2.5.A), indicando un cambio estructural significativo. Con la eliminación del EDTA y la adición de un exceso de Ca^{2+} se obtuvo un espectro casi idéntico al sin el tratamiento con EDTA, demostrando que la unión del metal es responsable del cambio estructural.

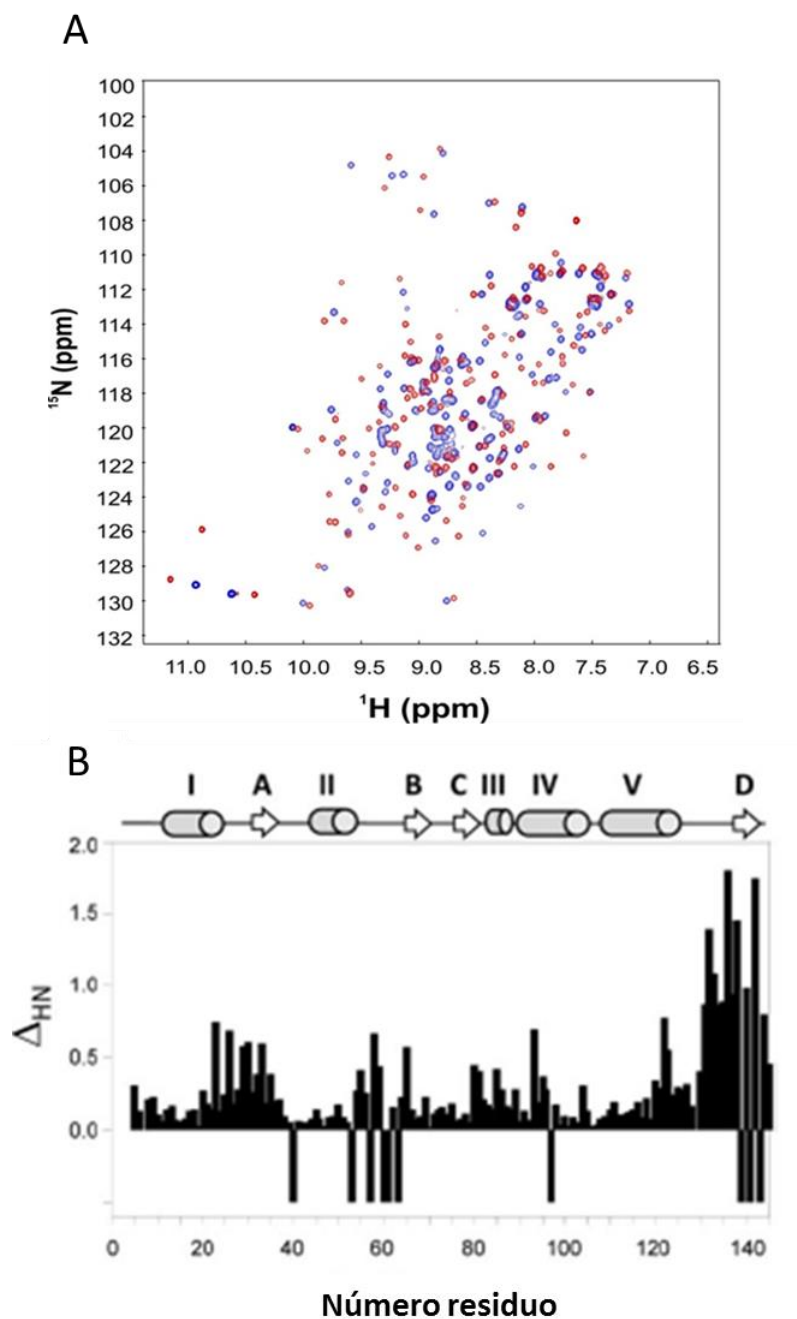


Figura 2.5. Cambios conformacionales de BA42 inducidos por la unión de calcio. (A) Espectros HSQC ^1H - ^{15}N de BA42. La proteína se incubó con un exceso de EDTA (espectro azul) y, después de la eliminación del EDTA por medio de diálisis, la proteína se trató con 3 mM CaCl_2 (espectro rojo). (B) Perturbación de los desplazamientos químicos del HN de BA42 en presencia o ausencia de Ca^{2+} . Los núcleos de amida no asignados se indican con un índice negativo. La estructura secundaria de la forma unida a metal se ilustra sobre el gráfico.

El espectro HSQC de BA42 libre de metal mostró un menor número de picos, lo que indica que algunos grupos amida están más expuestos al solvente y/o que pertenecen a regiones con menor estabilidad conformacional, lo que lleva a la obtención de señales más anchas. Paralelamente se llevaron a cabo experimentos 3D sobre una muestra de BA42 doblemente marcada, pudiendo lograrse una asignación casi completa de los núcleos de la cadena principal de BA42 en su forma Apo. Los residuos que no pudieron ser asignados (debido a que los picos no se detectaron en los diferentes experimentos) fueron E40 (*loop* A-II), N53 (último residuo de α -hélice II), D57, K60, L61, N63 (*loop* II-B), R97 (α -hélice IV), T141, y S143 (*loop* V-D). Para obtener más información, se calculó la perturbación del desplazamiento químico (CS) para el grupo HN (*“Chemical shift perturbation”*-CSP-) debido a la unión de metal, y el resultado se muestra en el gráfico de la Figura 2.5.B. Las regiones que presentaron la mayor perturbación del CS con la unión del metal fueron el *loop* V-D y la hebra- β D, seguida de dos regiones que están próximas a estas en la estructura de BA42: el *loop* I-A, la hebra- β A, y el *loop* II-B. La mayoría de los residuos cuyos picos no se detectaron pertenecen a estas regiones, con la excepción de E40 y R97. El protón amida del E40 está expuesto en la estructura de BA42 con metales, en línea con la señal de baja intensidad en el espectro HSQC ^1H - ^{15}N observada para este residuo.

A pesar del cambio general del CS de los H-N, el cálculo del índice de desplazamiento químico (CSI) utilizando los valores de CS de H^α , C^α , C^β , y C' en la forma sin metales de BA42, reveló que los residuos contienen los mismos elementos de estructura secundaria que en la forma con metales (Figura. 2.6). Todas las α -hélices y las hebras- β permanecieron inalteradas después del tratamiento con EDTA, en particular, la hebra- β A, que muestra importantes perturbaciones de CS.

Para completar la información sobre el cambio conformacional provocado por el tratamiento con EDTA, se realizó el experimento NOE H-N (Figura 2.7.A). A través del análisis de este experimento fue evidente que la región 131 a 145 es muy flexible en la proteína sin metales, ya que presenta valores de NOE H-N en el intervalo de 0,6 a -2,5. Los valores de NOE H-N inferiores a 0,7 son indicativos de flexibilidad en la escala de tiempo de ps.

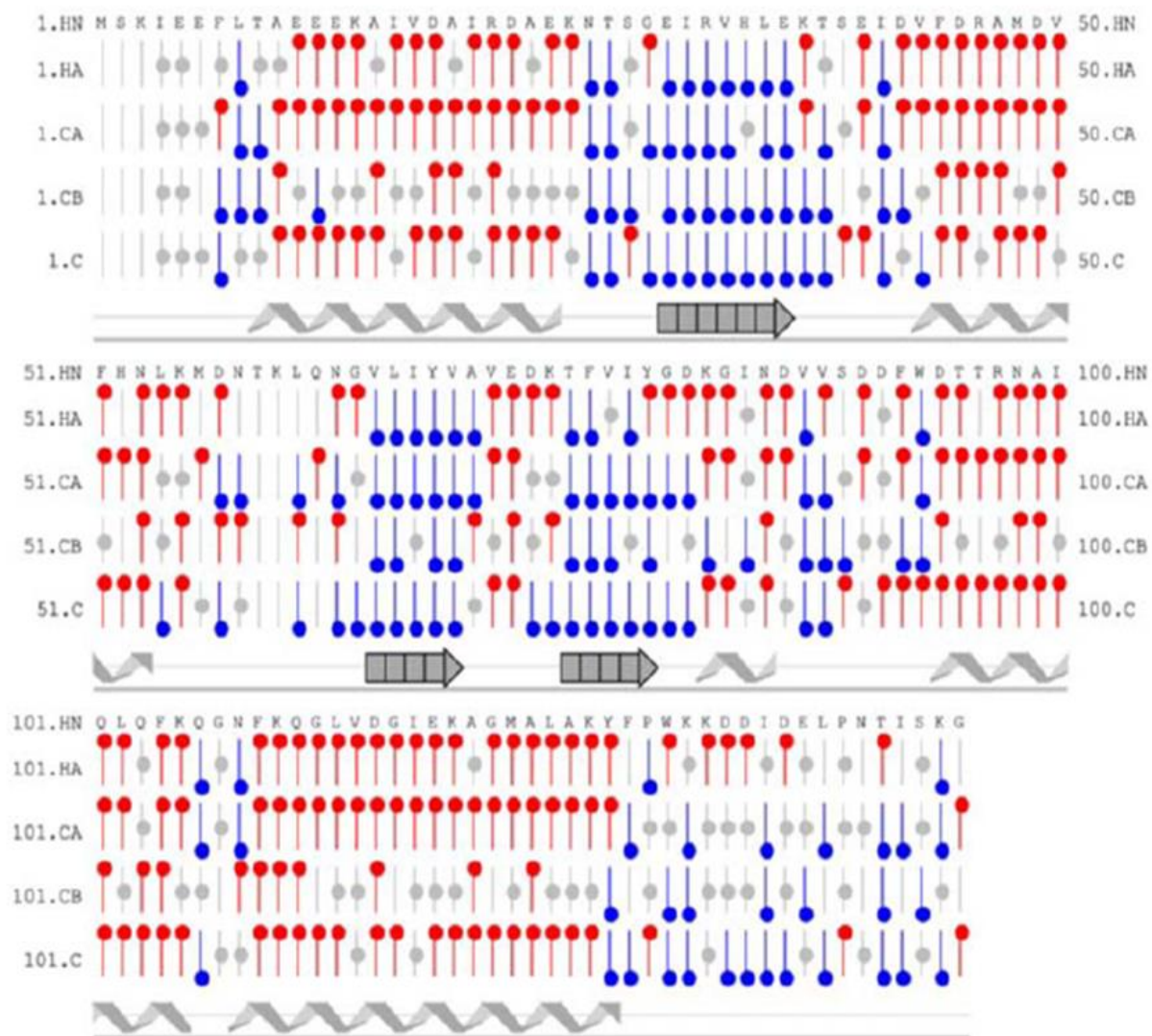


Figura 2.6. Predicción de estructura secundaria de BA42 libre de metal. La estructura secundaria se predijo para el valor de desplazamiento químico calculado para H^α , C^α , C^β , y C' . Los círculos rojos en la parte superior, señalan los desplazamientos químicos consistentes con estructuras helicoidales; los círculos azules en la parte inferior, los desplazamientos químicos consistentes con estructuras de hoja β ; y los círculos grises en el centro, los cambios intermedios.

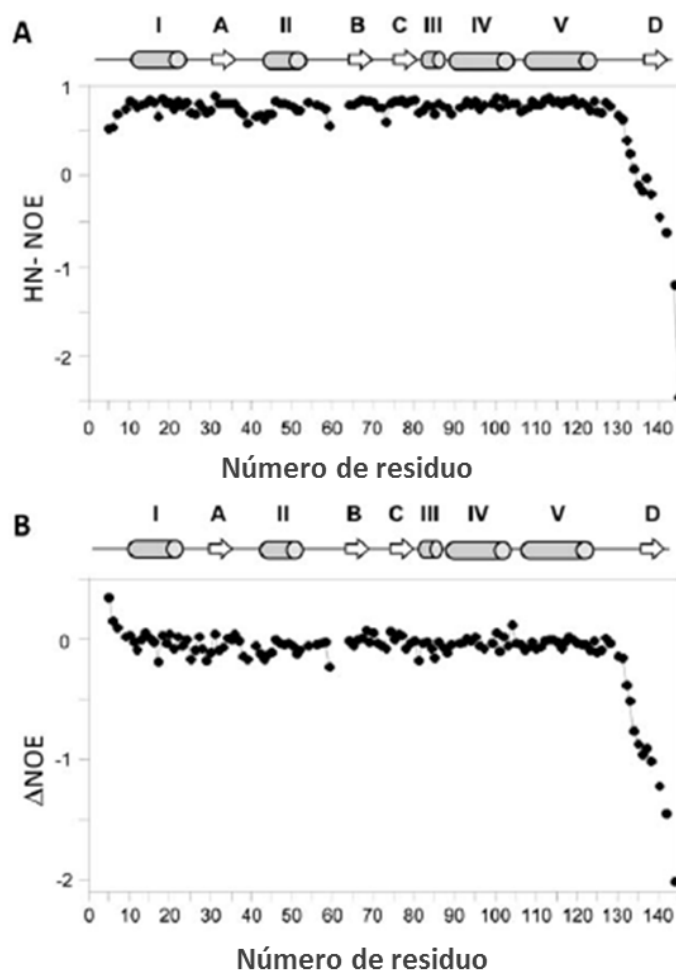


Figura 2.7. NOE ^1H ^{15}N heteronuclear de BA42. (A) Mediciones de NOEs H-N de la forma tratada con EDTA de BA42. (B) Diferencia entre NOEs H-N medidos para la forma tratada con EDTA y la forma unida a metal. La estructura secundaria de la forma unida a metal se ilustra sobre el gráfico.

La Figura 2.7.B muestra la diferencia entre los valores de NOE H-N medidos para la forma tratada con EDTA y la forma en presencia de metales (ΔNOE). Los residuos que muestran ΔNOE negativos se encuentran más rígidos en la forma con metales que en la forma Apo, y viceversa. En este sentido, se obtuvieron valores grandes y negativos de ΔNOE para el extremo C-terminal, lo que indica que el segmento 131-145 aumenta drásticamente su flexibilidad en ausencia de Ca^{2+} . Todos estos resultados tomados en conjunto indican que la eliminación de los dos iones metálicos de BA42 cambia principalmente la conformación del extremo C-terminal de la proteína, lo que conduce a que la región que comprende los residuos 131-145 se torne muy flexible, sin que esto afecte significativamente la estabilidad y la posición de las cinco α -hélices y las tres hebras- β .

BA42 presenta una constante de disociación micromolar para la unión de Ca^{2+}

Para evaluar la afinidad de unión de metal, se procedió a realizar la titulación de la forma libre de metal de BA42 mediante el agregado de Ca^{2+} . Para esto se siguió la unión mediante la inspección de las resonancias de alto campo en el rango de 0,5 a -1,5 ppm en espectros 1D ^1H . La Figura 2.8.A muestra dicha región espectral durante un experimento de titulación, evidenciándose los cambios en los desplazamientos químicos de los grupos metilo de los residuos V71, I78, V88, y I100, y un protón β del residuo M56. Todos estos residuos están resaltados en la estructura de BA42 con calcio (BA42- Ca^{2+} , Figura 2.8.B). El hecho de que los residuos que se localizan lejos del sitio de unión a los metales como V7, I78, I100, y V88 sean sensibles a la incorporación del metal refleja un ligero pero significativo cambio en la estructura terciaria de la región $\alpha\beta\alpha$. El gran efecto de apantallamiento observado en el desplazamiento químico del H^β del residuo M56 en la proteína con metales es particularmente indicativo de los cambios estructurales provocados por el Ca^{2+} . Este hecho se produce principalmente por la proximidad a la L136, creando un entorno altamente hidrófobo para la cadena lateral de M56. En la conformación libre de metales, el CS de los dos protones β de M56 se mueve a 1,84 ppm, un valor compatible con una cadena lateral expuesta al solvente.

Durante el curso del experimento de titulación, todos los picos pertenecientes a la proteína Apo disminuyeron sus intensidades a medida que la concentración de Ca^{2+} aumentó, con una aparición simultánea de los picos de la forma unida a los metales. Esto indica que la cinética de intercambio es lenta en la escala de tiempo de RMN (μs - ms). Este comportamiento se observa incluso para picos que muestran una pequeña diferencia en el desplazamiento químico entre las dos formas. Esta cinética de intercambio tan lento refleja un gran cambio conformacional asociado a la unión de los metales. A modo de evaluar la estequiometría y la afinidad de la interacción se utilizó la variación de las integrales de los picos mencionados en función de la concentración de Ca^{2+} (Figura. 2.8.C). La curva obtenida se ajustó a la ecuación de Hill variando el coeficiente n , obteniéndose el mejor ajuste para $n = 2$, lo que indica que existe una cooperación positiva de interacción entre ambos cationes. La constante de disociación media calculada para cada sitio fue de 43 μM , lo que equivale a una constante de afinidad de $2,3 \times 10^4 \text{ M}^{-1}$. Esta relativamente alta a moderada afinidad es compatible con el hecho de que en las muestras utilizadas para la cristalografía de rayos X y para RMN no se observó ninguna traza de la proteína en su forma libre de metal. Por lo tanto, se puede concluir que la proteína incorporó los dos iones Ca^{2+} del medio de cultivo, sin la necesidad de una posterior adición de metal a la proteína recombinante.

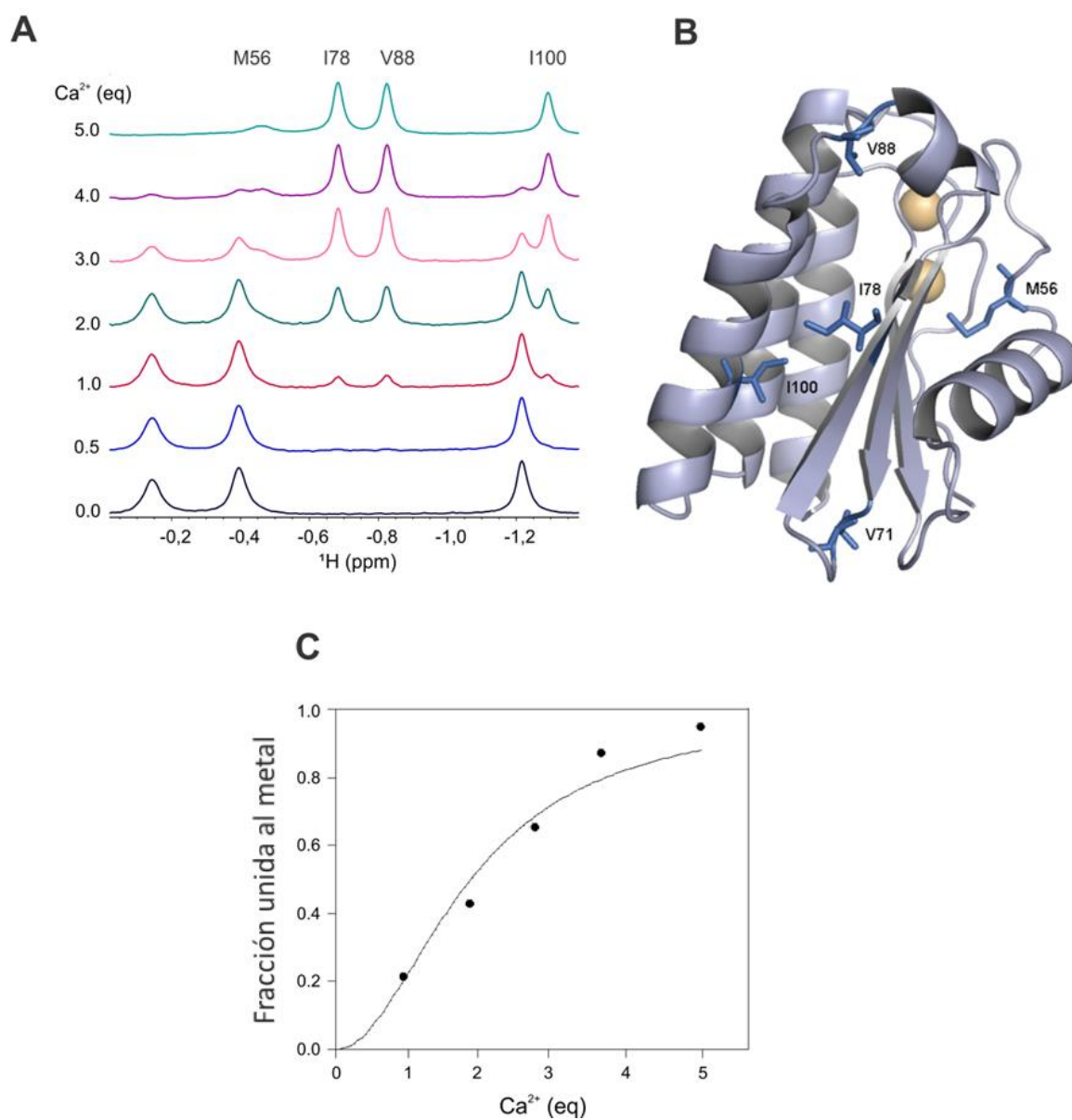


Figura 2.8. Afinidad de unión de los metales. (A) Titulación de BA42 libre de metal con Ca²⁺ monitoreada a través de espectros 1D ¹H de RMN en el intervalo de 0,5 a -1,5 ppm. Se indican los grupos metilo de V71, I78, V88, y I100, y el protón β de M56. (B) Vista en cinta de la estructura con metales señalando los residuos V71, I78, V88, I100, y M56. (C) Variación de la integral del pico en función de la concentración de Ca²⁺ en los espectros 1D ¹H NMR. La curva se ajustó utilizando la ecuación de Hill (línea continua) dando un coeficiente de n=2.

BA42 muestra una actividad hidrolasa de ésteres monofosfóricos

Estudios anteriores mostraron que los dominios TPM de Rv2345 y AtTLP18.3 son capaces de eliminar los grupos fosfato de diferentes sustratos [66, 70]. Aunque se desconoce la estructura del dominio TPM de Rv2345, una alineación de secuencias entre este dominio y el dominio TPM de AtTLP18.3 [70] mostró la conservación de los principales residuos propuestos en la unión del sustrato. Con el interés de comparar de igual forma la secuencia de BA42 y AtTLP18.3 (15% de identidad de secuencia), se realizó un alineamiento de estas secuencias, evidenciándose también en este análisis residuos conservados reportados como esenciales para la unión de sustrato en la fosfatasa ácida de *A. thaliana* (Figura 2.9.A). En la proteína AtTLP18.3 los residuos V101, D102 y K112 han demostrado interactuar con el sustrato en la estructura cristalina del complejo AtTLP18.3-Ser [66]. A partir del alineamiento, se pudo observar la conservación de los aminoácidos V17 y D18 en la secuencia primaria de BA42. Sin embargo, no se logró encontrar un residuo conservado en BA42 correspondiente a la K112 en AtTLP18.3. Con respecto a este hecho, vale la pena mencionar que K112 ha demostrado conservarse exclusivamente en plantas terrestres [66].

Estas observaciones anteriores nos llevaron a la hipótesis de que BA42 también podría tener un papel como hidrolasa de ésteres monofosfóricos. Sin embargo, a pesar de esta conservación en la secuencia, un alineamiento estructural entre BA42 y AtTLP18.3 mostró claramente que los residuos conservados se encuentran en diferentes elementos de la estructura secundaria del dominio TPM (Figura 2.9.C). Aunque los residuos de unión del sustrato en AtTLP18.3 están situados en la hebra- β I, los residuos hipotéticos de unión del sustrato conservados en BA42 se ubican en la α -hélice I. En este sentido, la disposición espacial de estos residuos en la estructura de BA42 es significativamente diferente de la configuración espacial observada en el sitio de unión del sustrato de AtTLP18.3. En consecuencia, esta característica dificulta la posibilidad de que ambos dominios puedan compartir el mismo sitio de unión del sustrato. Además, como se mencionó anteriormente, BA42 tiene una topología de plegado distintivo que introduce una notable diferencia conformacional en la región que abarca los sitios de unión del sustrato y del metal en la estructura de AtTLP18.3 (Figura 2.9.C) [63].

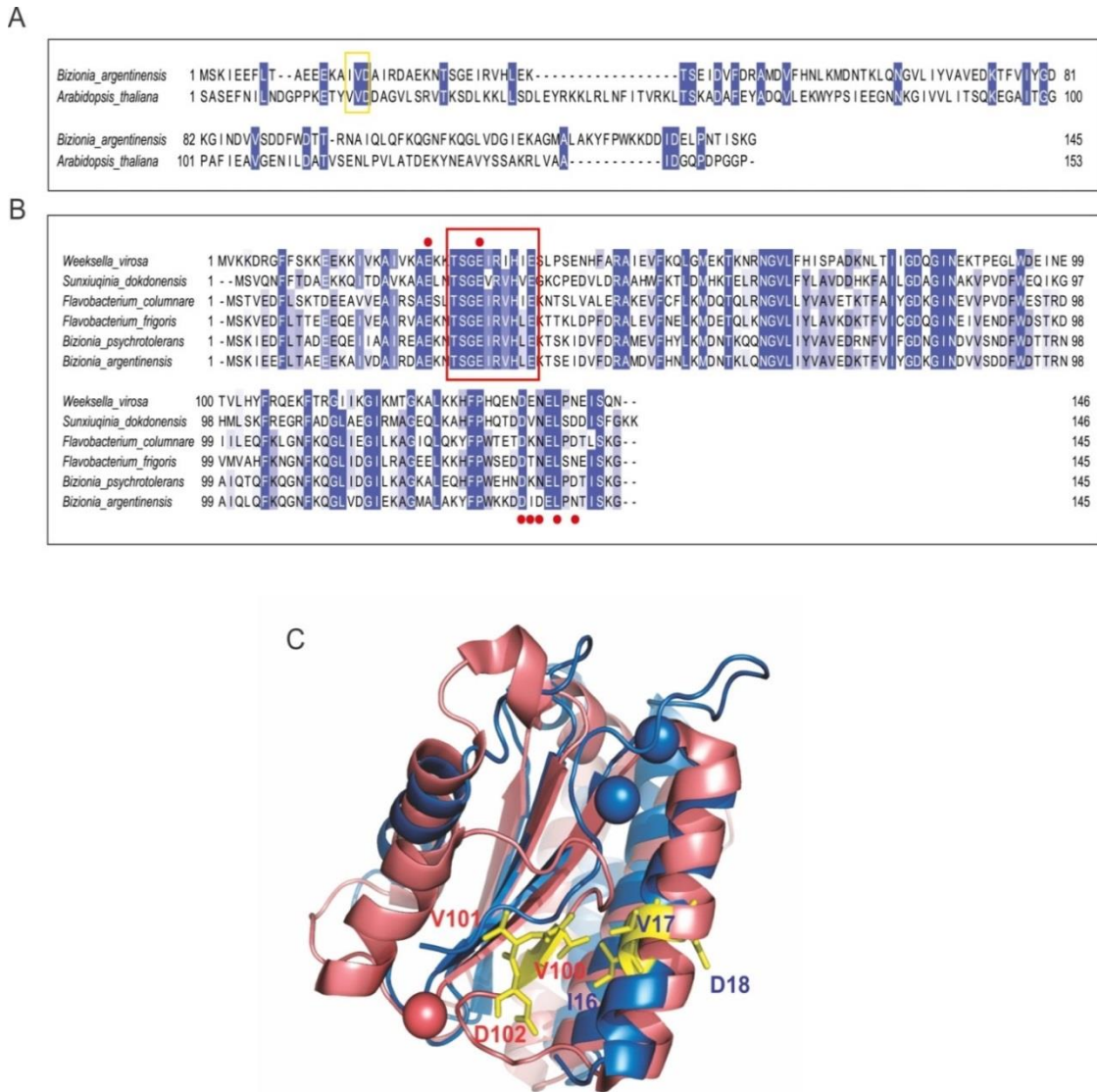


Figura 2.9. (A) Alineamiento de BA42 con la secuencia proteica del dominio TPM de AtTLP18.3 de *A. thaliana* generado por ClustalW (<http://www.ebi.ac.uk/clustalW>). El cuadro amarillo indica la conservación de los residuos V101 y D102 que están propuestos en la unión del sustrato en AtTLP18.3. Los residuos en azul oscuro están totalmente conservados. (B) Alineamiento múltiple de BA42 y cinco homólogos pertenecientes al filo Bacteroidetes. Los puntos rojos indican los residuos implicados en la coordinación de los metales en la estructura cristalina de BA42. La conservación del “crossing loop” inferior y de los residuos de la hebra-β se muestran recuadrados en rojo. Los residuos sombreados en azul se conservan estrictamente y los sombreados en azul claro se conservan parcialmente. (C) Superposición esquemática de las estructuras cristalinas de BA42 (código PDB 4OA3) (azul) y AtTLP18.3 (código PDB 3PTJ) (marrón). Las esferas sólidas representan los iones Ca^{2+} en ambas estructuras. Los residuos de unión al sustrato en AtTLP18.3 y los aminoácidos conservados en BA42 están resaltados en rojo.

Con respecto a esto, a partir de una inspección detallada de las proteínas clasificadas en la familia PF04536 se observó que la mayoría de los miembros (polipéptidos con un tamaño comprendido entre 270 y 290 aminoácidos) tienen una topología similar caracterizada por la presencia de una secuencia señal N-terminal seguida por el dominio TPM, un dominio transmembrana, y una región C-terminal de baja complejidad de longitud variable. Este grupo incluye a AtTLP18.3, y a otras proteínas que contienen dominos TPM. En contraste, como se mencionó anteriormente, BA42 es un dominio TPM independiente con una topología distintiva y con una predicción de localización citoplasmática. Respecto a estas observaciones, a partir de una búsqueda de similitud de secuencias usando el programa Blastp se detectó que las proteínas homólogas a BA42 (es decir, dominos TPM independientes) se conservaban casi exclusivamente dentro del filo Bacteroidetes (Figura 2.9.B). Interesantemente, este alineamiento de secuencias no mostró una estricta conservación de la tríada VDK correspondiente al sitio de unión del sustrato en AtTLP18.3. Sin embargo, se observó una conservación notable para los residuos implicados en el sitio de unión de metal, el *crossing loop* inferior y la hebra β -I (Figura 2.9.B).

A pesar de que BA42 puede carecer de un sitio activo homólogo a AtTLP18.3, no se podía descartar la hipótesis de que BA42 pudiese actuar como una hidrolasa fosfórica. En primer lugar, de igual forma a AtTLP18.3, BA42 contiene metales en su estructura, siendo este un rasgo característico de muchas hidrolasas fosfóricas de monoésteres [76, 77]. En segundo lugar, BA42 pertenece a un organismo psicrófilo y, por lo tanto, puede haber evolucionado un sitio activo alternativo con dinámica y características estructurales adecuadas para la catálisis en frío. Por último, a través de una comparación estructural con los programas DALI [78] (Tabla 2.1) y MATRAS [79] (Tabla 2.2), se encontró que BA42 poseía alta similitud estructural con hidrolasas, como ha sido descrito anteriormente para AtTLP18.3 [66].

Nº	Entry	Z	Rmsd	Lali	Nres	SqID	Description
1	4oa3-A	32,2	0	141	141	100	PROTEIN BA42
2	2lt2-A	23,5	1,2	141	145	100	PUTATIVE UNCHARACTERIZED PROTEIN
3	2mpb-A	23,3	1,2	141	145	100	BA42
4	2kw7-A	12	2,9	124	157	15	CONSERVED DOMAIN PROTEIN
5	2kpt-A	11,6	2,7	114	148	15	PUTATIVE SECRETED PROTEIN
6	3ptj-A	11,1	2,6	117	153	11	UPF0603 PROTEIN AT1G54780, CHLOROPLASTIC
7	3pw9-A	11	2,7	117	154	11	UPF0603 PROTEIN AT1G54780, CHLOROPLASTIC
8	3pvh-A	11	2,7	117	153	11	UPF0603 PROTEIN AT1G54780, CHLOROPLASTIC
9	3g98-A	6,4	2,7	85	111	15	ALANYL-TRNA SYNTHETASE
10	3g98-B	6,3	2,7	85	111	15	ALANYL-TRNA SYNTHETASE
11	1u2x-A	5,4	2,9	98	450	7	ADP-SPECIFIC PHOSPHOFRUCTOKINASE
12	3drw-B	4,6	2,9	98	450	10	ADP-SPECIFIC PHOSPHOFRUCTOKINASE
13	4op0-A	4,6	3,4	104	259	1	BIRA BIFUNCTIONAL PROTEIN
14	2cgh-B	4,6	3,5	103	242	2	BIOTIN LIGASE
15	3drw-A	4,6	2,8	96	460	9	ADP-SPECIFIC PHOSPHOFRUCTOKINASE
16	3l2z-B	4,6	3,4	101	243	3	BIRA BIFUNCTIONAL PROTEIN
17	3l2z-A	4,5	3,3	99	240	3	BIRA BIFUNCTIONAL PROTEIN
18	3zef-B	4,5	3,3	86	1398	15	PRE-MRNA-SPLICING FACTOR 8
19	4i43-B	4,5	3,4	87	1407	15	A1 CISTRON-SPLICING FACTOR AAR2
20	2zxo-A	4,4	3,3	79	648	9	SINGLE-STRANDED DNA SPECIFIC EXONUCLEASE RECJ
21	3jcm-A	4,4	3,4	86	2174	15	PRE-MRNA-SPLICING FACTOR 8
22	3jb9-A	4,4	3,5	86	1964	15	PRE-MRNA-SPLICING FACTOR SPP42
23	2dzc-A	4,4	2,9	88	222	9	BIOTIN--[ACETYL-COA-CARBOXYLASE] LIGASE
24	2cgh-A	4,3	3,3	101	238	1	BIOTIN LIGASE
25	4xu0-A	4,3	3,4	104	265	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
26	4xy-A	4,3	3,4	104	264	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
27	4xtu-A	4,3	3,4	104	266	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
28	4xtv-A	4,3	3,4	104	264	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
29	4xu1-A	4,3	3,3	102	263	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
30	3ruX-B	4,3	3,4	103	261	1	BIRA BIFUNCTIONAL PROTEIN
31	4xtv-B	4,3	3,4	103	262	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
32	4xu0-B	4,3	3,4	103	262	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
33	4xu2-B	4,3	3,4	102	258	1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
34	1ir6-A	4,3	3,1	78	385	10	EXONUCLEASE RECJ

Tabla 2.1. Comparación estructural de BA42 contra todas las estructuras depositadas en el banco de datos de proteínas (PDB) utilizando el servidor DALI (http://ekhidna.biocenter.helsinki.fi/dali_server). Las proteínas que resultaron tener homología estructural con BA42 se enumeran en orden decreciente de similitud. Entry: código PDB; Z: puntuación de Dali Z; RMSD: desviación de la media cuadrática, Lali: número de residuos alineados; Nres: número total de residuos; SqID: identidad de secuencia.

N°	Entry	Start	End	Rdis	Zsc	SqID	Description
1	4oa3A	5	145	79,2	90,76	100	PROTEIN BA42
2	2kw7A	26	156	40,3	44,39	15,5	CONSERVED DOMAIN PROTEIN
3	2kptA	62	177	38	41,02	16,5	PUTATIVE SECRETED PROTEIN
4	3pvhA	105	230	37,7	40,68	11,5	UPF0603 PROTEIN AT1G54780, CHLOROPLASTIC
5	3g98A	757	867	23	15,65	12,5	ALANYL-TRNA SYNTHETASE
6	2ej9A	1	174	9	13,45	10,9	PUTATIVE BIOTIN LIGASE
7	3jb9A	1679	1765	0,4	12,99	14,6	PRE-MRNA-SPLICING FACTOR SPP42
8	4i43B	1727	1813	0,6	12,85	13,4	PRE-MRNA-SPLICING FACTOR 8
9	1ir6A	327	427	3,8	12,49	9,3	EXONUCLEASE RECJ
10	2zvfa	801	906	14,7	12,31	13,5	ALANYL-TRNA SYNTHETASE
11	2dxuA	2	176	9,7	12,31	5	BIOTIN--[ACETYL-COA-CARBOXYLASE] LIGASE
12	2qb7A	270	397	4	12,23	15,6	EXOPOLYPHOSPHATASE
13	5cetA	234	340	5,4	11,83	7,1	BIFUNCTIONAL OLIGORIBONUCLEASE AND PAP PHOSPHATASE
14	2a9sA	2	167	15,1	11,4	7,4	COMPETENCE/DAMAGE-INDUCIBLE PROTEIN CINA
15	4m62S	7	161	15	11,31	11,1	T117
16	3oxpA	0	147	15	11,22	5,3	PHOSPHOTRANSFERASE ENZYME II, A COMPONENT
17	3devA	209	315	5,6	11,09	3,1	SH1221
18	4xtvA	4	206	8,9	10,69	5,1	BIFUNCTIONAL LIGASE/REPRESSOR BIRA
19	2gk4A	149	228	6,7	10,26	10,8	CONSERVED HYPOTHETICAL PROTEIN
20	1a3aA	4	148	14,2	10,07	9,7	MANNITOL-SPECIFIC EII
21	5ccfA	317	413	2,2	9,97	6,8	ADP-DEPENDENT GLUCOKINASE
22	3dmaA	232	343	4,8	9,79	7	EXOPOLYPHOSPHATASE-RELATED PROTEIN
23	5intA	317	399	7,5	9,16	15,3	PHOSPHOPANTOTHENOYL-CYSTEINE DECARBOXYLASE
24	4phzB	34	174	7,6	9,1	5,3	PARTICULATE METHANE MONOOXYGENASE SUBUNIT A
25	2eayA	1	175	8,4	8,89	10,6	BIOTIN [ACETYL-COA-CARBOXYLASE] LIGASE
26	2debA	317	397	1,2	8,74	11,5	CARNITINE O-PALMITOYLTRANSFERASE II,
27	3urrA	0	151	13,4	8,52	8,2	PTS IIA-LIKE NITROGEN-REGULATORY PROTEIN PTSN
28	1tluA	272	356	1,3	8,3	5,3	CHOLINE O-ACETYLTRANSFERASE
29	1ns5A	10	151	15,4	8,29	4,5	HYPOTHETICAL PROTEIN YBEA
30	3bfmA	9	176	7,4	8,14	7,5	BIOTIN PROTEIN LIGASE-LIKE PROTEIN OF UNKNOWN
31	4hvmA	78	160	2,7	8,13	7,6	TLMII
32	2dawA	71	144	8,7	8,11	1,6	RWD DOMAIN CONTAINING PROTEIN 2
33	4kavA	422	524	3,4	7,99	13,9	YHBX/YHJW/YIJP/YJDB FAMILY PROTEIN
34	3bs4A	48	147	6,7	7,86	3,2	UNCHARACTERIZED PROTEIN PH0321

Tabla 2.2. Comparación estructural de BA42 contra todas las estructuras en el banco de datos de proteínas (PDB) utilizando el servidor MATRAS (http://strcomp.protein.osaka-u.ac.jp/matras/cgi-bin/matras_libform.cgi). Las proteínas con homología estructural con BA42 se enumeran en orden decreciente de similitud. Entry: código PDB, Start: número del primer residuo alineado en la secuencia; End: número de secuencia del último residuo alineado; Rdis: Puntuación Sdis normalizada (%); Zsc: puntuación Z MATRAS; SqID: identidad de secuencia.

En este contexto, se procedió a ensayar la actividad hidrolasa de ésteres monofosfóricos (de ahora en adelante denominada fosfatasa) de BA42 mediante el uso dos sustratos generales, p-nitrofenil fosfato (pNPP) y fosfoserina (pSer). Los resultados iniciales indicaron que BA42 tiene actividad fosfatasa [80] (Figura 2.10.A). En estos ensayos se utilizó la proteasa TEV como control [81], una proteína que carece de actividad fosfatasa. BA42 presentó actividad enzimática con ambos sustratos, siendo más eficiente para hidrolizar pSer. Curiosamente, la actividad enzimática de BA42 se inhibió dramáticamente después del pretratamiento con EDTA. Por consiguiente, para caracterizar el efecto de los metales bivalentes sobre la actividad fosfatasa se realizó el ensayo con BA42 libre de metales en presencia de varios iones divalentes (Mg^{2+} , Ca^{2+} , Mn^{2+} y Zn^{2+}) usando pSer como sustrato (Figura 2.10.B). Cuando se incubó con Mg^{2+} la actividad de BA42 se incrementó significativamente (hasta cinco veces) en comparación con los otros iones metálicos divalentes ensayados. Posteriormente, se determinaron las condiciones óptimas enzimáticas de BA42 unida a Mg^{2+} (BA42- Mg^{2+}) para el sustrato pSer. La mayor actividad se observó a pH 7,0 y alrededor de 20° C (Figura 2.11.A y 2.11.B). Como se observa en las enzimas psicrófilas [38], la proteína conservó casi el 70% de la actividad máxima a 4 °C [80].

El estudio cinético de BA42 se realizó con pSer, tanto para BA42- Mg^{2+} como para la BA42- Ca^{2+} . Las curvas de actividad en función del sustrato mostraron un mejor ajuste a una sigmoidea alostérica (Figura 2.11.C). Las constantes de Michaelis-Menten aparentes ($K_{M\ ap}$), las velocidades catalíticas (V_{max}) y las constantes catalíticas (k_{cat}) fueron calculadas en la condición de 20° C y pH 7,0 (Tabla 2.3).

	$K_{M\ ap}$ (mM)	V_{max} (nmol min ⁻¹ mg ⁻¹)	k_{cat} (min ⁻¹)
BA42-Ca^{+2}	5,63 ± 1,82	0,25 ± 0,02	0,0038 ± 0,0002
BA42-Mg^{2+}	43,36 ± 3,29	3,95 ± 1,26	0,067 ± 0,002

Tabla 2.3. Parámetros cinéticos calculados para BA42- Mg^{2+} y BA42- Ca^{2+} utilizando pSer como sustrato a 20° C y pH 7,0.

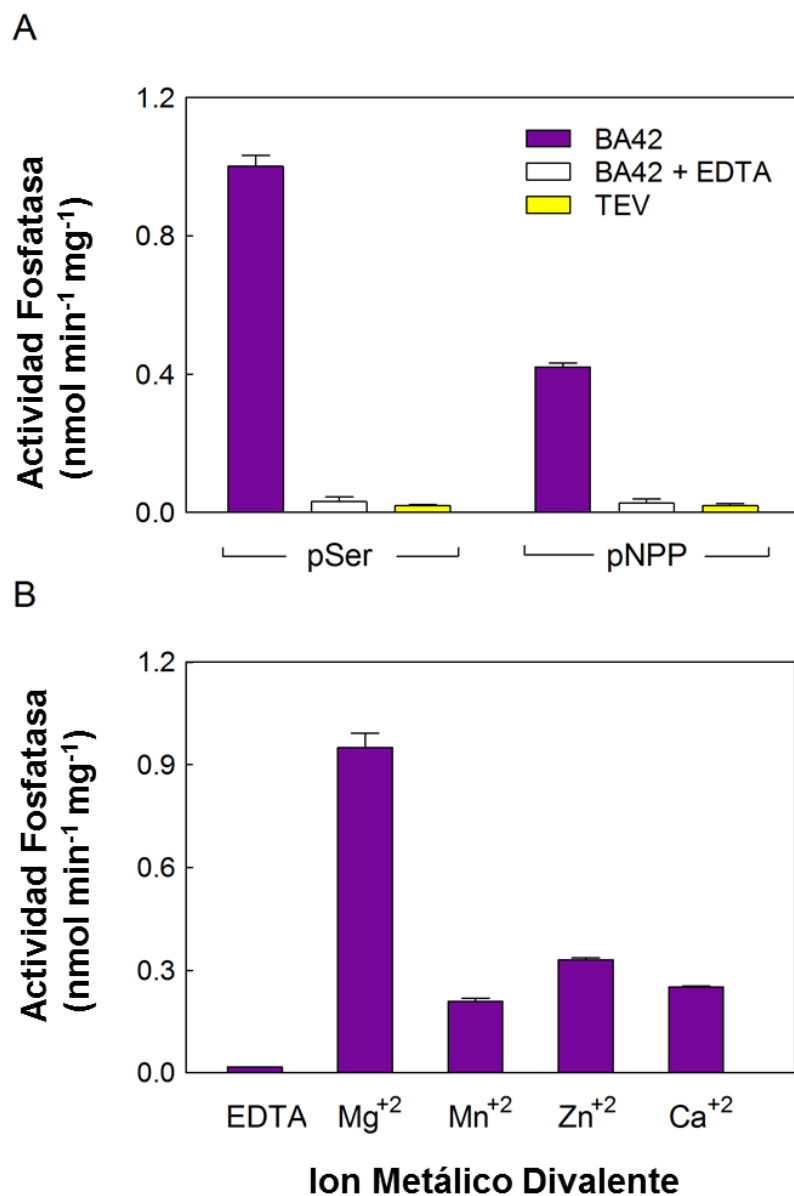


Figura 2.10. Actividad fosfatasa de BA42. (A) La actividad se ensayó con los sustratos generales p-nitrofenil fosfato (pNPP) y fosfoserina (pSer) incubando muestras de BA42 en Tris-HCl 20 mM (pH 7,5) a 25 °C en presencia o ausencia de EDTA. La proteasa TEV se utilizó como control negativo. (B) Efecto de los iones metálicos divalentes sobre la actividad fosfatasa. BA42 se trató con EDTA, y posteriormente fue sometida a una diálisis extensiva para eliminar el EDTA. Finalmente se ensayó su actividad fosfatasa en HEPES 25 mM (pH 7,0) a 22 °C y en presencia de varios metales divalentes.

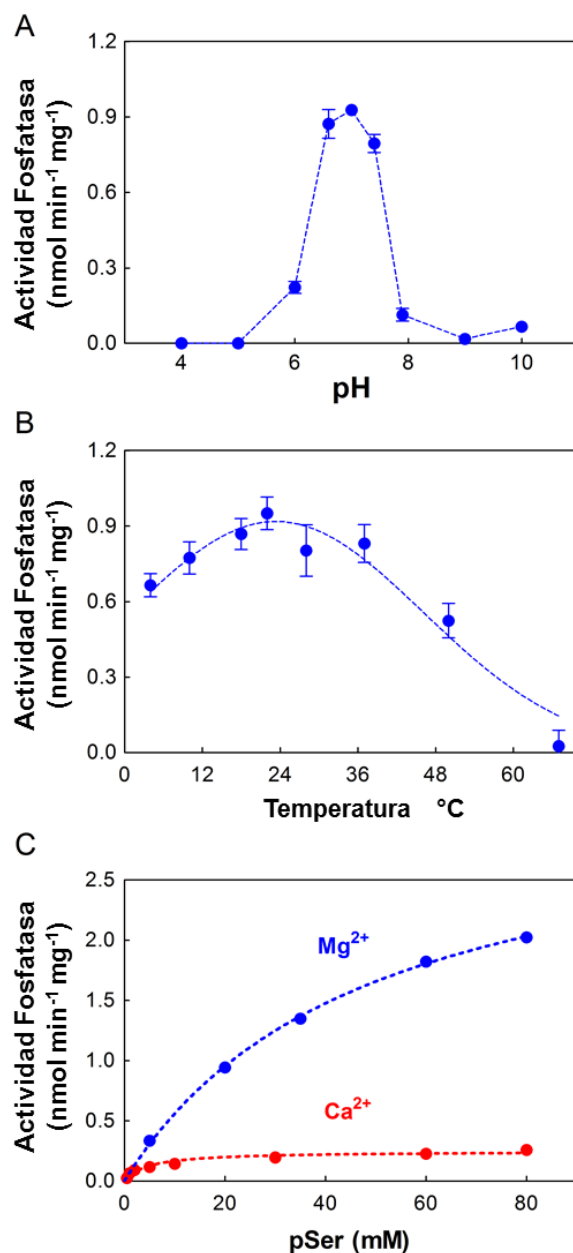


Figura 2.11. Caracterización bioquímica de la actividad fosfatasa de BA42. BA42-Mg²⁺ se incubó bajo diferentes condiciones con pSer y el fosfato liberado se cuantificó bioquímicamente. **(A)** Influencia del pH sobre la actividad fosfatasa. La actividad fosfatasa se ensayó a diferentes pH a 22 °C. La actividad máxima se presentó alrededor del pH 7. **(B)** Efecto de la temperatura en la actividad enzimática. La actividad fosfatasa se midió a diferentes temperaturas en HEPES 25 mM (pH 7,0). La actividad máxima se observó a 20 °C. **(C)** Cinética enzimática bajo condiciones catalíticas óptimas. La actividad fosfatasa de BA42-Mg²⁺ y BA42-Ca²⁺ se midió a concentraciones crecientes de sustrato a pH 7 y 20 °C. Todos los experimentos se realizaron por triplicado y los resultados se expresan como media ± EEM.

La unión de Mg^{2+} a BA42 se asocia con un aumento en la actividad enzimática

Cómo se describió previamente, es posible seguir los cambios conformacionales asociados a la unión de Ca^{2+} de la proteína BA42 en su estado libre de metales (BA42-Apo). Por consiguiente, se monitorearon los cambios conformacionales de BA42-Apo al unirse a otros metales divalentes (Figura 2.12.A). Todos los metales probados fueron capaces de unirse a BA42 y mostraron un comportamiento similar al observado previamente para Ca^{2+} . Por lo tanto, esto podría ser una evidencia de que la unión de los metales divalentes ensayados produciría un cambio conformacional global similar al observado para el Ca^{2+} . Sorprendentemente, a diferencia de los otros cationes ensayados, el Mg^{2+} presentó un comportamiento diferencial. La constante de disociación para este ion divalente fue de 87 μM mientras que los otros metales mostraron constantes de disociación de 48 μM en promedio. Pero lo más importante, la actividad máxima en presencia de Mg^{2+} fue en promedio cinco veces mayor que con Ca^{2+} , Mn^{2+} o Zn^{2+} (Figura 2.12.B). El comportamiento observado cuando la enzima se tituló con Mg^{2+} se apreció mejor mediante el trazado de la actividad y el cambio conformacional en el mismo gráfico (Figura 2.12.C). De esta forma, es evidente que la unión de Mg^{2+} , seguida por el cambio conformacional de BA42-Apo a BA42- Mg^{2+} , se asocia con un aumento en la actividad enzimática. Es interesante observar que aunque todos los metales divalentes probados fueron capaces de inducir el cambio conformacional de BA42-Apo, el Mg^{2+} fue el único capaz de aumentar la actividad enzimática. Esto sugiere una especificidad estricta en el tipo de metal en el mecanismo catalítico [80].

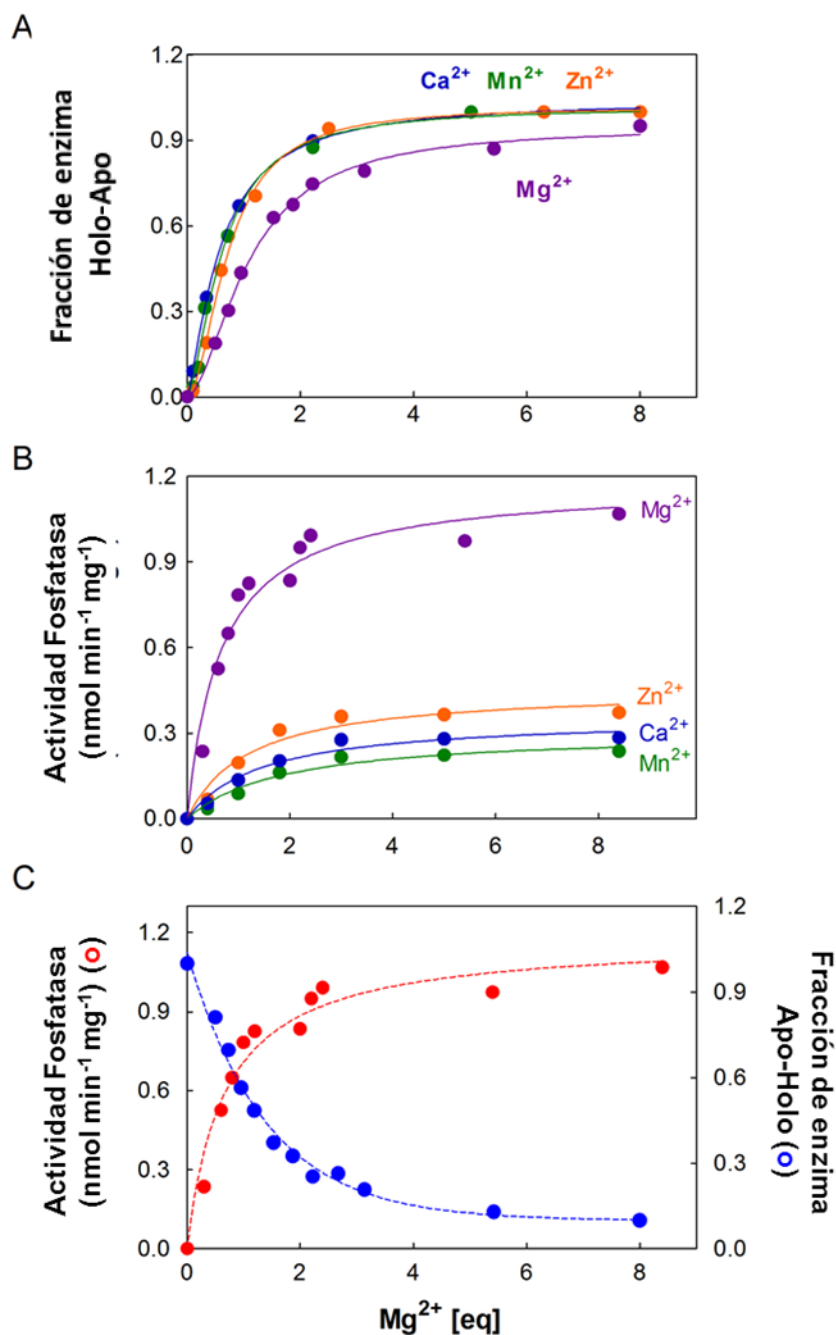


Figura 2.12. Relación entre la actividad enzimática y el cambio conformacional de BA42 inducido por la unión de metales divalentes. BA42-Apo se incubó en soluciones que contenían Tris-HCl 20 mM (pH 7), pSer 20 mM y concentraciones variables de diferentes metales divalentes. **(A)** Cambio conformacional seguido por los espectros 1D-¹H de BA42-Apo tras el agregado de diferentes Me²⁺. **(B)** Actividad fosfatasa de BA42-Holo a diferentes concentraciones de metales divalentes. **(C)** Actividad fosfatasa y cambio conformacional de BA42-Apo en función de la concentración de Mg²⁺.

BA42- Mg²⁺ presenta cambios conformacionales locales con respecto a BA42- Ca²⁺

Se demostró previamente que la unión de Ca²⁺ es crucial para la estabilización de la estructura de BA42. Dicho cambio conformacional implica el cierre del extremo C-terminal anclándose a la estructura $\alpha\beta\alpha$, generando así una conformación rígida. Para investigar si el Mg²⁺ induce el mismo cambio conformacional, se analizaron los espectros HSQC ¹H-¹⁵N de BA42 unido a Ca²⁺ y Mg²⁺. Curiosamente, aunque ambos espectros mostraron un patrón similar de picos con pequeñas variaciones, no se pudo detectar un grupo de picos en el espectro HSQC de BA42-Mg²⁺ (Figura 2.13).

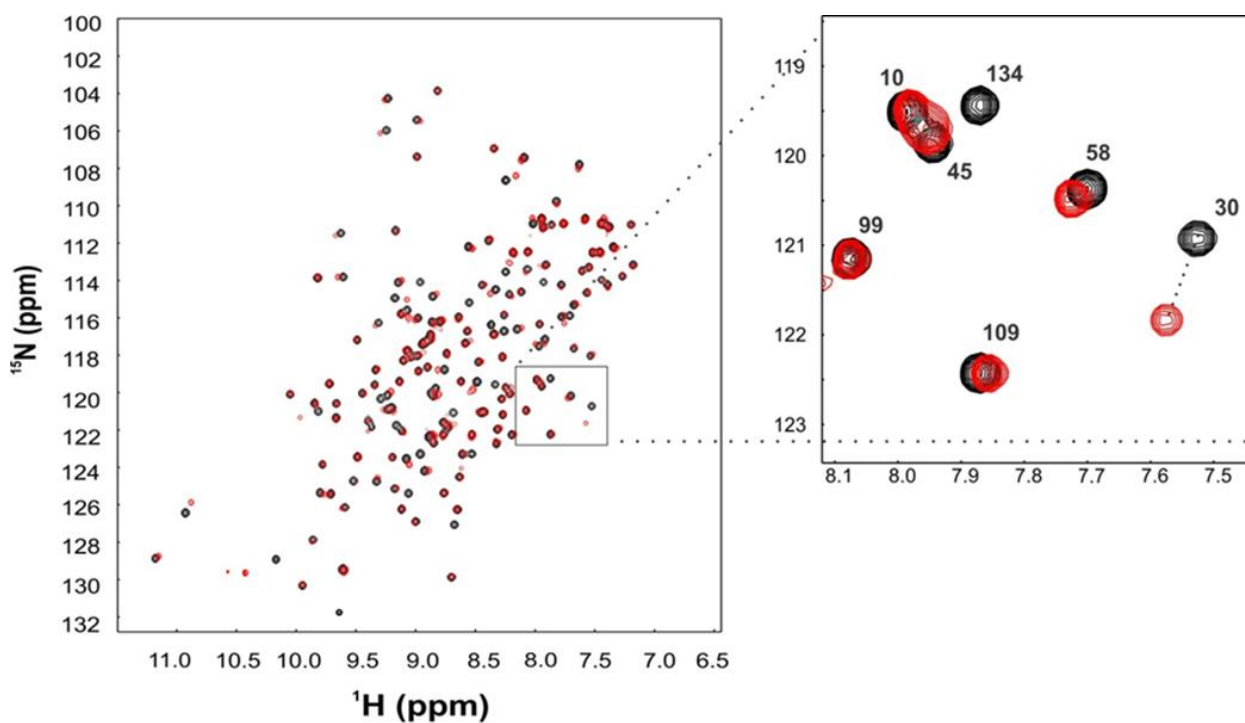


Figura 2.13. Cambios en el espectro HSQC ¹H-¹⁵N de BA42 inducidos por la unión de Mg²⁺ y Ca²⁺. Se muestra los espectros HSQC ¹H-¹⁵N de BA42-Mg²⁺ (rojo) y BA42-Ca²⁺ (negro) superpuestos. En el recuadro ampliado, se ejemplifica la variación de la señal de amida para los residuos seleccionados. El pico del residuo 134 no está presente en el espectro de BA42 unido a Mg²⁺ y hay un cambio significativo del desplazamiento químico de la señal de amida del residuo 30 entre los dos espectros.

Un análisis posterior mostró que los picos no observados correspondían a las señales de las amidas de los residuos 21, 26, 28, 31, adyacentes al sitio de unión de metal, y los residuos 131-137, que comprenden el “*crossing loop*” superior en la BA42-Ca²⁺. La desaparición de estas señales podría atribuirse a *i*) una mayor flexibilidad asociada a movimientos conformacionales lentos o intermedios, *ii*) una mayor exposición al solvente o *iii*) a cambios significativos en el entorno electrónico induciendo grandes diferencias de CS. Después de una inspección minuciosa del espectro HSQC de BA42-Mg²⁺, no se pudieron asignar otros picos adicionales a estos residuos. Además, no pudimos detectar la aparición de estas señales de amida en espectros de HSQC realizados a pHs más bajos (hasta el límite de precipitación de la proteína), descartando así la desaparición de las señales debido a un mayor intercambio con el solvente. Por lo tanto, nuestra hipótesis es que la unión del Mg²⁺ a BA42 desestabilizaría la conformación del par de “*crossing loops*”.

A fin de analizar más detalladamente las diferencias entre los espectros HSQCs de BA42-Mg²⁺ y BA42-Ca²⁺, se realizó un estudio de CSP (Figura 2.14.A). Como resultado se encontraron cambios importantes en las señales de amida de los residuos que interactúan directamente con los metales y otros cercanos al sitio de unión de estos, específicamente en el “*crossing loop*” inferior y el “*loop*” que conecta la hebra-β B con la α-hélice II (Figura 2.14.D y 2.14.E). Los residuos 22, 24, 25, 27, 30, 59, 60, 61 y 140 mostraron valores de CSP por encima de la línea umbral.

En conjunto, estos resultados indicaron claramente que la unión de Mg²⁺ a BA42 (en lugar de Ca²⁺) indujo cambios locales en el sitio de unión de metal y en residuos pertenecientes a los “*crossing loops*” [80]. A pesar de que la estructura global de BA42-Mg²⁺ parecía ser similar a la de BA42-Ca²⁺, el aumento de la flexibilidad de los “*crossing loops*” inducido por Mg²⁺ aparece como una característica importante que probablemente pueda explicar el aumento de la actividad fosfatasa.

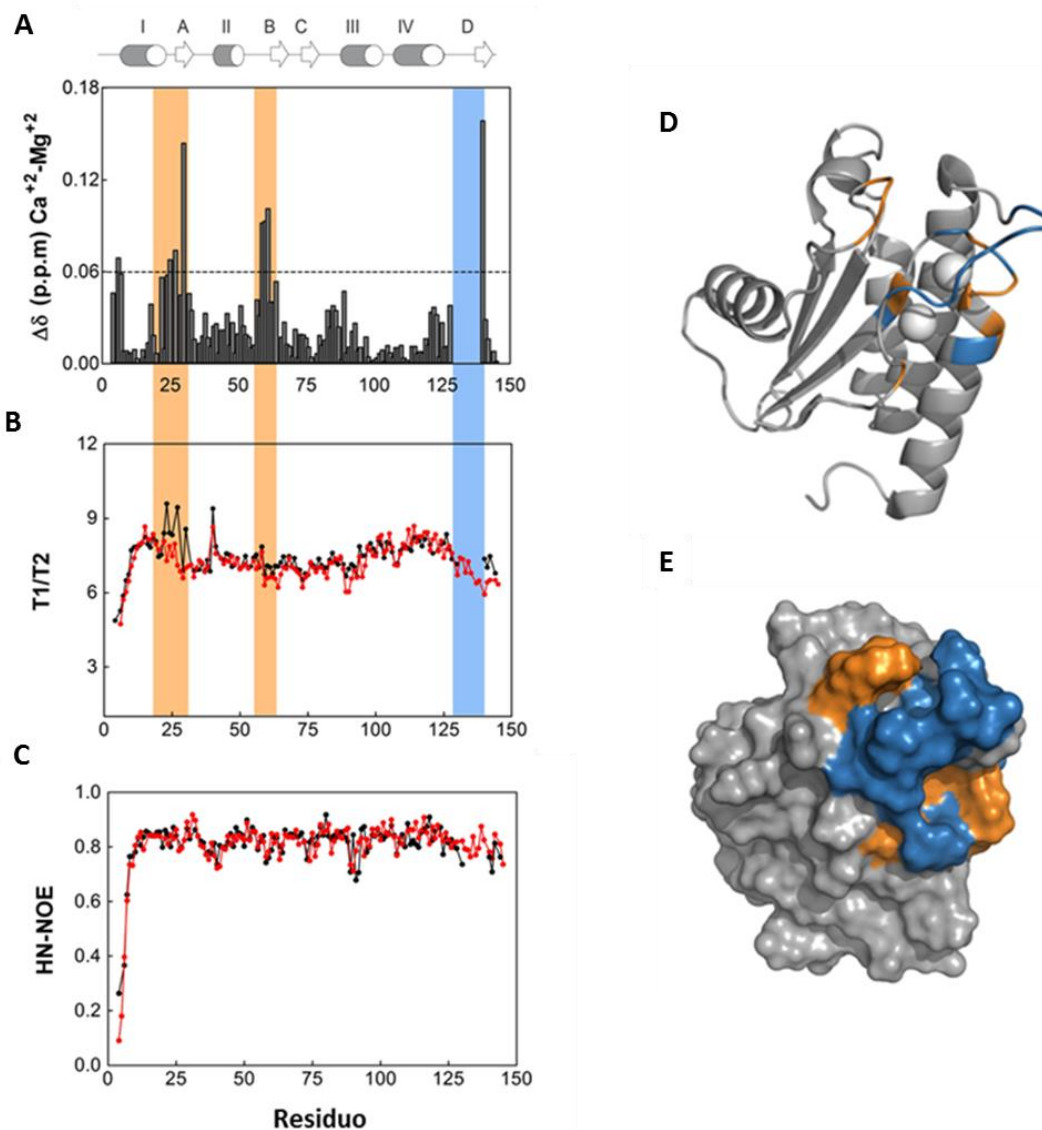


Figura. 2.14. Perturbación del desplazamiento químico del HN y dinámica de BA42-Mg²⁺ y BA42-Ca²⁺.

(A) Perturbación del desplazamiento químico del HN. Sobre el gráfico, se detalla la estructura secundaria de BA42; las α -hélices se representan por flechas blancas y las hebras β por cilindros grises. El valor umbral se estableció en 0,06. (B) Comparación de las relaciones T_1/T_2 obtenidas para BA42-Mg²⁺ (negro) y BA42-Ca²⁺ (rojo). Las barras asocian los residuos entre los gráficos A, B y la estructura secundaria de BA42. Las barras de color naranja claro resaltan los residuos que presentaron valores de perturbación de CS más altos. La barra de color azul claro resalta los residuos a los que no se les encontró señal de amida en el espectro HSQC ¹H-¹⁵N de BA42-Mg²⁺. (C) NOE ¹H-¹⁵N heteronuclear de BA42. Los NOE₃ H-N se midieron para BA42-Mg²⁺ (negro) y BA42-Ca²⁺ (rojo). (D) Representación en forma de cinta de la estructura de BA42 unida a los metales resaltándose los residuos con valores de perturbación de desplazamiento químico más altos (en naranja) y los residuos que faltan en el espectro HSQC ¹H-¹⁵N de BA42-Mg²⁺ (en azul). (E) Representación de la superficie de la estructura mostrada en el cuadro D.

BA42-Mg²⁺ presenta un incremento de movimientos internos lentos en el “crossing loop” inferior y la región C-terminal con respecto a BA42-Ca²⁺

Para estudiar el efecto del Mg²⁺ sobre la dinámica global de BA42 a nivel de los residuos individuales, se realizaron mediciones de relajación del ¹⁵N en la proteína BA42 unida a Ca²⁺ y Mg²⁺. Como se mostró anteriormente [63], BA42-Ca²⁺ exhibe valores relativamente uniformes de velocidades de relajación a lo largo de la secuencia proteica, siendo esto un signo distintivo de una proteína compacta, globular y monomérica. En el caso de BA42-Mg²⁺ los valores de T_1/T_2 para los diferentes residuos resultaron muy similares a los exhibidos por BA42-Ca²⁺, mostrando una similitud también en los efectos de “*tumbling*” anisotrópico como resultado de la forma de la molécula [61] (Figura 2.14.B). Sin embargo, algunos residuos en BA42-Mg²⁺ presentaron valores de T_1/T_2 relativamente altos y valores de T_2 más bajos. Particularmente los residuos que presentaron dichas características fueron los pertenecientes a la región D22-E30, que comprende el *crossing loop* inferior, y los residuos N140-K144, perteneciente a la región C-terminal, sugiriendo la presencia de movimientos de baja frecuencia (μ s-ms), escalas de tiempo a menudo asociadas con el intercambio conformacional. Además, el *loop* que conecta la hebra- β B con la α -hélice II, la cual mostró diferencias significativas en el experimento de CSP, presentó cambios menores en la dinámica entre BA42-Mg²⁺ y BA42-Ca²⁺.

Por otro lado, los valores obtenidos en los experimentos de NOE H-N obtenidos para el BA42-Mg²⁺ fueron casi idénticos a los obtenidos para BA42-Ca²⁺, indicando que no aparecieron nuevos movimientos internos rápidos (ps-ns) en la proteína unida a Mg²⁺ (Figura 2.14.C). Por lo tanto, en concordancia con nuestros resultados anteriores, los experimentos de dinámica del esqueleto proteico claramente muestran que BA42-Mg²⁺, cuando se compara con BA42-Ca²⁺, presenta mayor flexibilidad (en la escala de tiempos lentos, μ s-ms) en la región que comprende el *crossing loop* inferior y la región C-terminal [80].

Además, con el fin de estudiar con más profundidad las diferencias en la dinámica de BA42-Ca²⁺ y BA42-Mg²⁺, se llevaron a cabo experimentos de dinámica molecular (MD) de toda la proteína en solvente explícito y en presencia de uno de los dos iones (Ca²⁺ o Mg²⁺) en ambos sitios de coordinación. Este análisis mediante MD se realizó en colaboración con el Dr. Raúl E. Ithuralde y Dr. Adrián Turjanski del Departamento de Química Biológica de la Facultad de Ciencias Exactas y Naturales de la UBA. Las simulaciones de MD convergieron como se muestra en los gráficos de RMSD vs tiempo (Figura 2.15.B). Al analizar las fluctuaciones por residuos individuales, se observó un aumento significativo en la región C-terminal de BA42-Mg²⁺, como se puede ver en el gráfico de

RMSF (Figura 2.15.A) y en concordancia con nuestros experimentos de RMN.

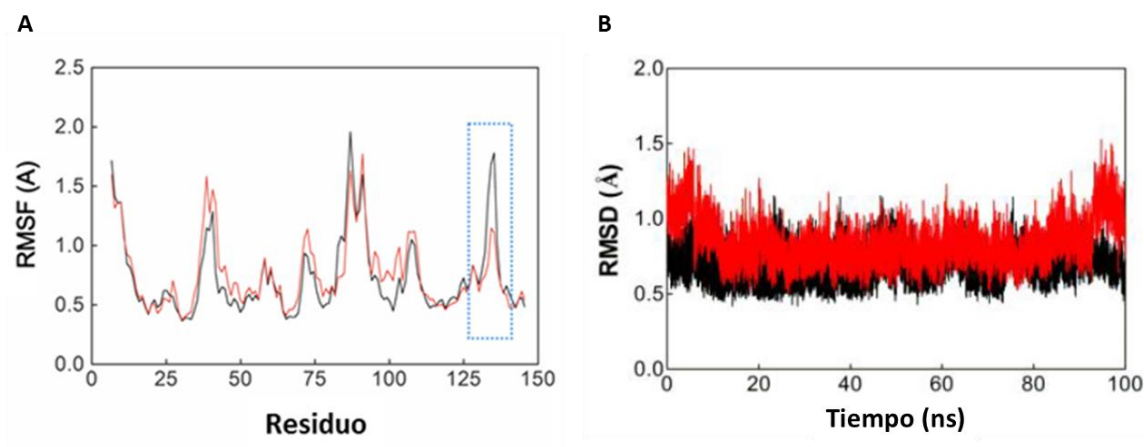


Figura 2.15. Simulaciones de dinámica molecular de BA42 unida a metales. (A) Se muestran los cálculos de RMSF para BA42 con Ca²⁺ (rojo) y con Mg²⁺ (negro). (B) RMSD calculado para BA42-Ca²⁺ (negro) y BA42-Mg²⁺ (rojo).

BA42-Mg²⁺ presenta cambios en la estabilidad térmica con respecto a BA42-Ca²⁺ y BA42-Apo

Como se detalló anteriormente existe una fuerte diferencia en la estabilidad conformacional entre BA42 en la conformación libre de metales y BA42 unida a Ca²⁺ [63]. Con el fin de evaluar si el cambio conformacional inducido por la unión de Mg²⁺ tenía un incidencia negativa sobre la estabilidad de la proteína, se llevaron a cabo experimentos de desnaturalización térmica monitoreados por dicroísmo circular (CD). En un primer experimento, se encontró que los espectros en el UV lejano de BA42-Ca²⁺ y BA42-Mg²⁺ eran casi idénticos entre sí, pero ligeramente diferentes al espectro de BA42 tratada con EDTA.

En los tres espectros se observó un amplio contenido de α -hélice con dos mínimos a 209 nm y 223 nm (Figura 2.16.A). A continuación se estudió la desnaturalización térmica de BA42-Apo, BA42-Ca²⁺ y BA42-Mg²⁺, registrando la elipticidad molar a 222 nm a diferentes temperaturas (Figura 2.16.B). Las tres variantes de BA42 ensayadas mostraron un proceso de desplegado-plegado reversible, evidenciándose un comportamiento cooperativo en el proceso de desnaturalización térmica. Sin embargo, BA42-Ca²⁺ resultó más estable que BA42-Mg²⁺, y esta última más estable que la BA42 libre de metales, siendo las T_{ms} calculadas $61,3 \pm 0,2$ °C, $57,0 \pm 0,2$ °C y $47,9 \pm 0,3$ °C, respectivamente (Figura 2.16.B).

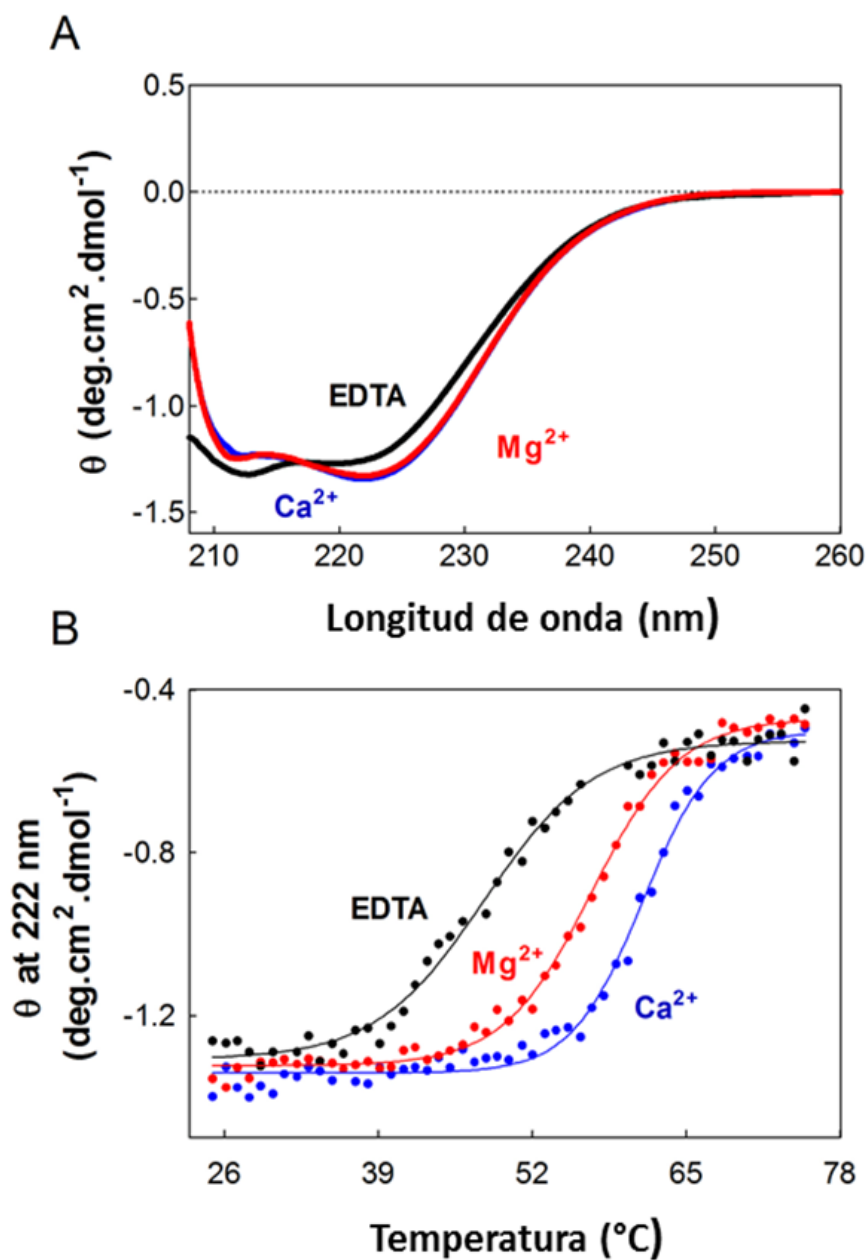


Figura. 2.16. (A) Espectros de dicroísmo circular de BA42. Los espectros de CD de BA42-Apo (línea negra), BA42-Mg²⁺ (línea roja) y BA42-Ca²⁺ (línea azul) se registraron a 25 °C. (B) Curva de desnaturalización térmica de BA42. Se incubaron BA42-Apo (línea negra), BA42-Mg²⁺ (línea roja) y BA42-Ca²⁺ (línea azul) a diferentes temperaturas desde 25 hasta 75 °C y la elipticidad fue registrada a 222 nm en intervalos de 1 °C.

Las mutantes de BA42 que son incapaces de unirse a Mg^{2+} revelan la localización del sitio activo

Como se mencionó anteriormente, la estructura cristalina del complejo AtTLP18.3-Ser [66] reveló la tríada VDK propuesta de constituir el sitio de unión al sustrato y un sitio de unión a calcio estabilizado por seis enlaces de hidrógeno probablemente implicado en la actividad fosfatasa. Sin embargo, de forma contraria a la proteína AtTLP18.3, la tríada VDK no está claramente presente en la estructura de BA42 y los dos iones de calcio están unidos al otro lado de la molécula (es decir, desplazados 20 Å del sitio de unión a Ca^{2+} correspondiente a AtTLP18.3) (Figura 2.9.B y C). En este punto, basándonos en nuestros experimentos, no podíamos excluir la coexistencia de dos sitios de unión de metales en BA42: uno estructural identificado en la estructura cristalina de BA42 y uno catalítico como se propone en AtTLP18.3.

Si el sitio estructural tuviera mayor afinidad que un posible sitio catalítico, sería difícil identificar los dos sitios a partir de las titulaciones. Por lo tanto, para evaluar estas posibilidades se procedió a la realización de experimentos de mutagénesis. Primero, se generó una variante de BA42 compuesta por los residuos 1-130 (BA42- Δ CT). BA42- Δ CT carece de la región C-terminal donde se encuentran la mayoría de los aminoácidos que constituyen el sitio de unión a los metales (Figura 2.17.A y B). Interesantemente, no se detectó actividad en presencia de Mg^{2+} para BA42- Δ CT en comparación con la proteína salvaje (*Wild type* – Wt) (Figura 2.17.D), descartando de esta forma la presencia de un sitio catalítico en BA42 en una región diferente de la constituida por el doble sitio de unión a metales. Además, los espectros 1D 1H de RMN de BA42- Δ CT se mantuvieron virtualmente sin cambios en presencia del ion divalente, en contraste con la proteína salvaje sin metales en la cual el cambio estructural generado por la unión de Mg^{2+} fue claramente evidente. (Figura 2.17.C).

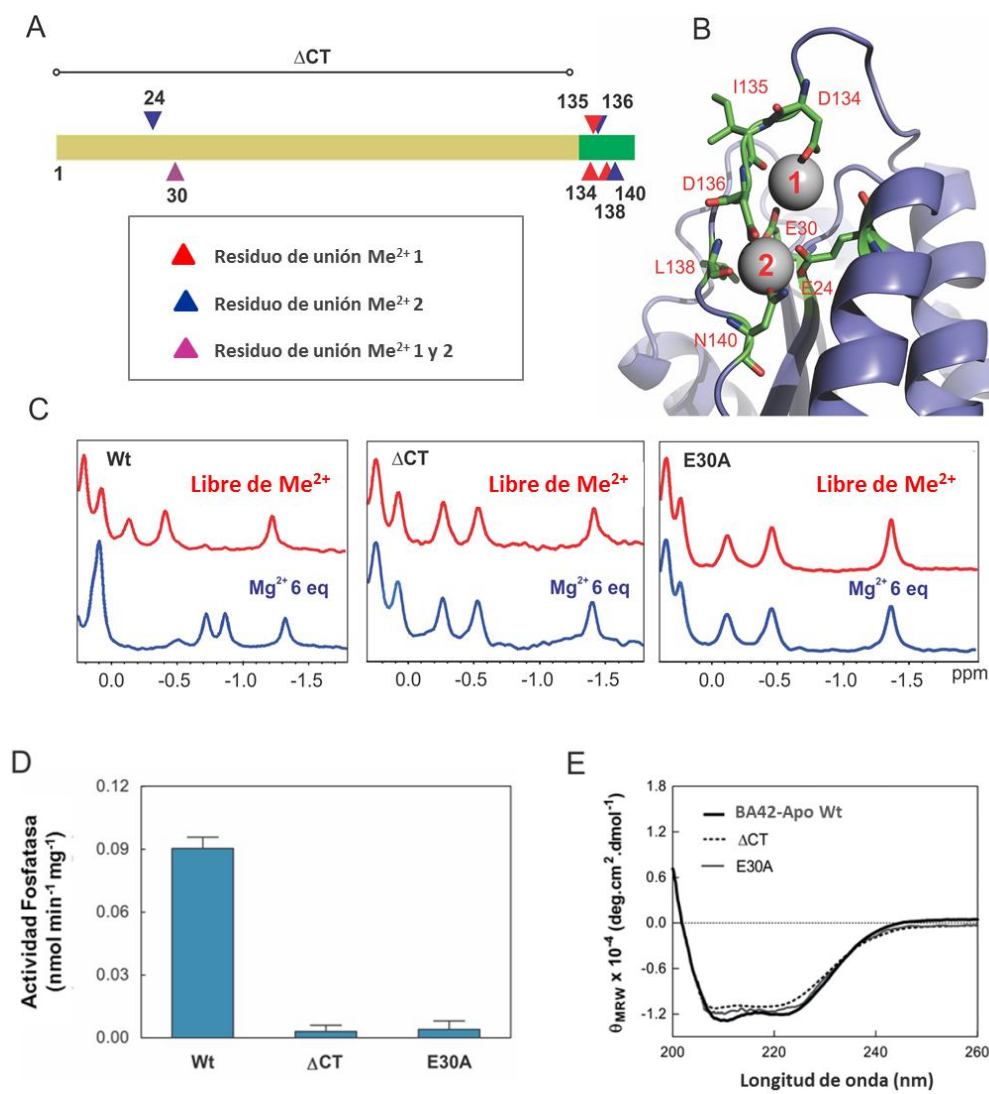


Figura. 2.17. Análisis de las mutantes de BA42. (A) Representación esquemática de la secuencia primaria de BA42 donde se indican los residuos que constituyen el sitio de unión metálico y la región C-terminal. Así mismo, se señala la constitución de la mutante BA42- Δ CT compuesta por los residuos 1-130 (rectángulo amarillo) careciendo de la región C-terminal (rectángulo verde). Los triángulos rojos indican los residuos involucrados en el sitio de unión del metal 1 y los triángulos azules resaltan aquellos que interactúan con el metal 2. El triángulo morado indica el ácido glutámico 30 que interactúa con ambos metales. (B) Representación de la región de unión a los metales en la estructura de BA42. Las esferas sólidas representan los iones Ca^{2+} (arbitrariamente marcados como 1 y 2) y los residuos involucrados en la unión de los metales se destacan en rojo. (C) Espectro 1D ^1H de las variantes de BA42 (Wt, Δ CT y E30A) en el intervalo de 0,25 a -1,75 ppm. Los espectros se registraron para las proteínas libres de metal y en presencia de 6 equivalentes de Mg^{2+} . (D) La actividad fosfatasa de BA42 Wt y las proteínas mutantes se ensayó en presencia de 6 equivalentes de Mg^{2+} . (E) Se registraron los espectros de CD de BA42-Apo Wt (línea negra), BA42- Δ CT (línea punteada) y BA42-E30A (línea gris) a 25°C.

Sin embargo, aunque los resultados obtenidos con BA42- Δ CT demuestran claramente la ausencia de un sitio catalítico fuera del sitio de unión de metales en BA42, la generación de esta mutante implica la eliminación de un gran número de aminoácidos que pueden conducir a un cambio conformacional sutil el cual podría eventualmente afectar de forma indirecta el potencial sitio catalítico. Por lo tanto, a modo de poder evaluar mejor la actividad enzimática de BA42 se consideró apropiado producir una variante de la proteína con un número menor de residuos mutados, que similar a BA42- Δ CT, conserve la característica de no unir el Mg^{2+} . De esta forma, se generaron variantes de BA42 que portaron mutaciones únicas en los residuos implicados en el sitio de unión de metales. Como resultado, se generó una simple mutante de BA42 del ácido glutámico 30 por alanina (BA42-E30A) la cual careció de la capacidad de unir Mg^{2+} [80] (Figura 2.17.C).

Es interesante observar que el ácido glutámico 30 tiene la particularidad de interactuar con los dos iones metálicos en la estructura de BA42 (Figura 2.17.B) y está altamente conservado entre los TPM que comprenden un dominio independiente (Figura 2.9.B). En línea con el comportamiento observado para BA42- Δ CT, la mutante BA42-E30A no mostró diferencias significativas entre los espectros 1D 1H de la proteína libre de metal y en presencia de Mg^{2+} (Figura 2.17.C). Pero aún más importante, BA42-E30A no mostró actividad fosfatasa (Figura 2.17.D). Esta observación fue congruente con los resultados anteriores con BA42- Δ CT, pero en este caso la pérdida de la actividad se asoció a una mutación puntual sobre un residuo específicamente implicado en el sitio de unión de metales de BA42, descartando cualquier artefacto debido a la eliminación de un gran número de aminoácidos.

Como control, se encontró que los espectros de CD de BA42-E30A y BA42- Δ CT fueron prácticamente idénticos al espectro de la proteína salvaje libre de metal (Figura 2.17.E), indicando así la integridad de la estructura secundaria y la ausencia de grandes cambios conformacionales de las proteínas mutantes con respecto a BA42-Apo salvaje.

Estudio *in vivo* de BA42

Con el objetivo de avanzar en la comprensión de la función biológica de BA42 se comenzó a estudiar la proteína BA42 a nivel *in vivo*. Para esto, en primera instancia se procedió a la producción de anticuerpos policlonales en ratones contra la proteína BA42, los cuales se lograron obtener en un buen título y resultaron tener una alta especificidad. Con el fin de comprobar que BA42 se expresa a nivel proteico se realizó un Western Blot utilizando un extracto de *B.*

argentinensis crecida en LB marino mediante los anticuerpos anti-BA42. El resultado evidenció la expresión de la fosfatasa en dichas condiciones de estudio (Figura 2.18.A).

Una forma de poder obtener información sobre el rol celular de una proteína es estudiar las posibles interacciones proteicas que ocurren a nivel celular. En este contexto, mediante la utilización de los anticuerpos policlonales contra BA42, se realizaron ensayos de co-inmunoprecipitación (IP) seguidos de un análisis por SDS-PAGE (Figura 2.18.B). Entre las proteínas co-inmunoprecipitadas se observaron diferentes bandas relativamente intensas que no estaban presentes en el control negativo (datos no mostrados).

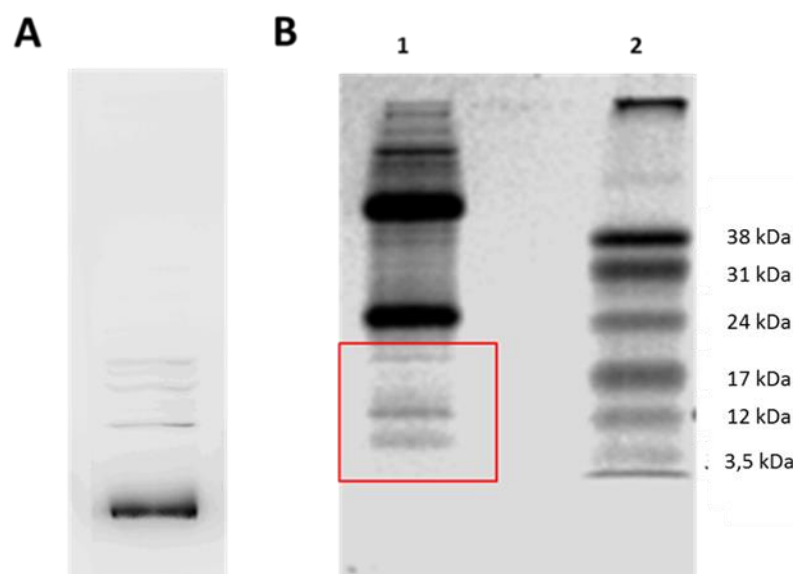


Figura 2.18. (A) Análisis de la expresión de la proteína BA42 mediante Western Blot en *B. argentinensis* en condiciones de crecimiento en LB marino. (B). Análisis por SDS-PAGE (15 % acrilamida) de la IP de extractos proteicos totales de *B. argentinensis* con el anticuerpo anti-BA42. En rojo se enmarcan las bandas que no aparecieron en el control negativo (datos no mostrados), las cuales fueron purificadas para un posterior análisis por espectrometría de masa (Carril 1). Marcador de peso molecular (Carril 2)

Estas fueron purificadas del gel de poliacrilamida y se les realizó un análisis por espectrometría de masa. Como resultado, se observaron proteínas de PM de entre 14 y 20 kDa, entre las cuales, tal como se esperaba, se encontró la fosfatasa BA42, proporcionando un control positivo de la co-inmunoprecipitación. Interesantemente, junto con BA42 se observó la presencia de otras proteínas de PM de entre 15 y 20 kDa. Entre estas se encontraron proteínas ribosomales pertenecientes a las subunidades mayor y menor junto con un factor de elongación TU. También se observaron proteínas de membrana pertenecientes a la superfamilia de las porinas, receptores

dependientes de TonB y proteínas de la familia Pfam OMP_b_brl, siendo este un dominio que se encuentra en una amplia gama de proteínas de la membrana externa. Pero de forma muy interesante, dentro de este grupo se encontró una proteína de membrana perteneciente a la familia Pfam LemA que se encuentra dentro del contexto genómico de BA42, formando un probable operón con la fosfatasa (ver más adelante).

BA42 integraría un operón

A fin de obtener información sobre el contexto genómico en el cual se encuentra la proteína BA42, se realizó un análisis mediante un conjunto de programas de predicción de genes bacterianos y operones (GENESB, <http://linux1.softberry.com>). Estos programas realizan la predicción de los operones sobre la base de las distancias entre los ORFs contiguos y las frecuencias de éstos en diferentes genomas bacterianos, utilizándose también la predicción de las regiones promotora y terminadora para definir el mismo. Mediante esta herramienta bioinformática se concluyó que esta proteína formaría parte de un operón compuesto de cinco proteínas y un promotor divergente (Figura 2.19). Entre las proteínas que integran el probable operón se puede encontrar: una proteasa perteneciente a la familia M23, la cual podría tener similitud con proteasas que degradan elastina en el ambiente marino [82]; una proteína de membrana que contiene un dominio TPM, el cual fue recientemente caracterizado como una ATPasa/ADPasa [83]; una proteína perteneciente a la familia Pfam LemA, la cual se desconoce su función; y un probable factor de transcripción perteneciente a la familia de reguladores MerR [84], relacionados a la regulación de procesos celulares en situaciones de estrés. Esta última proteína fue nombrada como BA40 y se inició su caracterización con el fin de obtener mayor información sobre el rol biológico del operón y, en consecuencia, de BA42. El estudio inicial de BA40 se detalla a continuación en el capítulo 3.

En el sentido opuesto de transcripción del operón se presenta la proteína alanil-tRNA sintetasa, que sintetiza la incorporación de alanina al RNA de transferencia correspondiente y podría eventualmente encontrarse relacionada a la función de los genes del operón, aunque todavía no se cuenta con suficiente información para establecer un vínculo entre ellos.

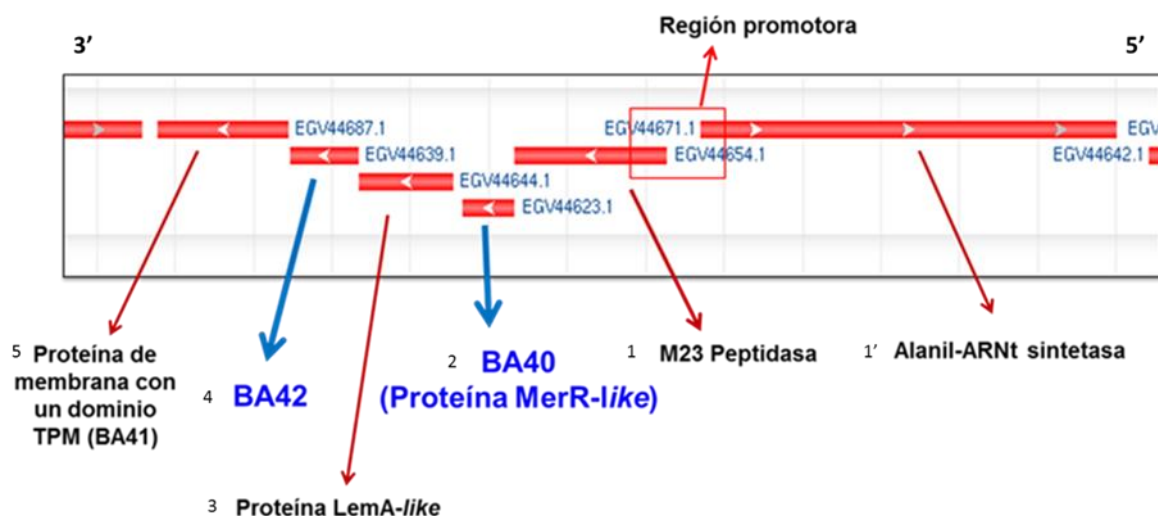


Figura 2.19. Esquema del operón predicho bioinformáticamente. En la figura se señalan en su orden de transcripción 5'-3' los 5 genes que integrarían el operón: 1, *peptidasa* perteneciente a la familia M23; 2, *BA40*, probable factor de transcripción perteneciente a la familia de reguladores MerR; 3, *proteína* perteneciente a la familia Pfam LemA; 4, fosfatasa *BA42*; 5, *proteína* de membrana que contiene un dominio TPM (BA41); promotor divergente de dicho operón. En el sentido de transcripción 3'-5': gen de la proteína *alanil ARNt sintetasa*.

Discusión

BA42 es el primer miembro de la familia PF04536 que comprende un dominio TPM independiente que ha sido estructural y bioquímicamente caracterizado hasta la fecha. La estructura de BA42 descubre una nueva variante topológica de las cuatro hebras β que constituyen la hoja β central de la arquitectura $\alpha\beta\alpha$ y un sitio de unión a metales que estabiliza un par de "crossing loops", no observado en estructuras previas de proteínas pertenecientes a esta familia. Además, en contraste con otros miembros caracterizados de la familia TPM que son proteínas de membrana, BA42 tendría una localización citoplasmática.

Anteriormente, la proteína AtTPL18.3 que contiene un dominio TPM caracterizado estructuralmente reveló poseer actividad fosfatasa [66]. En este contexto, a partir de la comparación de las estructuras de BA42 y el dominio TPM de AtTPL18.3 se reveló que a pesar del hecho de que BA42 contiene residuos conservados involucrados en el mecanismo catalítico de AtTPL18.3, estos no tienen una disposición espacial compatible con su sitio activo [66]. Sin embargo, en este capítulo se demostró que BA42 posee igualmente actividad fosfatasa en las

condiciones ensayadas. Aunque los valores absolutos de la actividad son bajos en comparación con otras enzimas hidrolasas [85], están en línea con los informados anteriormente para otros miembros de la familia TPM [66, 70]. Además, dado que no se conoce el sustrato fisiológico de BA42, la baja actividad observada *in vitro* también puede atribuirse a condiciones de reacción subóptimas.

En esta tesis, se muestra la estricta dependencia de la actividad de BA42 de la presencia de iones divalentes en su estructura, observándose que la proteína no presenta actividad cuando es pretratada con EDTA y recuperado la misma después del agregado de diferentes cationes metálicos. Interesantemente, se observó un aumento en la actividad de aproximadamente cinco veces en presencia de Mg^{2+} en comparación con los otros metales probados. Este aumento en la actividad enzimática se asocia con un cambio conformacional de BA42 desde una estructura con una región C-terminal desplegada a una estructura cerrada con el extremo C-terminal unido al núcleo proteico [80]. Sin embargo, todos los metales probados inducen un cambio conformacional similar en BA42 e incluso con mayor afinidad que el Mg^{2+} . Las principales diferencias en la conformación de BA42- Mg^{2+} residen en la dinámica de los “*crossing loops*” y el sitio de unión de metales. Dos experimentos independientes indican un aumento en la flexibilidad de estas regiones. En primer lugar, los espectros HSQC de la proteína unida a Mg^{2+} revelan la desaparición de las señales de amida de residuos adyacentes al sitio de unión de los cationes y los residuos 131-137, que comprenden el “*crossing loop*” superior en la estructura de BA42 con Ca^{2+} . Dado que no se detectó en esta región una gran variación de CS ni existe un aumento en la exposición al solvente, se concluyó que la desaparición de estos picos se debe a un aumento de la flexibilidad (en la escala de movimientos lentos o intermedios, μ s-ms). En segundo lugar, los experimentos de relajación del ^{15}N muestran claramente un aumento en los movimientos lentos (μ s-ms) de los residuos D22-E30, que comprenden el “*crossing loop*” inferior, y N140-K144, que pertenecen al extremo C-terminal. Además, la menor estabilidad conformacional de BA42- Mg^{2+} con respecto a BA42- Ca^{2+} , detectada en experimentos de desnaturalización térmica, podría atribuirse a un aumento en la flexibilidad de esta región. En concordancia con esto, las simulaciones de MD proporcionan una evidencia adicional de que existe una mayor flexibilidad en los residuos que pertenecen al par de “*crossing loops*” y que comprenden el sitio de unión de metal. En conjunto, estos resultados indican que la unión de Mg^{2+} a BA42 induce un aumento de la flexibilidad local en la región que comprende el par de “*crossing loops*” y los residuos implicados en la unión de los metales. Esta característica se asocia con una disminución de la estabilidad conformacional global

y una mejora de la actividad catalítica, en comparación con BA42-Ca²⁺. En este sentido, al comparar los valores de $K_{M\ ap}$ y k_{cat} determinados para BA42-Ca²⁺ y BA42-Mg²⁺ se observa un aumento en el valor de dichas constantes cuando la proteína está unida a Mg²⁺. Este hecho concuerda con lo observado en muchas enzimas de organismos psicrófilos en las cuales la generación de estructuras más flexibles se asocia a aumentos en los valores de k_{cat} y K_M [39, 40].

Con el fin de evaluar aún más la relación entre la flexibilidad y la actividad catalítica, se comparó la actividad de la enzima conteniendo Ca²⁺ y Mg²⁺ en función de la temperatura (Figura 2.20). Aunque la actividad de BA42-Ca²⁺ fue inferior a la obtenida para BA42-Mg²⁺ en todo el rango de temperatura estudiado, las principales diferencias se presentaron por debajo de 26 °C. Interesantemente, se encontró que la actividad máxima de BA42-Ca²⁺ se observó a 37 °C, desplazada alrededor de 17° C con respecto a la máxima actividad observada para BA42-Mg²⁺ (alrededor de 20-22°C). En consecuencia, estos resultados fueron compatibles con la idea de que la presencia de un sitio activo flexible en BA42 genera una enzima más eficiente en la catálisis a bajas temperaturas.

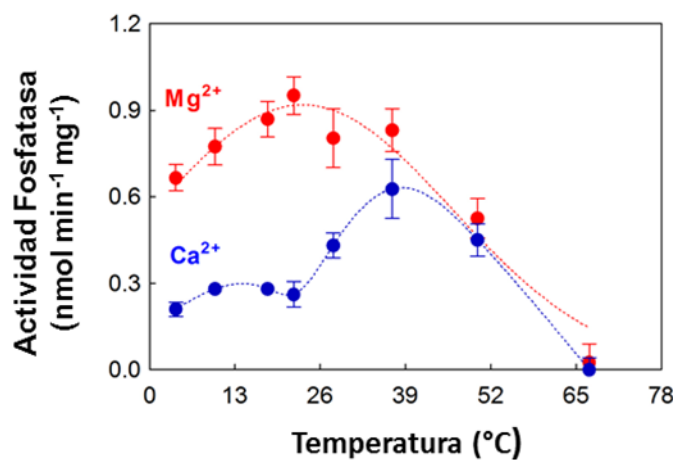


Figura 2.20. Efecto de la temperatura sobre la actividad enzimática de BA42. La actividad fosfatasa se midió a diferentes temperaturas en HEPES 25 mM (pH 7,0) y en presencia de 1 mM CaCl₂ (puntos azules) o 1 mM MgCl₂ (puntos rojos). Todos los experimentos se realizaron por triplicado y los resultados se expresan como media ± EEM.

En respuesta a las bajas temperaturas, los psicrófilos utilizan diversas estrategias de adaptación [36, 86, 87]. En la mayoría de los organismos la flexibilidad de los elementos estructurales que están implicados en el ciclo catalítico de las enzimas se ajustan casi siempre

dando lugar a una estructura más flexible caracterizada por una baja estabilidad conformacional local y/o por la desestabilización de la molécula completa [35, 40]. Esto implica una reducción en el número y la fuerza de todos los tipos de interacciones débiles o la desaparición de los factores de estabilidad, dando como resultado una dinámica aumentada de los residuos del sitio activo en el frío. En este contexto, en base a nuestros resultados, podemos plantear la hipótesis de que la región que comprende el par de “*crossing loops*” y los residuos involucrados en la unión de los metales en BA42 están directamente relacionados con el mecanismo catalítico y puede contener el sitio activo [80]. Además, se ha observado en varias enzimas que las estructuras de “*crossing loops*” forman parte del sitio activo [88, 89].

En este sentido, los resultados obtenidos con BA42- Δ CT y BA42-E30A fueron cruciales para circunscribir el sitio activo de BA42 en una región constituida por el sitio de unión de metales y eliminar la existencia de un sitio activo similar al descrito para AtTLP18.3 [66]. En este punto, dadas las diferencias observadas entre BA42 y AtTLP18.3 con respecto a la topología de plegamiento, el sitio de unión de metales y la localización del sitio activo, sería útil que para estudiar más adecuadamente las proteínas que contienen dominios TPM, éstas se clasifiquen en al menos dos grupos: TPMs asociados a membranas y TPMs independientes citoplasmáticos. Los TPMs citoplasmáticos parecen estar casi exclusivamente conservados dentro del filo Bacteroidetes, mientras que los TPMs asociados a membranas parecen conservarse en una amplia gama de organismos (incluyendo plantas, hongos y bacterias). En particular, ya que ambos grupos de proteínas que contienen dominios TPM se encuentran en el grupo Bacteroidetes, se plantea la hipótesis de que estos dominios podrían haber evolucionado de forma independiente. Además, los TPMs citoplasmáticos independientes se encuentran principalmente en organismos marinos adaptados a bajas temperaturas y por lo tanto, suponemos que los miembros de este grupo conservan características estructurales compatibles con la actividad en estas condiciones ambientales, similares a las descritas para BA42.

Por lo tanto, es interesante plantear la hipótesis de que en BA42 podría haber evolucionado un sitio catalítico donde la actividad a baja temperatura esté regulada, al menos en parte, mediante la variación de la flexibilidad del sitio activo. En presencia de Ca^{2+} el sitio activo se volvería rígido y la actividad bajaría considerablemente, mientras que con la unión de Mg^{2+} aumentaría la flexibilidad de una manera compatible con la catálisis a bajas temperaturas [80] (Figura 2.21).

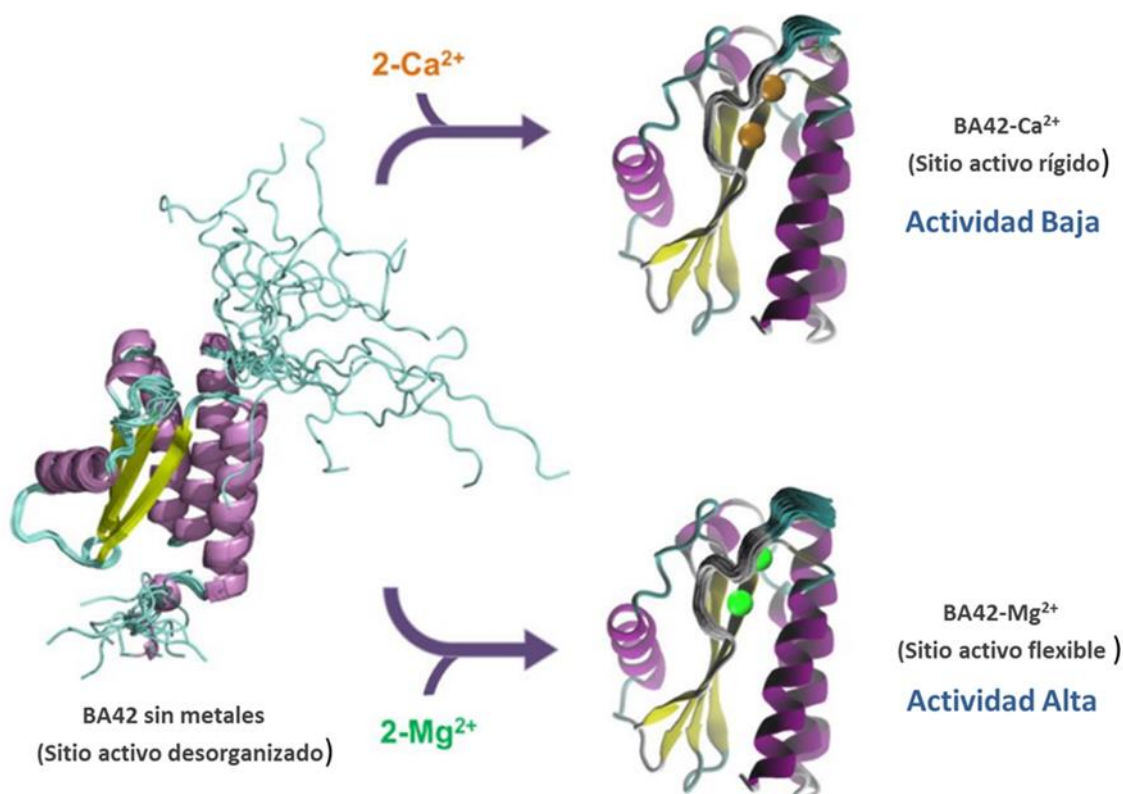


Figura 2.21. Representación esquemática del comportamiento de la región catalítica de BA42 dependiente de metales. La figura ilustra el mecanismo postulado de la modulación de la actividad fosfatasa variando la flexibilidad del sitio activo. En ausencia de iones metálicos divalentes, la región C-terminal de BA42 está extremadamente desordenada y no se forma el sitio activo. En presencia de Ca²⁺ el sitio activo adquiere una configuración rígida y la actividad es baja. La unión de Mg²⁺ induce un incremento de la flexibilidad del sitio activo, en comparación con la proteína unida a Ca²⁺, y la actividad de la proteína se incrementa.

En este sentido, El Mg²⁺ es el metal divalente más abundante en las células vivas. Se ha reportado que los niveles intracelulares de Mg²⁺ en las bacterias marinas pueden variar de 130 a 710 mM en condiciones normales de crecimiento [90] e incluso hasta 910 mM bajo la limitación de carbono o latencia [91]. Además, la concentración intracelular de Ca²⁺ en estos organismos oscila entre 10 y 80 mM [90]. Por lo tanto, dada la alta concentración de estos dos metales, se puede predecir que BA42 debería encontrarse en el citoplasma en la conformación unida a metales y los niveles relativos de Ca²⁺ y Mg²⁺ podrían ser un factor importante para modular su actividad fosfatasa.

En conclusión, los resultados mostrados aquí contribuyen a la comprensión de la relación entre la estructura y la función de esta familia recientemente caracterizada de dominios TPM. Este trabajo, junto con otros anteriormente descritos hasta la fecha, sugiere que la actividad fosfatasa en los dominios TPM se ha conservado a lo largo de la evolución. Sin embargo, se necesitan nuevas investigaciones que proporcionen más datos para continuar indagando en los mecanismos catalíticos y de regulación subyacentes.

La mayor parte de los resultados expuestos en este capítulo se encuentran publicados en Aran et al. [63] y Pellizza et al. [80].

Capítulo III

Caracterización estructural y funcional de la proteína BA40

Para obtener más información del probable operón integrado por BA42 se comenzó a estudiar bioquímicamente y estructuralmente al probable factor de transcripción perteneciente a la familia de reguladores MerR, al que denominamos BA40. Todos los estudios y resultados que se detallan en este capítulo abarcan la caracterización inicial de esta proteína.

La familia MerR está constituida por reguladores transcripcionales los cuales contienen regiones de unión al ADN altamente conservadas conteniendo el motivo hélice-vuelta-hélice (HTH-*Helix Turn Helix*). Este motivo, que es un dominio en sí mismo, es extremadamente común en los factores de transcripción y se presenta en los organismos de todos los reinos de la vida [92]. Particularmente, en los factores de transcripción tipo MerR el motivo HTH se ubica en el extremo N-terminal donde es parte fundamental del dominio de unión al ADN, el cual se continúa de una región coiled-coil propuesta de estar involucrada en la dimerización de estos factores de transcripción. Estas dos regiones (Dominio de unión al ADN+ coiled-coil) son el distintivo de esta familia y en conjunto involucran alrededor de 100 aminoácidos. Adicionalmente, la mayor parte de las proteínas de esta familia contienen en sus extremos C-terminales regiones con sitios de unión específicos para efectores que modulan su actividad de unión al ADN (Figura 3.1). Contrariamente a la región N-terminal (dominio HTH), las regiones C-terminales efectoras no están conservadas a nivel de sus secuencias y son específicas de diferentes efectores tales como iones metálicos, drogas, sustratos orgánicos o diferentes proteínas [84].

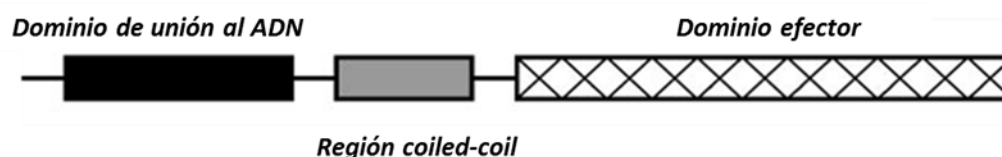


Figura 3.1 Arquitectura canónica de la familia MerR. Se muestran los diferentes dominios que constituyen la arquitectura general de las proteínas pertenecientes a la familia MerR. El dominio de unión al ADN se muestra como un rectángulo negro, las regiones capaces de formar estructuras coiled-coil se representan con rectángulos grises. Los rectángulos cruzados representan los dominios de transducción de señales. El tamaño de cada rectángulo es proporcional al tamaño de cada dominio en las proteínas.

Los factores de transcripción MerR se encuentran distribuidos fundamentalmente en todos los géneros de eubacterias y en algunas arqueas. Con el creciente reconocimiento de los miembros de esta familia en la última década, se ha observado que la mayor parte de estos reguladores median respuestas a diferentes situaciones de estrés incluyendo la exposición a

diferentes estímulos ambientales, metales pesados, especies reactivas de oxígeno, antibióticos, etc. La activación de la transcripción que producen estas proteínas ocurre a través de la distorsión del ADN. Se han resuelto varias estructuras de los dominios HTH de unión al ADN pertenecientes a reguladores transcripcionales de tipo MerR. Específicamente, este dominio de unión al ADN presenta un motivo HTH con alas (wHTH - *winged helix-turn-helix*), que consiste generalmente de cuatro α -hélices (H1, H2, H3 y H4) y una hoja- β anti paralela de tres hebras- β (B1, B2 y B3) en la topología: B1-H1-H2-B2-B3-H3-H4 [93, 94], uniéndose al ADN en el surco mayor (Figura 3.2). El motivo HTH está comprendido por las α -hélices H1 y H2, y es la segunda α -hélice (H2) por la que ocurren el mayor número de contactos con el ADN, por lo que se la denomina hélice de reconocimiento. El segundo elemento estructural de unión al ADN, denominado ala W1 (*Wing 1*), se compone por la segunda y tercera hebras- β y el *loop* por el cual están unidas (B2-B3). Existe un tercer elemento estructural de unión a ADN, llamado ala W2, el cual está formado por las α hélices tres y cuatro (H3-H4) (Figura 3.2).

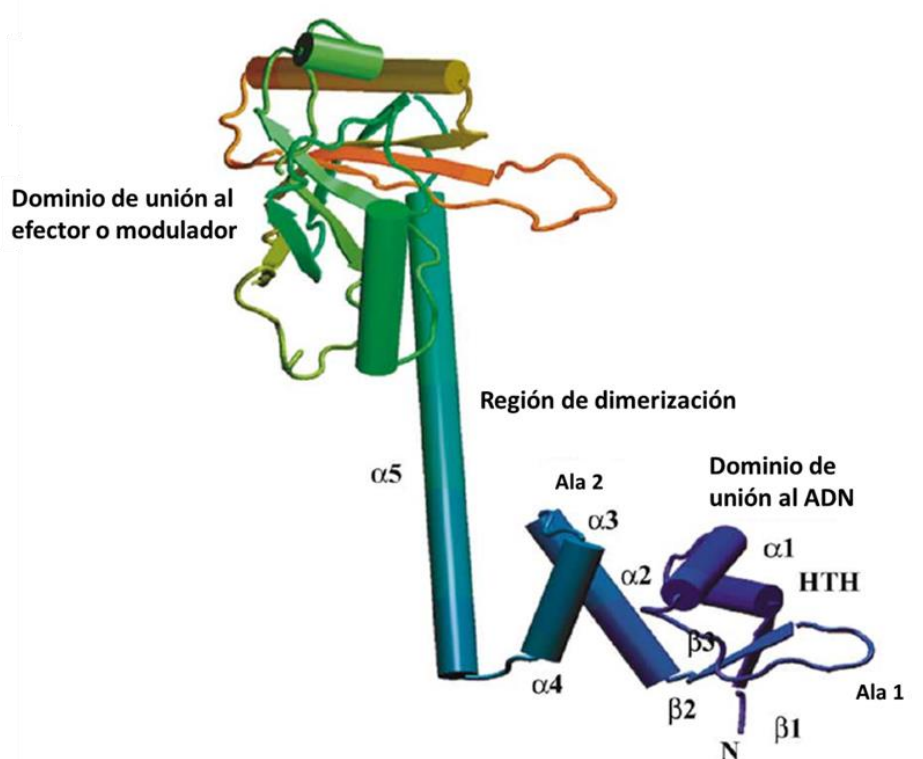


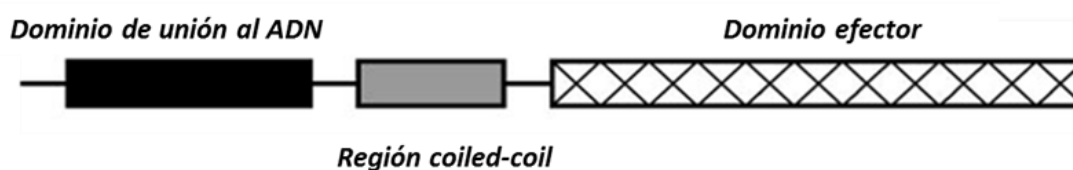
Figura 3.2. Estructura del monómero del factor de transcripción BmrR perteneciente a la familia MerR. Se detallan el dominio de unión al ADN con su motivo HTH y las dos alas (azul y lila), la región de dimerización (turquesa) y el dominio de unión al efector (verde y naranja). Tomado de Heldwein [95].

En un regulador MerR típico, el HTH y las dos alas W1 y W2 se unen al promotor del operón regulado en las secuencias -35 y -10 las cuales están separadas por una región espaciadora de 19/20 pb (más larga que en la mayoría de los promotores de otros factores de transcripción), distorsionando el ADN del operador y haciendo que la ARN polimerasa inicie la transcripción o inhibiendo el comienzo de la misma [84]. La mayoría de los reguladores transcripcionales tipo MerR son dímeros.

A partir del análisis de la secuencia de BA40 se observó que esta proteína posee una arquitectura diferente a la canónica de la familia MerR (Figura 3.1). BA40 está compuesta por 109 aminoácidos y, por lo tanto, es más pequeña que la mayor parte de los miembros de esta familia (~240 aminoácidos). En consecuencia, BA40 posee sólo dos de los dominios característicos de estos factores de transcripción; el dominio HTH y la región *coiled-coil* de dimerización (Figura 3.3.B). Este hecho contrasta con la mayoría de las proteínas pertenecientes a esta familia, las cuales poseen dominios C-terminales transductores de señales específicos para diferentes efectores luego de la región de dimerización (Figura 3.1 y 3.3.A). En este sentido, las únicas dos proteínas estudiadas hasta el momento que presentan una arquitectura similar a BA40, aunque con una muy baja identidad de secuencia, son TnrA y GlnR de *Bacillus subtilis*. La proteína TnrA es un regulador maestro de la asimilación de nitrógeno durante el crecimiento de la bacteria en condiciones limitantes del mismo [96, 97] mientras que GlnR reprime la expresión génica durante el crecimiento bacteriano en exceso de nitrógeno [96, 98].

Diferentes estudios sobre TnrA y GlnR han observado que en estos casos el dominio efector o transductor de señales está formando parte de la misma región de dimerización. Más aun, se ha demostrado en ambos factores que la modulación de la actividad de unión al ADN ocurre por la interacción de la glutamina sintetasa (GS) con sus regiones C-terminales [97].

A



B



Figura 3.3. Dos tipos de arquitecturas en las proteínas de la familia MerR. Los rectángulos negros denotan dominios de unión al ADN. Los rectángulos grises denotan regiones de proteína capaces de formar estructuras coiled-coil. Los rectángulos cruzados representan los dominios de transducción de señales. El tamaño de cada rectángulo es proporcional al tamaño de cada dominio de proteína. **(A)** Se utilizó la proteína BmrR como modelo para derivar de su estructura cristalina la arquitectura mostrada [95]. Se observa el dominio C-terminal de transducción de señales independiente. Este tipo de arquitectura es mayoritaria en la familia **(B)** Se utilizó la proteína TnrA como modelo para representar la arquitectura mostrada. Se observa la característica sobresaliente de la superposición de la región de dimerización con el dominio efector. Tomado de Wray et al, 2007 [99].

Un estudio bioinformático de la proteína BA40 en búsqueda de secuencias ortólogas usando el programa BLASTp, reveló que BA40 se encuentra principalmente conservada dentro del filo Bacteroidetes. Sin embargo, se observó que su región N-terminal HTH está más ampliamente conservada en otros filos bacterianos. En particular, este dominio presentó mayor identidad de secuencia con el dominio HTH de una subfamilia de factores de transcripción MerR denominada MlrA. En este sentido, se ha demostrado que la proteína MlrA tiene un rol clave en la regulación positiva de la transcripción del factor CsgD el cual controla la producción de matriz extracelular y curli en *Escherichia coli* y *Salmonella typhimurium* modulando la agregación célula-célula [100].

Contrario a lo observado para el dominio HTH, la comparación de las secuencias primarias de los ortólogos de BA40 mostró que la región C-terminal, en donde se localizaría la α -hélice H5, está exclusivamente conservada en el filo Bacteroidetes (Figura 3.4). Sin embargo, se observa que el porcentaje de identidad (ID) de esta región, abarcando los últimos 20 aminoácidos del extremo C-terminal, es de 25 %, menor al que presentan el dominio HTH (55 %) y otras regiones de la cadena polipeptídica. Esta observación indicaría que estas dos regiones (N-terminal y C-terminal) estarían siendo sometidas a distintas presiones de selección a nivel de la secuencia primaria. Posiblemente, la región C-terminal haya evolucionado en base a una selección estructural más que de secuencia.

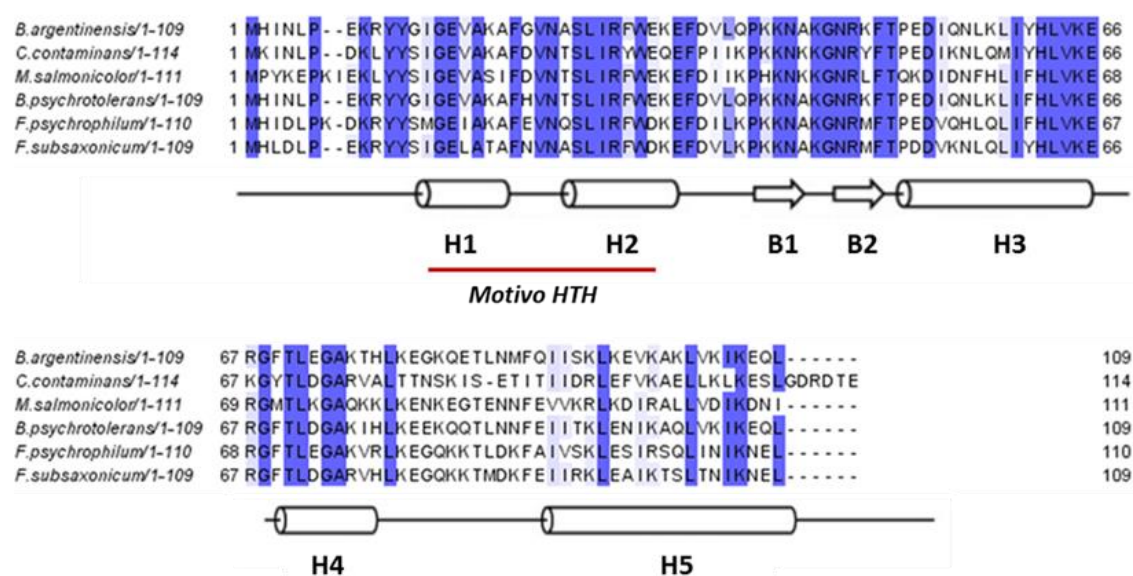


Figura 3.4. Alineamiento múltiple de BA40 y cinco homólogos pertenecientes al filo Bacteroidetes y predicción de su estructura secundaria. El alineamiento fue generado por ClustalW (<http://www.ebi.ac.uk/clustalW>). Los residuos sombreados en azul se conservan estrictamente y los sombreados en azul claro se conservan parcialmente. Debajo, se representan los elementos de estructura secundaria correspondientes obtenidos mediante el programa JPred. Se muestran las α -hélices como flechas y las hebras- β como cilindros.

Para proseguir con el estudio de BA40 se realizó la predicción de su estructura secundaria mediante el programa JPred. Se observó que BA40 se compone de cinco α -hélices (H1, 2, 3-5) y dos hebras- β (B2, 3) probablemente en la topología H1-H2-B2-B3-H3-H4-H5. En concordancia con los diferentes factores de transcripción tipo MerR que han sido estudiados estructuralmente, el

probable dominio de unión al ADN de BA40 contiene un motivo HTH, y las dos alas descritas anteriormente, formadas por las hebras- β B1-B2 (ala 1) y por las α -hélices H3-H4 (Figura 3.2), situándose este dominio entre los residuos 12 y 79 (Figura 3.4). La α -hélice H5 se compondría de 20 aminoácidos y estaría conectada al dominio de unión al ADN por una región no conservada (Figura 3.4). Como ya se mencionó, esta α -hélice C-terminal ha sido descrita de formar una estructura coiled-coil implicada en la dimerización de proteínas pertenecientes a esta familia [84, 94].

BA40 muestra actividad de unión a ADN

Como se detalló anteriormente, BA40 contiene un dominio de unión a ADN con una alta similitud a la subfamilia MlrA, lo que implica que esta proteína podría actuar como un potencial factor de transcripción. A partir de esto, se propuso la realización de diferentes ensayos *in vitro* a fin de develar si BA40 podría presentar actividad de unión al ADN. Por lo tanto, en primera instancia el gen *BA40* se amplificó por PCR utilizando como molde el ADN genómico de *B. argentinensis* y consecutivamente se precedió al clonado del mismo mediante el uso del sistema GATEWAY (Invitrogen, Carlsbad, CA). La construcción de BA40 se diseñó para que contenga en su extremo N-terminal un sitio de clivaje para la proteína TEV seguido de un “tag” de 6X histidinas para su purificación mediante cromatografía de afinidad. Posteriormente, BA40 fue sobreexpresada en células de *E. coli* en medio LB empleándose la cepa BL21 DE3 (Stratagene, La Jolla, CA). Finalmente, BA40 fue purificada a homogeneidad mediante cromatografía de afinidad, exclusión molecular e intercambio iónico (ver Materiales y Métodos).

A continuación, luego de conseguir la proteína recombinante pura, se llevó a cabo un ensayo de retardo en la movilidad electroforética (EMSA) para analizar la capacidad de la proteína BA40 de unirse a un fragmento de ADN aleatorio. El resultado mostró que la proteína BA40 interactúa con el ADN doble cadena (Figura 3.5), evidenciando que probablemente tenga el rol predicho como factor de transcripción.

Como se observa en el gel, este ensayo produce condiciones muy disociantes reflejadas en la forma de las bandas retardadas de ADN-Proteína. Cabe mencionar que el valor del pI de BA40 es de 9,8 lo cual resulta en una mayor dificultad del complejo ADN-BA40 para ingresar al gel en las condiciones ensayadas (pH 7,5).

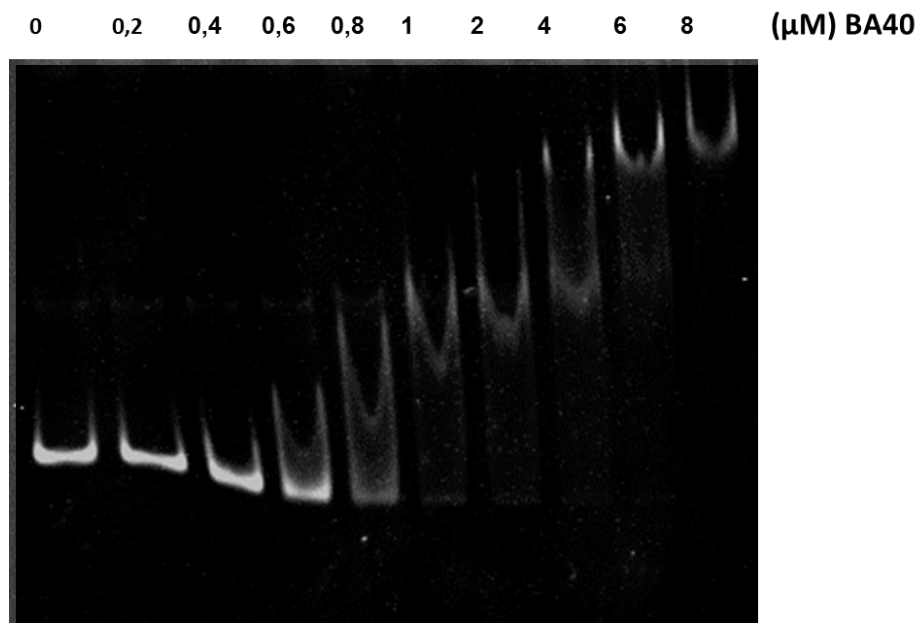


Figura 3.5. EMSA de la proteína BA40 en gel de poliacrilamida teñido con bromuro de etido. Se incubó un fragmento de ADN de 200 pb con concentraciones crecientes de BA40 en el rango de 0 a 10 μ M.

A partir de este resultado y con el objetivo de encontrar una secuencia de ADN a la cual BA40 se uniera de forma específica, comenzamos a estudiar diferentes secuencias de ADN pertenecientes a regiones promotoras de genes homólogos a los regulados por la proteína MlrA, dado que podrían estar relacionados con BA40 por la alta identidad de secuencia con su dominio de unión a ADN. De esta manera, se ensayó la secuencia promotora de un gen ortólogo a *CsgD* [101] (principal gen regulado por MlrA) en *B. argentinensis*. También, dado que se ha observado que muchos factores de transcripción tipo MerR controlan su propia transcripción, se ensayó la secuencia de ADN del promotor del operón *BA42*. En todos estos casos el resultado obtenido fue el mismo que para una secuencia de ADN aleatoria, concluyéndose que probablemente BA40 tenga una función regulatoria diferente a la de MlrA.

BA40 es un monómero en solución

Cómo ya se mencionó, los miembros de la familia MerR son en general dímeros. Esa estructura cuaternaria se estabiliza gracias a la formación de una estructura coiled-coil anti-paralela entre las hélices H5 de cada monómero [84]. En este sentido, con el fin de estudiar el estado de oligomerización de BA40 se realizó inicialmente una cromatografía de exclusión

molecular (SEC). En estos ensayos, se observó que la proteína eluyó en un solo pico con un tamaño aparente cercano a 13 kDa (ver Materials y Métodos), siendo el peso molecular teórico para el estado monomérico de 12,7 kDa. Sorprendentemente, a diferencia de la mayoría de los miembros de la familia MerR, BA40 parecería encontrarse como monómero en solución.

A fin de estudiar este comportamiento particular, en primera instancia realizamos una predicción bioinformática mediante el programa COILS, que estima la probabilidad de adoptar una conformación coiled-coil en proteínas. Este motivo estructural proteico contiene dos o más hélices enrolladas sobre sí mismas y generalmente presenta un patrón repetido de residuos de aminoácidos hidrofóbicos y polares, llamado héptada repetitiva. Las posiciones relativas de estos residuos siguen un patrón *abcdefg*, donde a y d son las posiciones hidrofóbicas, que forman la interfaz entre las hélices [102]. Como resultado, se observó que los últimos 14 aminoácidos de BA40 pertenecientes a la α -hélice H5 tienen una alta probabilidad de adoptar una estructura tipo coiled-coil (Figura 3.6).

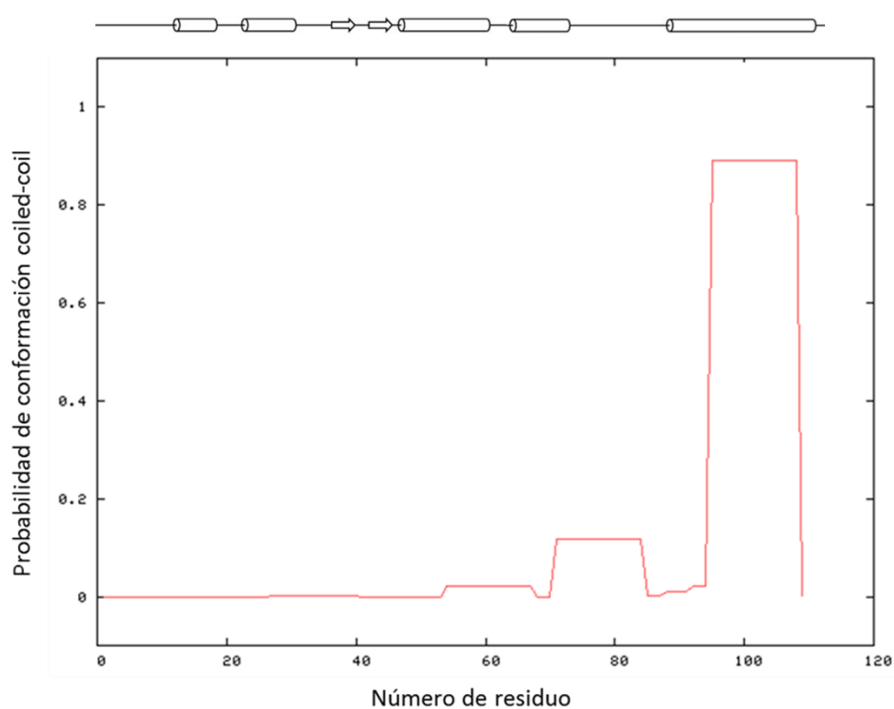


Figura 3.6. Predicción de regiones coiled-coil en la proteína BA40 utilizando el programa COILS. Se observa el aumento de la probabilidad de adoptar una conformación coiled-coil en los últimos 14 residuos de la secuencia de BA40. Sobre el gráfico se muestra la estructura secundaria de BA40. Se muestran las α -hélices como flechas y las hebras- β como cilindros.

Por lo tanto, a modo de corroborar que la proteína realmente se encuentra como monómero en solución, se procedió a medir el radio hidrodinámico de BA40 mediante RMN. Realizando los cálculos teóricos para el radio hidrodinámico de un monómero y un dímero [103], se obtienen los valores de 18,56 Å y 22,7 Å, respectivamente. El valor experimental obtenido usando el experimento de PFG fue de $19,69 \pm 0,22$ Å. De esta forma, se observó que BA40 se comporta como una proteína monomérica en solución. Más aun, para complementar estos datos se realizó la técnica de dispersión estática de luz (SLS- "*Static light Scattering*") sobre BA40, observándose que la proteína se encuentra como monómero en solución, obteniéndose un peso aproximado de 12,5 kDa.

En este contexto, diferentes estudios en las proteínas TnrA y GlnR han mostrado que la región C-terminal de ambas (que contendría la α -hélice H5) no es necesaria para la dimerización, siendo importantes para este evento algunos residuos situados dentro del dominio N-terminal de unión al ADN [99]. Además, en el caso particular de GlnR la región C-terminal inhibe la formación del dímero en solución, uniéndose con baja afinidad al ADN [104].

Este hecho nos motivó a realizar una construcción de la proteína BA40 la cual careciera de la región C-terminal, en dónde se localizaría la α -hélice H5, la cual denominamos BA40- Δ H5. Luego de clonar, expresar y purificar la proteína mutante, se realizó una cromatografía de exclusión molecular obteniéndose nuevamente un solo pico pero en este caso de aproximadamente 11 kDa (ver Materiales y Métodos). Dado que el peso molecular teórico de BA40 Δ H5 es de 10 kDa parecería que la proteína sin esta región C-terminal continúa siendo un monómero en solución. También se realizaron experimentos de radio hidrodinámico por RMN de BA40- Δ H5, obteniéndose valores que se condicen con una proteína en estado monomérico.

El estado oligomérico de BA40 y BA40- Δ H5 también se investigó mediante un ensayo de entrecruzamiento (*crosslinking*) con glutaraldehído. Para esto se incubaron las proteínas a una concentración de 8 μ M exponiéndolas luego al agente de entrecruzamiento. Después del tratamiento, las muestras se analizaron mediante SDS-PAGE. Se detectó una banda muy poco intensa correspondiente al tamaño del dímero de BA40 (Figura 3.7.A). En BA40- Δ H5 no se observó ninguna banda de peso mayor a la proteína misma. Cuando se incubaron las proteínas con ADN se pudo observar la desaparición de BA40 y BA40- Δ H5 debido a la formación de oligómeros que no ingresaron al gel. En este sentido, esto puede ser resultado de que el glutaraldehído como agente de entrecruzamiento tiene la capacidad de fijar, además de las interacciones proteína-proteína, las interacciones proteína-ADN. Sin embargo, BA40 mostró ser más eficiente que BA40- Δ H5 para unir

ADN. Se puede observar en la calle correspondiente a BA40- Δ H5-ADN la presencia de una banda que coincide con el monómero (Figura 3.7.B), en cambio, en la calle BA40-ADN no se observa tinción de proteínas (Figura 3.7.A). Este resultado podría indicar que BA40 es más afín al ADN que la variante carente de la región C-terminal.

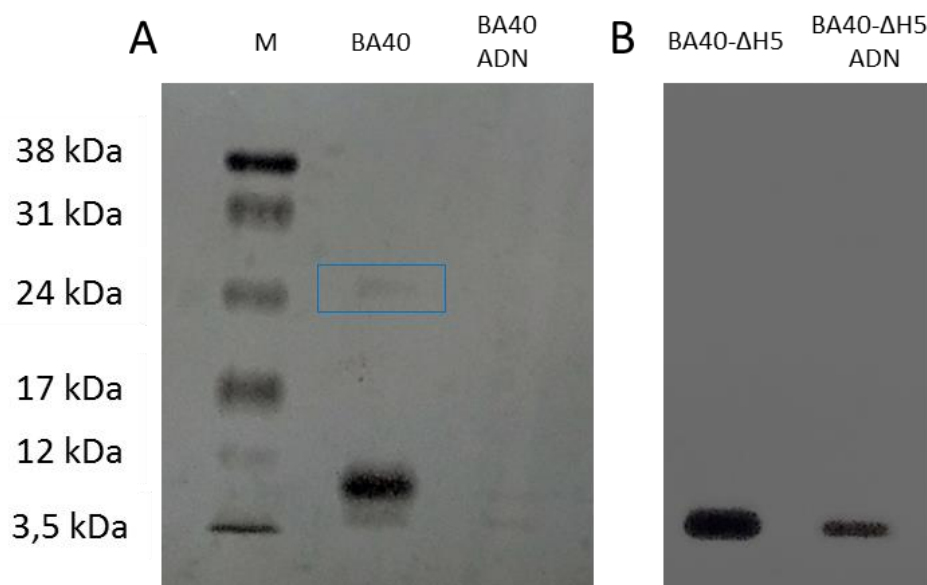


Figura 3.7 Experimentos de entrecruzamiento de las proteínas BA40 y BA40- Δ H5. Las reacciones de entrecruzamiento se separaron mediante SDS-PAGE. **(A)** SDS-PAGE de los ensayos de entrecruzamiento de la proteína BA40. En los diferentes carriles de este gel se muestra el marcador de peso molecular (M), la proteína BA40; BA40 incubada con ADN. **(B)** SDS-PAGE de los ensayos de entrecruzamiento de BA40- Δ H5. En los carriles se muestra la construcción truncada BA40- Δ H5 y BA40- Δ H5 incubada con ADN. Se indica en el recuadro azul la banda correspondiente al dímero de BA40, con un peso aparente de 25 kDa.

La presencia de una banda correspondiente al dímero de BA40 y la ausencia de una banda similar para BA40- Δ H5 en los ensayos de entrecruzamiento sugieren que la región C-terminal sería importante para promover la dimerización de la proteína. Sin embargo, se observa que la formación del dímero está muy poco favorecida en las condiciones ensayadas (el equilibrio se encuentra casi completamente desplazado hacia el monómero). Por este motivo, estos resultados son consistentes con los obtenidos previamente por RMN, SLS y SEC y confirman que BA40 se encuentra mayoritariamente como monómero en solución.

Análisis de BA40 y BA40-ΔH5 mediante RMN

Con el fin de estudiar a BA40 y BA40-ΔH5 a nivel atómico, se comenzó un análisis mediante experimentos de RMN. Con el objetivo de conocer si las proteínas eran buenas candidatas para su estudio por RMN se realizaron experimentos de HSQC ^1H - ^{15}N . En una primera inspección del espectro de BA40 se observó claramente que la proteína es una muy buena candidata para estudiar mediante esta técnica. Sin embargo, el número de picos esperados (106 HN, sin contar los NH_2 de cadenas laterales y el del indol del Trp) fue menor al observado en los espectros de BA40. En total, no se observan 17 resonancias HN, indicando que algunas regiones de la proteína tendrían mayor intercambio con el solvente o presentarían mayor flexibilidad en la escala de movimientos lentos o intermedios (μs - ms). Además, existe un conjunto de picos en el espectro de BA40 que presentan poca dispersión de CS denotando pertenecer a una región no estructurada. Una superposición de los espectros HSCQ de BA40 y BA40-ΔH5 muestra que estos son muy similares a excepción del centro del espectro, donde se observa la desaparición en el espectro de BA40-ΔH5 de aquellos picos que no presentaban dispersión en BA40 (Figura 3.8). Interesantemente, este resultado sugiere que dichos picos pertenecerían a un grupo de residuos que integran la región predicha como α -hélice H5, planteando la hipótesis de que dicha región C-terminal se encuentre parcialmente desestructurada, dificultándose la formación de una estructura coiled-coil. Por otro lado, también desaparece un conjunto de picos con mayor dispersión de CS relativa a los residuos antes mencionados, sugiriendo que otra porción de la región C-terminal podría tener algún tipo de estructura. Sin embargo, más estudios serán necesarios para dilucidar la conformación de esta región C-terminal en solución.

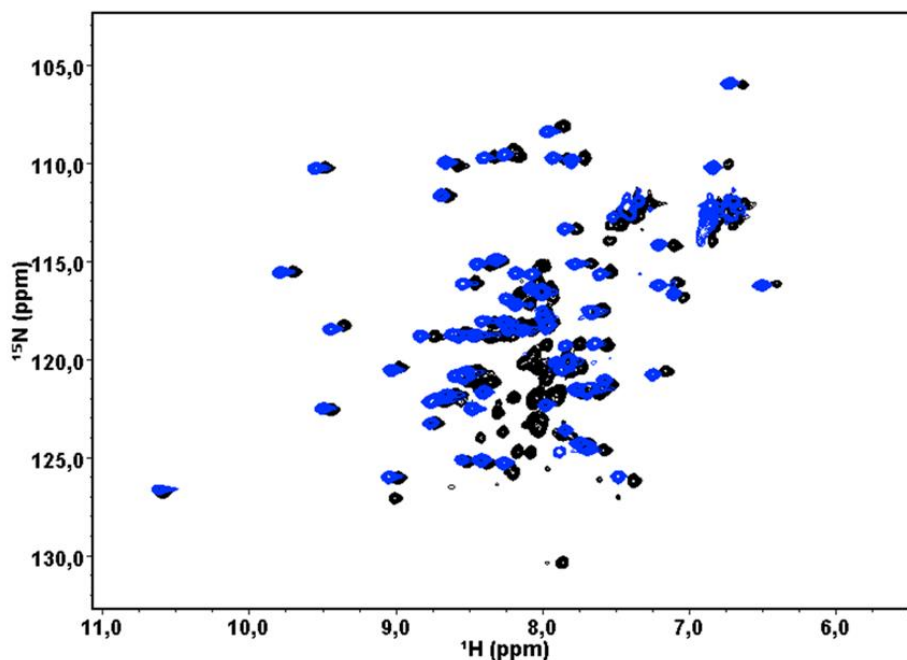


Figura 3.8. Superposición de los espectros HSQC ^1H - ^{15}N de las proteínas BA40 y BA40- ΔH5 . El espectro HSQC ^1H - ^{15}N de BA40 se muestra en negro y el de BA40- ΔH5 en azul. En ambos espectros se observan picos homogéneamente intensos y con buena dispersión de CS, aunque se denota un grupo de picos con muy baja dispersión en el centro del espectro de BA40. Interesantemente, dicha región desaparece en el espectro HSQC de BA40- ΔH5 indicando que dichos picos pertenecerían a la región predicha como la α -hélice H5. Los espectros se encuentran ligeramente desplazados para una mejor observación de las diferencias entre ambos.

Elucidación del sitio de unión específico y caracterización del tipo de regulación implicada

Con la finalidad de poder elucidar una secuencia consenso de ADN a la cual BA40 se uniera de forma específica se utilizó un método de selección de alta afinidad a partir de secuencias aleatorias de ADN (SELEX) [105]. La técnica SELEX (del inglés, Systematic Evolution of Ligands by Exponential enrichment), tiene como fin original identificar secuencias o estructuras de ácidos nucleicos, denominadas aptámeros, que se unen firmemente a las moléculas diana [105]. El procedimiento imita la evolución natural empleando un proceso iterativo de selección y amplificación de secuencias de una biblioteca aleatoria de ácidos nucleicos. La biblioteca se crea normalmente mediante síntesis química y cada secuencia tiene una región aleatoria central, que define el alcance y la complejidad de la biblioteca, y dos regiones constantes flanqueantes para la

amplificación por PCR. Los resultados de SELEX generalmente muestran secuencias que poseen motivos comunes, por lo que siendo aplicado sobre bibliotecas de ADN doble cadena resulta particularmente útil para definir secuencias consenso de unión al ADN de proteínas como los factores de transcripción.

Particularmente, nuestra metodología consistió en inmovilizar la proteína recombinante BA40 en una columna de Ni-NTA y realizar incubaciones con una biblioteca de oligonucleótidos con una región central aleatoria de 18 pares de bases. Las secuencias de menor afinidad fueron separadas de la proteína y la columna mediante lavados con un gradiente escalonado de creciente fuerza iónica. Finalmente, las secuencias que quedaron retenidas por la proteína se eluyeron y fueron sometidas a una ronda de amplificación por PCR, para comenzar un nuevo ciclo de maduración de la afinidad (Figura 3.9). En la Figura 3.10 se puede observar la primera ronda de selección. La elución a una concentración de 300 y 400 mM de imidazol fue utilizada para la posterior amplificación de las secuencias y realizar la siguiente ronda de selección. Este conjunto de pasos se repitieron durante seis ciclos. Finalmente, los oligonucleótidos seleccionados se purificaron y clonaron en un vector pGEM®-T Easy (Promega), transformándose luego células competentes XL1-Blue (Stratagene).

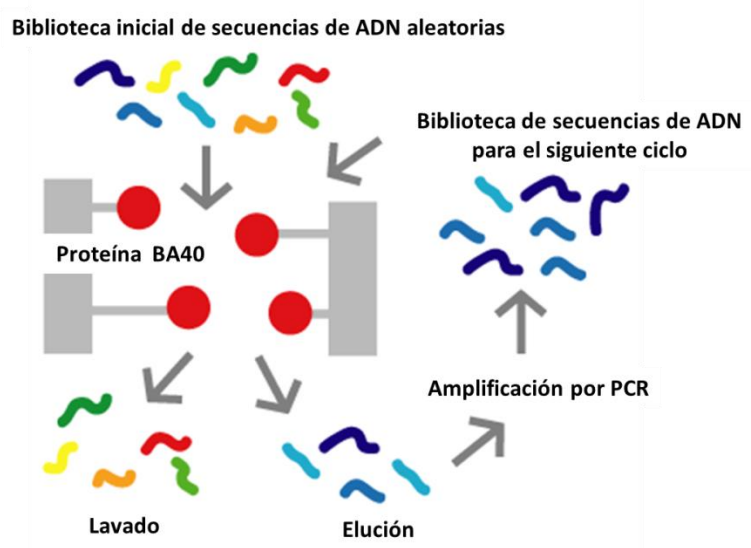


Figura 3.9. SELEX en columna de afinidad. Se parte de una biblioteca de secuencias aleatorias de ADN, las cuales se incuban con la proteína inmovilizada en la columna de afinidad. Las secuencias de menor afinidad son separadas de la proteína mediante lavados. Finalmente, las secuencias que quedan retenidas por la proteína se eluyen y son sometidas a una ronda de amplificación por PCR, para comenzar un nuevo ciclo de maduración de la afinidad.

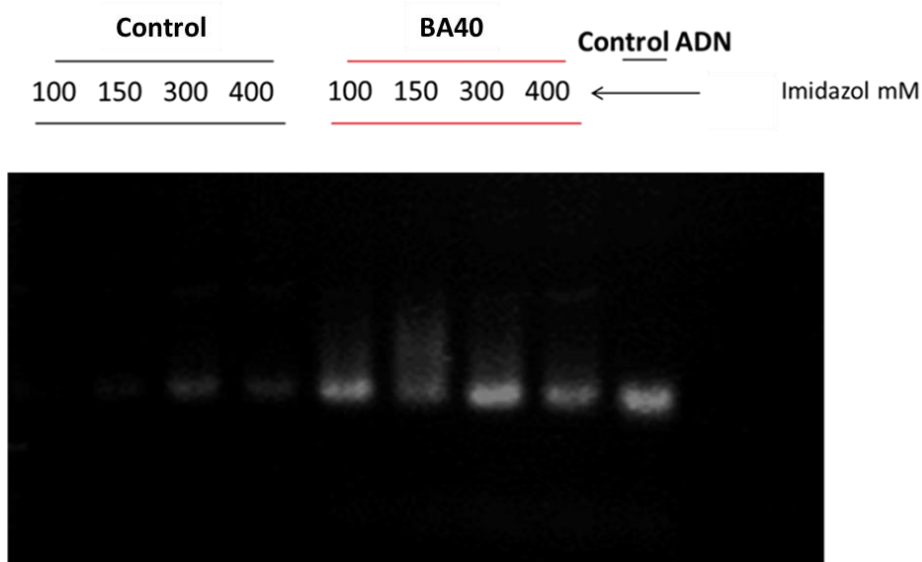


Figura 3.10. Gel de agarosa mostrando la amplificación por PCR de la primera ronda de selección por el método SELEX. En primera instancia se inmovilizó la proteína a la resina de Ni-NTA para luego incubarla con la biblioteca de oligonucleótidos aleatorios. En el control no se utilizó proteína. Luego de una incubación de 30 minutos se realizaron diferentes lavados con concentraciones crecientes de Imidazol manteniendo la concentración de NaCl constante a 500 mM. Finalmente, se realizaron PCRs de las diferentes fracciones obtenidas. En el carril señalado como Control ADN se muestra la PCR de la biblioteca sin haberla sometido al método de SELEX. Para la segunda ronda de selección se utilizaron las muestras obtenidas a 300 y 400 mM de imidazol.

De esta forma, se secuenciaron 30 colonias y las secuencias de ADN seleccionadas fueron sometidas a la búsqueda de motivos consenso utilizando el programa MEME suite [106]. Con esta metodología de selección *in vitro* con la proteína purificada e inmovilizada en la matriz de afinidad obtuvimos una secuencia consenso proveniente de la interacción del monómero de BA40 con el ADN. Cabe aclarar que en este caso no se obtiene una secuencia palindrómica, como se describe para este tipo de factores de transcripción, dado el diseño del experimento. Como resultado se obtuvo un motivo consenso de 8 pares de bases el cual se presentó en 22 de las 30 secuencias ingresadas en el programa MEME (Figura 3.11).



Figura 3.11. Motivo de ADN consenso de unión de la proteína BA40. Motivo de ADN de 8 pb generado por el programa MEME con 30 secuencias de ADN que fueron obtenidas mediante el método de SELEX.

Evaluación de la secuencia consenso obtenida por SELEX

Con el fin de validar la secuencia consenso de 8 pb obtenida mediante la técnica de SELEX decidimos comparar la afinidad de BA40 por esta secuencia con respecto a una secuencia aleatoria del mismo tamaño. Para esto se utilizó la técnica de espectroscopía de fluorescencia, y se llevaron a cabo ensayos de titulación utilizando oligonucleótidos conteniendo el motivo SELEX y de secuencia aleatoria unidos covalentemente a fluoresceína (6-FAM). Se determinó el cambio en el valor de la anisotropía de fluorescencia de la molécula acoplada a 6-FAM luego de la unión de BA40. Una vez determinada la estequiometría del complejo se procedió a calcular la constante de disociación al equilibrio (K_D). Este parámetro común y frecuentemente fácil de calcular en condiciones *in vitro* es útil para abordar numerosas preguntas al examinar una reacción de unión específica entre moléculas. Sin embargo, hay que tener presente que las condiciones de reacción *in vitro* pueden estar muy alejadas de la ambiente intracelular, con lo que se debe tener precaución al analizar los valores de dichos parámetros. Para calcular la K_D es necesario realizar las curvas en condiciones disociantes, esto es, a concentraciones entre 1/5 y 5 veces el valor de la K_D , por lo que los ensayos se realizaron a concentraciones de ADN de 50 a 500 nM, y la temperatura utiliza en todos los experimentos fue 20 °C.

Las curvas de titulación mostraron el mejor ajuste a una sigmoidea alostérica a partir de las cuales se calcularon los valores de K_D . De esta forma, se obtuvo una K_D de 90 nM para la secuencia con el motivo obtenido mediante SELEX y de 1,9 μ M para una secuencia aleatoria de la misma longitud (Figuras 3.12 y 3.13). Como conclusión, es interesante mencionar que mediante esta técnica se consiguió obtener una secuencia de ADN por la cual BA40 es 21 veces más afín que para una secuencia aleatoria.

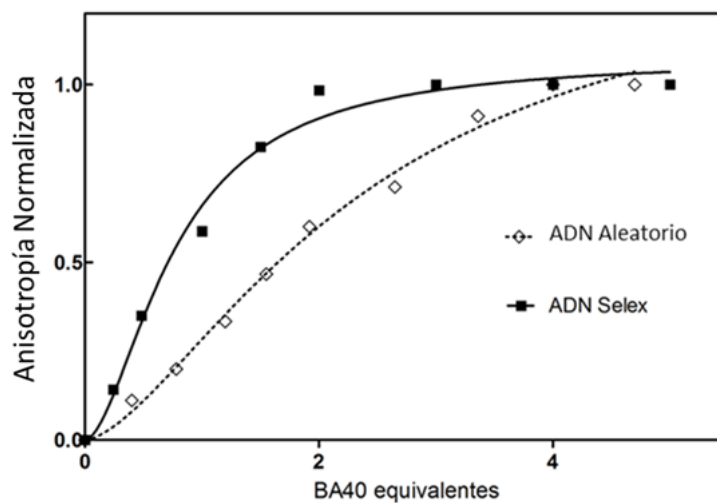


Figura 3.12. Cambio en la anisotropía de fluorescencia normalizada de 6-FAM-SELEX y 6-FAM-aleatorio frente a la adición de concentraciones crecientes de BA40. Cambio en la anisotropía de fluorescencia de 500 nM 6-FAM-SELEX (cuadrados negros) y 500 nM 6-FAM-aleatorio (cuadrados vacíos). La línea muestra el ajuste a un curva sigmoidea alostérica.

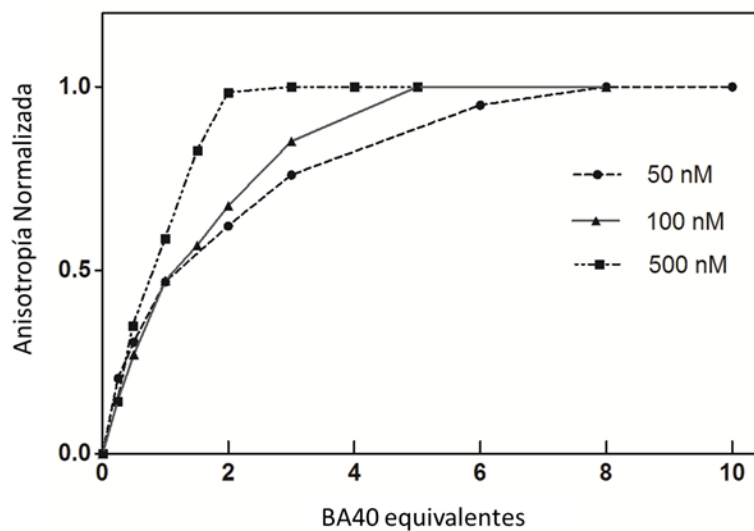


Figura 3.13. Cambio en la anisotropía de fluorescencia normalizada de 6-FAM-SELEX y cálculo de la K_D asociada. Se registró el cambio de anisotropía de 6-FAM-SELEX en concentraciones de 50, 100 y 500 nM frente a la adición de concentraciones crecientes de BA40. La línea muestra el ajuste a un curva sigmoidea alostérica. a partir de la cual se obtuvo la constante de disociación al equilibrio del complejo [6-FAM-SELEX:BA40] : $K_D = 90 \pm 10$ nM.

La región C-terminal de BA40 incrementa la afinidad por el ADN

Para averiguar si la región C-terminal tiene alguna implicancia en la actividad de unión al ADN de BA40, llevamos a cabo ensayos de titulación utilizando el oligonucleótido con el motivo SELEX unido covalentemente a fluoresceína (6-FAM). Los experimentos se realizaron agregando concentraciones crecientes de BA40 o BA40- Δ H5 y midiendo el cambio en la anisotropía de fluorescencia del oligo acoplado a 6-FAM. Los resultados obtenidos mostraron que la eliminación de una porción de la región C-terminal tiene un efecto negativo en la actividad de unión al ADN (Figura 3.14). Comparando los valores de las constantes de disociación de BA40 y BA40- Δ H5 se observó un aumento de 35 veces para BA40- Δ H5 con respecto a BA40, obteniéndose un valor de K_D de 3,2 μ M para la proteína sin la región C-terminal. Este valor, además de ser mucho mayor que el obtenido para el motivo SELEX, también está incrementado con respecto al calculado para una secuencia aleatoria inespecífica.

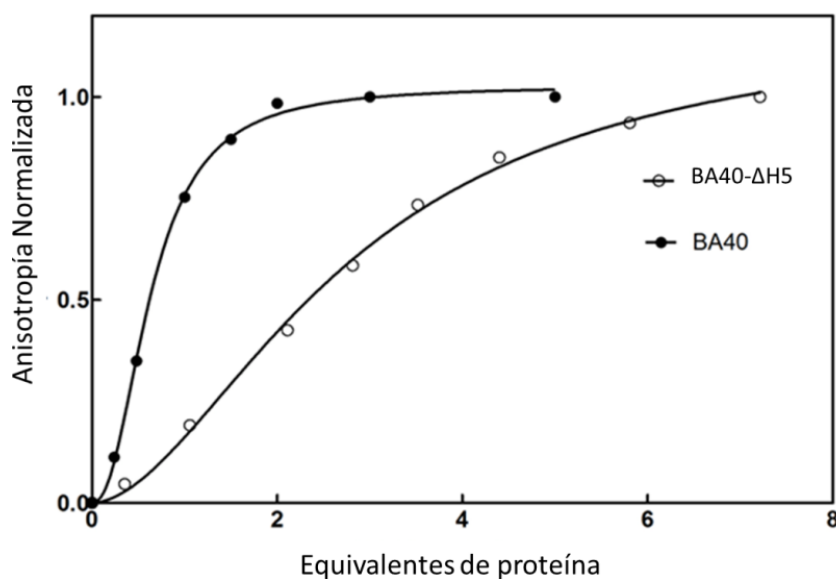


Figura 3.14. Cambio en la anisotropía de fluorescencia normalizada de 6-FAM-SELEX luego del agregado de BA40 y BA40- Δ H5 Se registró el cambio de anisotropía de 6-FAM-SELEX a una concentración de 500 nM de 6-FAM-SELEX frente a la adición de concentraciones crecientes de BA40 y BA40- Δ H5. La línea muestra el ajuste a un curva sigmoidea alostérica.

Estudio in vivo de BA40

El siguiente objetivo fue comenzar a obtener información sobre rol biológico de la proteína BA40. De esta forma, la primera propuesta fue estudiar la expresión del factor de transcripción a nivel del ARN mensajero analizando diferentes condiciones de crecimiento de la bacteria *B. argentinensis*. En este contexto, se realizaron PCRs semicuantitativas a partir de preparaciones de cDNA de la bacteria expuesta a diferentes tratamientos. Estas preparaciones se adquirieron luego de crecer la bacteria en condiciones óptimas (medio rico en nutrientes, 20 °C) para posteriormente someterla a diferentes situaciones de estrés en forma de *shock*. Las diferentes condiciones de estrés incluyeron: medio desprovisto de nutrientes, altas y bajas temperaturas, alta y baja salinidad. Como resultado se observó un aumento del mensajero de BA40 luego de una hora de haber expuesto a la bacteria a un medio desprovisto de nutrientes (oligotrófico) (Figura 3.15). Más aun, luego de transcurridas siete horas en este medio se observó una disminución del ARNm volviendo a presentar los valores iniciales en condiciones de crecimiento óptimas (Figura 3.15).

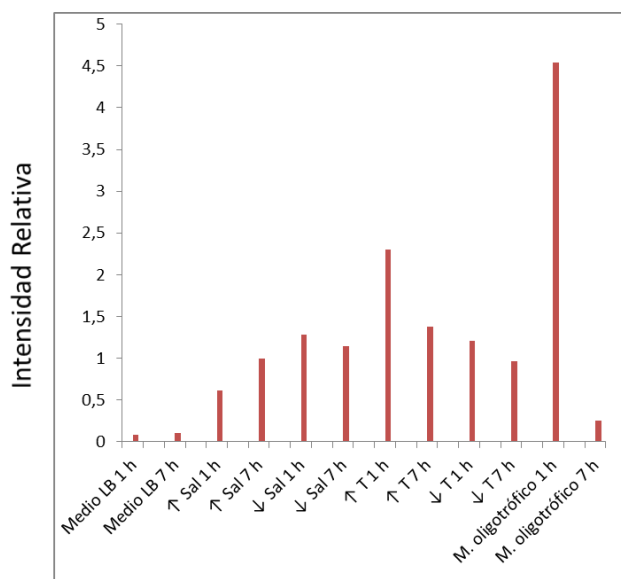


Figura 3.15. PCRs semi-cuantitativas a partir de cDNA de *B. argentinensis* sometida a diferentes tratamientos. Se observó un aumento en la expresión de BA40 luego de una hora de haber expuesto a la bacteria a un medio desprovisto de nutrientes. Los valores de ARNm fueron comparables a los iniciales (Medio LB 7 h) luego de 7 horas de exposición al medio oligotrófico. Las intensidades fueron normalizadas al cDNA total.

Con el fin de corroborar esta observación, se analizó la expresión de la proteína BA40 mediante Western Blot. Para ello, se produjeron anticuerpos policlonales contra la proteína BA40 los cuales resultaron ser altamente afines y específicos. En primera instancia, se realizó el Western blot sobre dos muestras, un cultivo crecido en medio rico y uno obtenido luego de una hora de exposición a un medio oligotrófico. Los resultados estuvieron en línea con la información previa mostrando un aumento de la expresión de BA40 luego de la exposición al medio oligotrófico (Figura 3.16).

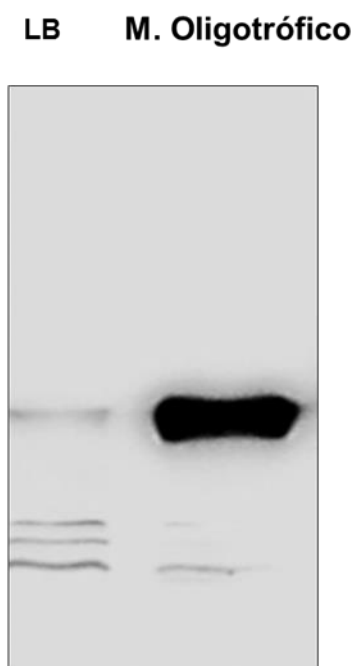


Figura 3.16. Análisis de la expresión de la proteína BA40 mediante Western Blot en *B. argentinensis* en condiciones normales de nutrientes (LB marino) y luego de 30 minutos de exposición a un medio oligotrófico.

Posteriormente se estudió el perfil temporal de expresión de la proteína en dicho medio. De esta forma, se realizó un Western blot de muestras obtenidas a diferentes tiempos de exposición al medio oligotrófico. Interesantemente, se reveló que la expresión de la proteína BA40 fue máxima a los 5-15 minutos luego de transferir a la bacteria desde un medio rico a un medio oligotrófico (Figura 3.17).

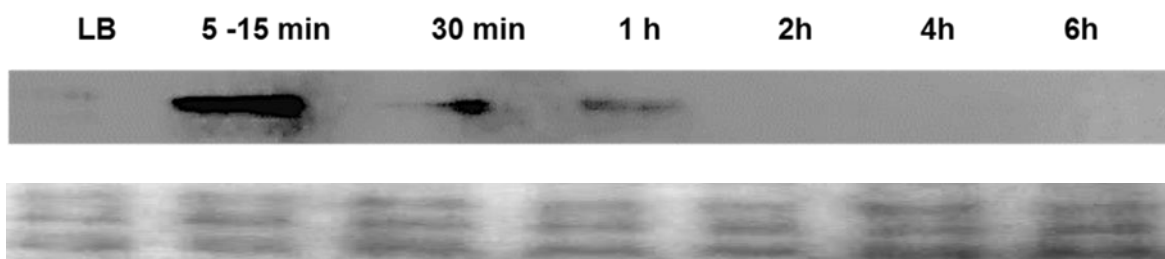


Figura 3.17. Análisis de la expresión de la proteína BA40 en la bacteria en diferentes tiempos de exposición a un medio oligotrófico mediante Western Blot. Se observa la máxima expresión a los 5-15 minutos de privar a la bacteria de nutrientes. En la parte inferior se observa el control de carga del gel.

Estos resultados indicaron que BA40 se encontraría sujeta a una estricta regulación en respuesta a señales del medioambiente, por lo cual podría actuar como una proteína de shock relacionada al estrés por exposición a un medio oligotrófico. En este sentido, existe aún muy poca información sobre los mecanismos moleculares que dirigen los sistemas de monitoreo frente a la presencia o disminución de nutrientes asociados comúnmente al ciclo de vida de bacterias del género *Bizionia* y en general de organismos de ambientes antárticos. Por lo tanto, este hallazgo es de gran relevancia y abre una nueva perspectiva para abordar la comprensión de los mismos.

Discusión

Los estudios previos habían revelado que la fosfatasa BA42 pertenecía a un operón compuesto por cinco proteínas, una de ellas, la proteína BA40, fue clasificada como un probable factor de transcripción ya que presentó una alta identidad de secuencia con miembros de la familia MlrA. Con el objetivo de comenzar a estudiar el rol biológico del operón se propuso iniciar un estudio estructural y bioquímico de la proteína BA40. De esta forma, en base a los resultados obtenidos se evidencia que BA40 presenta características consistentes con su rol predicho como factor de transcripción dado que la proteína mostró tener actividad de unión al ADN. En este contexto, con la finalidad de poder elucidar una secuencia consenso de ADN a la cual BA40 se uniera de forma específica, se utilizó un método de selección de alta afinidad a partir de secuencias aleatorias de ADN (SELEX). Mediante esta técnica pudimos seleccionar una secuencia consenso de 8 pb por la cual BA40 mostró tener 21 veces más afinidad que hacia una secuencia aleatoria de la misma longitud. En particular, BA40 presentó una K_D de 90 nM para la secuencia consenso. Los factores de transcripción pertenecientes a la familia MerR son proteínas que se unen a secuencias palindrómicas separadas por una región variable de pares de base. En este

sentido, el oligonucleótido utilizado en los ensayos de anisotropía de fluorescencia no está constituido por un palíndromo ya que se desconoce la separación entre los motivos del mismo. Así mismo, la secuencia obtenida por SELEX puede diferir de la secuencia biológica específica. Estos hechos nos hacen considerar que la afinidad de BA40 por su secuencia blanco podría ser aún mayor de la obtenida en los ensayos descriptos. Además, si bien la mayoría de los factores de transcripción que se unen específicamente a pocos sitios en el genoma tienen valores de K_D en el intervalo de 0,1 pM-10 nM, aquellos que se unen a un alto número de sitios regulatorios presentan una menor especificidad de secuencia y por lo tanto valores de K_D un poco más elevados [107], comparables con el valor de K_D obtenido para BA40 y la secuencia consenso. Tampoco podemos descartar la existencia de un probable modulador positivo de su actividad y que la ausencia del mismo en la reacción tuviese como resultado la obtención de valores de K_D elevados de BA40 frente a su secuencia consenso específica. Con respecto a lo dicho, diferentes estudios mostraron que la proteína tipo MerR denominada GlnR presenta una K_D de 77 nM hacia su secuencia promotora *glnRA* en ausencia de su efector o modulador. Sin embargo, cuando la glutamina y la glutamina sintetasa (efectores) están presentes en conjunto en la reacción de unión, se observa un aumento de 32 veces en la afinidad de unión al ADN de GlnR [98]. En este contexto, la K_D observada para la secuencia del SELEX muestra que este motivo podría ser una secuencia cercana a la biológica y la falta del efector necesario para BA40 generaría este valor de K_D .

Como se mencionó anteriormente, la mayoría de los miembros de la familia MerR se comportan como dímeros en solución y de la misma forma en el complejo ADN-Proteína. Se ha mostrado que en muchos de los casos la región C-terminal, más concretamente la α -hélice H5 sería fundamental para que ocurra el evento de dimerización. Nuestros estudios muestran que si bien la predicción bioinformática de estructura secundaria concluye que BA40 posee la α -hélice H5, la proteína se encuentra como monómero en solución y que su región C-terminal se encontraría parcialmente desestructurada careciendo total o parcialmente de la α -hélice H5. En este contexto, diferentes trabajos han observado un rol diferencial de la α -hélice H5. Ejemplos de esto son algunos factores de transcripción MerR en los cuales la región C-terminal que contiene la hélice H5 inhibe la formación del dímero en solución [104], o no es necesaria para la dimerización siendo importantes para este evento diferentes residuos situados dentro del dominio N-terminal de unión al ADN [99, 108]. Para investigar estas posibilidades se analizó en primera instancia el estado multimérico de la construcción sin la región predicha como la α -hélice H5, BA40- Δ H5. Los

resultados obtenidos muestran que la proteína truncada también se encuentra como monómero en solución. Sin embargo, al estudiar la unión de BA40- Δ H5 con el ADN fue interesante notar que presentó una menor afinidad por la secuencia consenso generado por SELEX, mostrando una K_D de 3,2 μ M, 35 veces menos afín que la proteína salvaje. Estos resultados están en línea con los descriptos previamente en los experimentos de entrecruzamiento e indican que la región C-terminal de BA40 podría participar en la estabilización de la unión con el ADN. En este punto, el aumento de afinidad podría ser consecuencia de: i) la dimerización de BA40 cuando interacciona con el ADN o ii) un cambio conformacional en BA40- Δ H5 que afecta el sitio de unión al ADN, ocasionando una disminución de la afinidad.

Una posible explicación para el primer caso podría ser que tras la unión de BA40 al ADN la región C-terminal logre estructurarse generando la α -hélice H5, favoreciendo la formación del dímero de BA40 y aumentando la estabilidad del complejo ADN-Proteína. En base a esta hipótesis, se puede suponer que esta región podría contener un posible rol regulatorio de la actividad de unión al ADN de BA40, tal como se ha reportado en los trabajos antes mencionados [97, 104].

La carencia de un dominio C-terminal específico (adicional a la región de dimerización), como se ha descrito para proteínas que poseen la misma arquitectura que BA40, plantea la posibilidad de que el modulador de la actividad sea una proteína [97, 98]. Un posible mecanismo de regulación podría involucrar la estructuración de la región C-terminal mediante la interacción con una proteína accesoria. Dado que normalmente los ligandos que interaccionan con miembros de la familia MerR están relacionados con el papel de los genes cuya expresión es regulada [84, 98, 109], la búsqueda futura de proteínas que interactúen con el factor de transcripción BA40 será muy importante para develar su función regulatoria.

Finalmente, los estudios de la expresión de BA40 en *B. argentinensis* mostraron que la proteína presenta un aumento de su expresión luego de someter la bacteria a un medio desprovisto de nutrientes. Más aun, los resultados obtenidos mostraron que la expresión fue máxima entre los 5 a 15 minutos luego de la exposición a dicho medio oligotrófico. De esta forma, se puede concluir que la proteína BA40 se encontraría sujeta a una estricta regulación respondiendo a señales del medioambiente, por lo cual podría actuar como una proteína de shock relacionada al estrés por exposición a un medio oligotrófico. En este sentido, se ha reportado para muchas especies bacterianas que en respuesta a diversas presiones ambientales se expresan conjuntos de genes que inician la transición de un estilo de vida planctónico a un comportamiento multicelular que ofrece distintas ventajas en la tolerancia al estrés, la adquisición de nutrientes y la

supervivencia [110]. Como mencionamos en la introducción, una de las principales características del filo Bacteroidetes es su rol como degradador y su preferencia por crecer unido a superficies de las cuales obtener nutrientes. En este contexto, el pasaje a una situación estresante como puede ser un estilo de vida libre en el medio marino debe encender la expresión de múltiples proteínas para poder perpetuarse en este ambiente hostil hasta que se consiga colonizar otra superficie. Interesantemente, un análisis bioinformático preliminar para detectar regiones homólogas a la secuencia consenso obtenida por SELEX en el genoma de *B. argentinensis* mostró que dicha secuencia presenta homología con regiones promotoras de muchas proteínas relacionadas con el metabolismo energético, gluconeogénesis, síntesis de proteínas, adherencia a superficies, entre otras. Estos datos, resultan muy interesantes a fin de continuar con el estudio de BA40 y la regulación río abajo que esta pueda desencadenar. En el futuro, con el fin de identificar de manera global los blancos genéticos en el genoma de *B. argentinensis* de este nuevo factor de la subfamilia MlrA, se propone utilizar unas de las técnicas más importantes en la actualidad y que ha demostrado ser exitosa en un gran número de estudios de este tipo, el ChIP-seq, buscando obtener mediante esta técnica información valiosa que contribuirá a la comprensión de la adaptación de *B. argentinensis* al ambiente marino.

Discusión general

Un reto importante en la biología es encontrar estrategias para poder determinar de forma confiable y rápida las funciones moleculares y celulares de las proteínas. De esta forma, uno de los caminos que conducen a la asignación de la función molecular de proteínas aún no caracterizadas es la determinación de su estructura tridimensional y el posterior estudio de la misma. En este contexto, el proyecto de genómica estructural de la bacteria *Bizionia argentinensis* descrito en esta tesis, tuvo como principal objetivo determinar la estructura de proteínas de función desconocida, con el fin de obtener información sobre el plegamiento y función de éstas. Los resultados muestran que a través de esta metodología se consiguió identificar nuevas asociaciones secuencia-estructura, lo que permitió acceder en última instancia a la función molecular de diferentes proteínas, abriéndonos un camino hacia el estudio de su rol biológico. Esto es una confirmación de la importancia que pueden tener estos proyectos cuando están dirigidos a proteínas “hipotéticas conservadas” y dominios de función desconocida (DUF), ya que tales descubrimientos aportan información a una fracción significativa del universo inexplorado de proteínas. Si bien eventualmente esta clase de estrategia para el estudio de las proteínas puede no tener éxito, cuando se obtienen resultados positivos, se pueden generar líneas de estudio interesantes que de otra forma no hubieran sido desarrolladas. En nuestro caso particular, BA42 es el primer miembro de la familia Pfam PF04536 que comprende un dominio TPM independiente que ha sido estructural y bioquímicamente caracterizado hasta la fecha, revelándose en la estructura de BA42 una nueva variante topológica. De esta forma, queda en evidencia que la genómica estructural es mucho más sensible que la comparación de secuencias primarias para la identificación de una función, aportando diferentes niveles de información. También, en el caso particular de organismos pertenecientes a familias poco estudiadas, como lo es *B. argentinensis*, la biología estructural combinada con técnicas de genómica y proteómica brinda una muy buena estrategia para poder descubrir la función molecular de diferentes proteínas y así comenzar a comprender la función celular en diferentes procesos biológicos inexplorados. En este sentido, a través del estudio genómico estructural de *B. argentinensis* se logró identificar y comenzar a caracterizar un nuevo factor de transcripción al cual denominamos BA40, asociado a la respuesta a condiciones oligotróficas y perteneciente a la familia de factores transcripcionales MerR.

Los resultados obtenidos en esta tesis muestran el papel fundamental que puede ocupar la genómica estructural en estudios funcionales de ciencia básica. Así mismo, sientan las bases para continuar con el análisis de la función de BA42 y del rol regulatorio del factor de transcripción BA40, los cuales sin dudas aportarán novedosa información sobre la respuesta de la bacteria frente

a condiciones de bajos nutrientes y contribuirá al entendimiento de los mecanismos moleculares que regulan el ciclo biológico de este grupo de organismos marinos.

Finalmente, la realización de este proyecto de genómica estructural contribuyó al desarrollo de nuevas líneas de investigación destinadas a obtener información sobre este grupo de organismos marinos, sus ciclos biológicos, las interacciones con el ecosistema y los mecanismos moleculares que los regulan.

Materiales y métodos

Clonado Molecular

Todos los genes en esta tesis se amplificaron por PCR utilizando como molde ADN cromosómico de la cepa JUB59 de *B. argentinensis*. Los primers utilizados para la amplificación de los ORFs seleccionados se anexan en la tabla M.1. Las amplificaciones se realizaron en dos rondas de PCR, obteniendo los genes flanqueados por los respectivos sitios de recombinación. Los productos de PCR se recombinaron en el plásmido pDONR-201 y posteriormente en el plásmido de expresión pDEST-527 (Invitrogen). Las construcciones finales contuvieron en el extremo N-terminal un sitio de corte específico para la proteasa TEV seguido de una región de 6X-histidinas para su purificación por afinidad. Todos los diferentes plásmidos recombinados fueron sometidos a secuenciación a fin de analizar la calidad de los mismos. Finalmente, se transformaron células competentes de *E. coli* BL21 (DE3) (Stratagene, La Jolla, CA) con el plásmido de expresión recombinado utilizando el método de shock térmico [111].

ORF ID	Primer 5'	Primer 3'
BA_3_15	GAGAACCTGTACTTT CAGGGTATGAAATACGCAAGATT AAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAACTAATGAGTTGATGTATG
BA_5_34	GAGAACCTGTACTTT CAGGGTATGAAAAAAGAAATGATAATTTAA AAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAAATC TATTTCTGCTTCAGC
BA_6_46	GAGAACCTGTACTTT CAGGGTATGGCTAGAGACGAAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATTTATACTCAGCTTCC
BA_8_56	GAGAACCTGTACTTT CAGGGTATGATTTTCGATAAAGAGAAA AAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAGAAACCTAAATCGCG
BA_13_18	GAGAACCTGTACTTT CAGGGTATGTTGATAGATCATTAGAAATACAA TTC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAACTCTTTTCTACTCTATTT
BA_27_18	GAGAACCTGTACTTT CAGGGTATGAAAAACTCCGACTTATTTTAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATAAATTAACCGTTTTAAATCCCAATTTTC
BA_30_17	GAGAACCTGTACTTT CAGGGTATGTTTAAAGTAGGAGACG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAAAGTACTCACATTTTGG
BA_31_44	GAGAACCTGTACTTT CAGGGTATGTTTTCAGAAAAAGCCAAATAAAATATTT C	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTTTTGAGAACGC AATACTGG
BA_47_29	GAGAACCTGTACTTT CAGGGTATGGTGGAAAAGGAAATGAAAGC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAGCCATTATCTTTGGTTTTATTTAC
BA_53_45	GAGAACCTGTACTTT CAGGGTATGAGTTT AGAAAAAGTATTATCAATAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTCTCTCTCTCAGTAGC
BA_212_10	GAGAACCTGTACTTT CAGGGTATGAAATAAACAAAACTTTTTACAAAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATCC TATATATAATGTATGCTAATG
BA_300_2	GAGAACCTGTACTTT CAGGGTATGAAAAACAACGGAAATAATAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTATTTTAGAGTAAATCTGATGTG
BA_2_7	GAGAACCTGTACTTT CAGGGTATGAAGCTATCAGAAGTAAAAAAAAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAAACACGCTACC GCC
BA_5_69	GAGAACCTGTACTTT CAGGGTATGTTGGAGGAAGAAGATTTAAAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTTCTGCTGTTTAAATG
BA_6_27	GAGAACCTGTACTTT CAGGGTATGCTTTGGCCACATTATGC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATAAATAATCAGTAACAACTTATG
BA_7_45	GAGAACCTGTACTTT CAGGGTATGAAAAATTTACTTGC TTTTACTGC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTCTCGTTTCTTAAAGTCC
BA_40_17	GAGAACCTGTACTTT CAGGGTATGGATAATACAATTAT AATTGTAGGGG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTTCTAGTTC CAATACGTTG
BA_43_2	GAGAACCTGTACTTT CAGGGTATGACGGGCCATGGG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTAGTTTTCAA AACGTATTTATAAT
BA_58_6	GAGAACCTGTACTTT CAGGGTATGAAAAAGCTTTACGTTATC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACTAATC GACCTGAATAATAACG
BA_64_66	GAGAACCTGTACTTT CAGGGTATGCAACCGAGTTTCTG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATGC CATAGTAAGTTTAAACAAC
BA_80_14	GAGAACCTGTACTTT CAGGGTATGTACGAAAAAACC TATCCAAATAAAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATAGTTTGATTTCTCCGCATAAAC
BA_123_14	GAGAACCTGTACTTT CAGGGTATGTTTCGCTTTTACTATGC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATTTAAAATTAACATTCCTTTATC
BA_119_21	GAGAACCTGTACTTT CAGGGTATGGCAGAACACAAACG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAAAAATGATAAAACGCATTTCTC
BA_121_2	GAGAACCTGTACTTT CAGGGTATGGTGAATATTACATCGTATTGGG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATTTCAAATAACTCAGTTATTGTAG
BA_42_3	GAGAACCTGTACTTT CAGGGTATGCTAAAATAGAAGAGTTTTTAACAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTATCC TTTTGAAAATTTGTTGGTAATTC
BA_101_99	GAGAACCTGTACTTT CAGGGTATGAAAAATGAAATAATCAATAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTACTATTTCTTTTAAATAAAGGTAC
BA_75_21	GAGAACCTGTACTTT CAGGGTATGTTTCAATTAGGAAAAACCATAG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTACTATTGAAATCTTTTCAGCCAC
BA_150_2	GAGAACCTGTACTTT CAGGGTATGACCGTTATTGATCCGAATGC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATAAATAATTTTTCACCATTTAAATAG
BA_169_10	GAGAACCTGTACTTT CAGGGTATGAACACACAAAACCAAACGTTG	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTAC AAGCTGAAAAGTTAAGTG
BA_237_2	GAGAACCTGTACTTT CAGGGTATGCAATCGAATCAGAAAAC	GGGGACC ACTTTGTACAAGAAAGCTGGGTTATTACCACCAACAAATAAAG

Tabla M.1. Primers utilizados para la amplificación de los ORFs seleccionados mediante el filtro bioinformático.

Construcción de mutantes

Las mutantes puntuales se obtuvieron por el método del “megaprimer” [112]. Las amplificaciones de las mutantes puntuales de BA42 y las proteínas truncadas de BA42 Δ CT y BA40 Δ H5 se obtuvieron utilizando el mismo método de clonado descrito anteriormente.

Reacciones de PCR

Las reacciones de PCR se realizaron utilizando la ADN polimerasa Pfu en un termociclador (Perkin Elmer), utilizando el molde correspondiente, 1 μ M de primer 5', 1 μ M de primer 3', 0,2 mM de cada dNTPs, 1mM Mg²⁺, 0,05 U/ μ l de Pfu y el buffer de la misma polimerasa. Se utilizó el siguiente ciclo de PCR, comenzando con 4 min de desnaturalización a 94 °C, seguida de 30 ciclos de 30 seg de desnaturalización a 94 °C, 30 seg de annealing a la Tm calculada para los primers en la reacción, y 2 min de extensión a 68 °C. Por último se realizó una extensión final a la misma temperatura por 10 min.

En las PCR semi-cuantitativas se utilizó como molde el cDNA preparado de cada una de las muestras, siendo la única modificación en el programa de PCR detallado anteriormente el número de ciclos de amplificación utilizándose 20.

Extracción de ARN y síntesis de cDNA

La extracción y síntesis de cDNA fue realizada por la Dra. Susana Vazquez. En primera instancia se realizó la extracción de RNA de las diferentes muestras mediante la utilización del MasterPure™ RNA Purification Kit de Epicentre. Posteriormente se digirió el DNA en las muestras utilizando la DNasa (Promega) en un volumen final de 15 μ l. Se Incubaron las muestras por 40 min a 37 °C y terminando la reacción con 1,5 μ l de solución stop a 65 °C por 10 min. La síntesis de cDNA se realizó a partir de las digestiones anteriores, usando la enzima SuperScript III (Invitrogen). El control negativo se realizó sin la polimerasa. Por último se corroboró que las muestras no estuvieran contaminadas con ADN genómico. Se realizó una PCR de las muestras, con los primers para amplificar el gen del ARN ribosomal de 16S.

Screening de solubilidad

Se realizó un ensayo de expresión a pequeña escala para determinar el estado de solubilidad de las proteínas recombinantes. Para esto se transformaron células competentes de *E. coli* BL21 (DE3) (Stratagene, La Jolla, CA) con los plásmidos de expresión recombinados utilizando el método de shock térmico [111]. Se realizaron cultivos en 10ml de LB con ampicilina los cuales se incubaron a 37 °C a 220 rpm, hasta una densidad óptica medida a 600 nm ($DO_{600\text{ nm}}$) de 0,4-0,6. La expresión de las proteínas recombinantes se indujo mediante la adición de isopropil-L-D-tiogalactósido (IPTG) a una concentración final 0,5 mM. Los cultivos inducidos fueron incubados 4hs a 37 y 28 °C y 16 hs a 20°C. Luego, cada cultivo se centrifugó 10 minutos a 6000 rpm y el pellet bacteriano se resuspendió en buffer de lisis (20 mM Tris-HCl, 0.5 M NaCl, pH 7,5). La lisis de las células se realizó por ultrasonido. La suspensión fue sometida a 20 pulsos de 1 minuto utilizando un sonicador (Cole Parmer, 4710 Series) en un baño de hielo y agua a una potencia de 25 watts. Los pulsos se separaron por un tiempo de reposo de 1 minuto. Finalizado el protocolo de lisis, se centrifugó la suspensión obtenida a una velocidad de 30000 g por 30 minutos a 4 °C. Las fracciones soluble e insoluble fueron analizadas mediante SDS-PAGE para determinar el porcentaje de solubilidad de las proteínas recombinantes.

Expresión de proteínas recombinantes solubles y purificación

Las proteínas que resultaron solubles se expresaron de igual forma que lo detallado para el screening de solubilidad, a partir de precultivos de 10 ml de LB con ampicilina los cuales se incubaron a 37 °C a 220 rpm, hasta $DO_{600\text{ nm}}$ de 0,4-0,6. Posteriormente se inocularon 500 ml de medio LB (ampicilina) usando la totalidad de los precultivos y se incubaron con una agitación de 220 rpm a 37 °C. La inducción de la expresión se realizó a una $DO_{600\text{ nm}}$ de 0,4-0,6 a través del agregado de IPTG (0,5 mM final) dejando el cultivo a 20°C toda la noche. Los cultivos se centrifugaron, resuspendieron y sonicaron como se detalló anteriormente.

Las proteínas recombinantes fueron purificadas en un equipo de FPLC (JASCO) mediante una columna IMAC HiTrap-HP (GE, Healthcare). La columna se equilibró con el buffer de lisis y se le aplicó la fracción soluble del lisado bacteriano a un flujo de 1 ml/min. Posteriormente, la columna fue lavada con un buffer (20 mM Tris-HCl, 0.5 M NaCl, 30 mM imidazol, pH 7,5), realizando como último paso la elución de la proteína con un gradiente lineal de imidazol (30 mM – 700 mM). Se recogieron todas las fracciones que presentaron señales de absorbancia a 280 nm y a continuación,

se analizaron mediante SDS-PAGE. Las fracciones que resultaron contener las proteínas recombinantes correspondientes se dializaron extensivamente contra buffer (20 mM Tris-HCl, 0.5 M NaCl, pH 8) a 4 °C. La región hexa-histidina se escindió usando la proteasa TEV (Sigma-Aldrich) según las recomendaciones del proveedor. Luego de analizar el porcentaje de digestión mediante SDS-PAGE, la proteasa TEV y la cola de histidina se eliminaron a través de un nuevo pasaje por una columna IMAC HiTrap-HP (la construcción de la proteasa TEV contiene un tag de histidinas) y su pureza se constató mediante SDS-PAGE. Las proteínas la cuales se necesitaron en altas concentraciones para su estudio mediante RMN o cristalografía de rayos X se concentraron mediante la utilización de tubos centrífuga (MILLIPORE) con cut off de 3 kDa. La concentración de cada proteína se estimó mediante absorbancia a 280 nm empleando el coeficiente de absorptividad molar teórico calculado mediante el servidor on-line ExPASy ProtParam (<http://web.expasy.org/protparam/>).

Ensayos de actividad fosfatasa

La actividad fosfatasa de BA42 se analizó mediante el uso de pNPP. (New England Biolabs) y O-fosfoserina (pSer) (Sigma-Aldrich). Las reacciones fosfatasa se llevaron a cabo en HEPES 25 mM, pH 7 a 37 ° C. Las reacciones se iniciaron por la adición de pNPP o pSer 20 mM, y se interrumpieron con el agregado de 10 µl de NaOH 1M. En el caso de la actividad mediante el uso del sustrato pNPP, se realizó un ensayo espectrofotométrico monitoreando el cambio de absorbancia a 405 nm del pNPP (incoloro) al p-nitrofenol (PNP, amarillo), $\epsilon_{405\text{nm}} = 18300 \text{ M}^{-1} \text{ cm}^{-1}$. Para la pSer se utilizó el método de Eibl and Lands [113] de detección del Pi liberado. Todos los ensayos se midieron en muestras por triplicado. La cinética enzimática de BA42 se calculó utilizando diferentes concentraciones de pSer a 20 °C, ajustando los datos a un modelo sigmoideo alosterico.

Con el fin de analizar el efecto en la actividad de los iones de metales divalentes, la proteína se trató con EDTA 1 mM durante 10 min y después se dializó extensivamente frente a un buffer HEPES 25 mM, pH 7,0. La enzima pretratada con EDTA se incubó con CaCl_2 1 mM, MgCl_2 1 mM, MnCl_2 1 mM o ZnCl_2 1 mM en el buffer de ensayo a 37 °C durante 10 min antes de iniciar la reacción. Las condiciones óptimas de actividad BA42 se determinaron utilizando pSer como sustrato. El pH óptimo se determinó mediante el uso de 25 mM de buffer acetato (pH 5-7), 25 mM Tris-HCl (pH 7 a 9), y 25 mM buffer glicina-NaOH (pH 9 a 12).

Análisis por Dicroísmo Circular (DC)

El perfil de estructura secundaria de las distintas proteínas se evaluó por la técnica de DC, midiendo sus espectros característicos en el UV lejano (200-260 nm). Todas las mediciones se llevaron a cabo a una concentración de proteína de 2 μM en HEPES 20 mM pH 7,0 en una cubeta de cuarzo de 0,2cm. Los valores de elipticidad (θ) en grados fueron convertidos a valores de elipticidad residual media (θ_{WRD}) según la ecuación M.1,

$$\theta_{WRD} = \frac{\theta(mdeg)}{L(cm) \cdot [P](M) \cdot np} \quad \text{Ecuación M.1}$$

donde θ es la elipticidad, L la longitud del paso de luz, P la concentración de proteína y np el número de residuos. El porcentaje de estructura secundaria fue analizado con el software CDpro. Los resultados obtenidos con los programas SELCON3, CONTILL y COSSTR fueron promediados para calcular el porcentaje de estructura secundaria de cada espectro. Adicionalmente, se realizó una predicción de estructura secundaria mediante el servidor on-line Jpred (<http://www.compbio.dundee.ac.uk/www-jpred/>). Las curvas de desnaturalización térmica se lograron siguiendo la θ a 222 nm. Se varió la temperatura de 25 a 90 °C a una velocidad constante de 1 °C.min⁻¹ y se tomaron mediciones a intervalos de 1 °C.

Dispersión estática de luz (SLS)

Los pesos moleculares promedio (PM) de BA40 y BA40- Δ H5 se determinaron utilizando un instrumento de dispersión de luz Precision Detectors PD2010 90° conectado en tándem a un equipo de HPLC, incluyendo un detector UV Waters 486 y un difractor diferencial LKB 2142. Para la determinación del peso molecular promedio se realizaron las corridas cromatográficas en una columna Superdex 75 GL 10/300 con un buffer que contenía Tris-HCl 25 mM, NaCl 200 mM, pH 8,0 a un flujo de 0,5 ml min⁻¹. En la elución de la proteína se midió la señal SLS a 90°, su absorción UV a 280 nm, y su índice de refracción (RI). Los datos se analizaron con el software Discovery 32 suministrado por Precision Detectors. El peso molecular de cada muestra se calculó comparando los valores de SLS y RI con los obtenidos para la BSA (MW 66,5 kDa). Los PM se

calcularon utilizando el software DTS v.7.11 utilizando el modelo de Rayleigh-Debye-Gans. Los datos representan un promedio de al menos cuatro mediciones.

Marcación selectiva con ^{15}N y ^{13}C

Las células transformadas con los plásmidos de expresión recombinados se cultivaron a 37 °C en medio mínimo M9 no marcado hasta un $\text{DO}_{600\text{ nm}}$ de 0,4-0,6 y la expresión de la proteína heteróloga se indujo mediante el agregado de 0,5 mM de IPTG. Después de 15 minutos, se añadió rifampicina a una concentración de 20 $\mu\text{g ml}^{-1}$. Luego de una incubación adicional de 15 minutos, las células se sedimentaron y se resuspendieron en un medio mínimo M9 conteniendo 1 g l^{-1} de $^{15}\text{NH}_4\text{Cl}$ y 2 g l^{-1} $^{13}\text{C}_6$ -glucosa (cuando se realizó el marcado doble), 1 mM de IPTG y 20 $\mu\text{g ml}^{-1}$ de rifampicina [60]. Las células se cultivaron durante toda la noche a 20 °C antes de su centrifugación y posterior lisis para la purificación de la proteína recombinante marcada.

Experimentos de RMN

Las muestras utilizadas en los experimentos de RMN contenían proteína con marcación simple ^{15}N o doble $^{15}\text{N}/^{13}\text{C}$ a concentraciones de 0,3-0,8 mM o proteína no marcada a 0,05-0,1 mM. Las muestras se prepararon en un buffer que contenía NaCl 25 mM, HEPES 20 mM y D_2O al 5%, pH 7. Los experimentos de RMN se realizaron en un espectrómetro Bruker Avance III de 600 MHz (Bruker Instruments, Inc., Bellerica, MA, EE.UU.) equipado con una criosonda de triple resonancia de 5 mm. Los datos de RMN se procesaron utilizando NMRPIPE [114] y se analizaron utilizando NMRVIEWJ [115].

HSQC y perturbación de desplazamientos químicos

Las variaciones en las resonancias de ^1H y ^{15}N se siguieron a 298 K mediante la realización de experimentos de HSQC. Las perturbaciones del cambio químico combinadas se calcularon con un factor de escala α_N de 0,17 utilizando la siguiente ecuación M.2:

$$\Delta_{CS} = \sqrt{((\Delta\delta_{HN})^2 + (\Delta\delta_N \cdot \alpha_N)^2)}$$

Ecuación M.2

Cálculo del “Chemical Shift Index”

Se realizó una predicción de estructura secundaria a través del cálculo de “Chemical Shift Index” con el programa CSI 3.0 (<http://csi3.wishartlab.com>), usando los valores de CS del H^α, C^α, C^β y C’.

Espectros 1D ¹H

Se utilizó la secuencia de impulsos de supresión de agua *Jump-and-return* 1-1 para los experimentos de ¹H 1D RMN. Este experimento se utilizó en la selección de los candidatos que mostraron bajos niveles de expresión observando la dispersión de los picos. En el estudio de la proteína BA42 y con el fin de seguir el cambio conformacional en presencia de metales divalentes se midieron las integrales de los picos de los grupos metilo de los residuos I100 y V88.

Relajación del ¹⁵N y NOE ¹H-¹⁵N

Las mediciones de T_1 , T_2 ¹⁵N y NOE ¹H-¹⁵N (hnNOE) sobre BA42 se realizaron a una frecuencia de nitrógeno de 60,82 MHz y 301 K, usando esquemas de impulsos estándar de una manera entremezclada (*interleaved*). Los datos de T_1 y T_2 se adquirieron con seis *delays* de relajación. Los *delays* utilizados fueron 7, 300, 600, 900, 1200 y 1600 ms para los cálculos de T_1 y 15,7, 31,4, 62,8, 125,5, 156,9 y 219,7 ms para los cálculos de T_2 . Las velocidades de relajación se analizaron mediante las herramientas del NMRVIEWJ [115]. Los valores de hnNOE se determinaron por la relación de intensidad de los picos registrados con y sin saturación del ¹H, empleando un *delay* de relajación de 4 s para cada exploración en ambos experimentos.

Típicamente, los errores fueron de aproximadamente 1% para T_1 y T_2 y 2-5% para las mediciones de hnNOEs.

Mediciones de radio hidrodinámico

Los experimentos de pulsed field gradient de RMN (PFG-RMN) se realizaron usando la secuencia PFG-SLED [116]. Se añadió dioxano (6% en H₂O) a todas las muestras como patrón interno y se registraron los espectros 1D ¹H antes y después del agregado del dioxano para asegurar que no hubo cambios en las resonancias de la proteína. Las longitudes de los impulsos y *delays* en la secuencia utilizada se mantuvieron constantes, adquiriéndose 19 espectros variando

la fuerza del gradiente de difusión entre 5% y 95% del valor máximo para el dioxano, BA42 y BA40. La longitud del *delay* (d_{20}) de difusión se calibró para las muestras con el fin de dar un decaimiento máximo del 85-90% para las señales de BA42, BA40 y el dioxano. Se utilizó un filtro de T_2 para observar selectivamente la señal de dioxano, sin interferencia de la proteína en cuestión. Los espectros de RMN de dioxano se adquirieron con 16 K puntos complejos, y los espectros de RMN de proteína con 4 K puntos complejos. Se observaron las intensidades de resonancia procedentes de la región aromática/amídica del espectro.

La intensidad del pico de resonancia (I/I_0) se relaciona con la fuerza del gradiente según la ecuación de Stejskal-Tanner (Ecuación M.3):

$$\frac{I}{I_0} = e^{-dg^2\left(\Delta - \frac{\delta}{3}\right)} \quad \text{Ecuación M.3}$$

donde Δ es el delay de difusión (que se estableció en 150 ms para BA42 y BA40 y 15 ms para el dioxano), δ es la duración del gradiente (4 ms), g es la fracción de la fuerza de gradiente máxima utilizada en cada experimento (de 0,05 a 0,95) y d es una constante directamente proporcional al coeficiente de difusión traslacional D . Según la ecuación de Stokes-Einstein, el coeficiente de difusión traslacional de una molécula está inversamente relacionado con su radio hidrodinámico efectivo (R_h). Esta magnitud se define como el radio de una esfera rígida no cargada que exhibe el mismo comportamiento hidrodinámico que la molécula solvatada en solución [117].

La constante d para la proteína y el dioxano se obtuvieron ajustando la Ecuación M.3 a los valores determinados experimentalmente de la intensidad de resonancia, graficando $\ln(I/I_0)$ en función de g^2 . El radio hidrodinámico de la proteína se calculó de acuerdo con la ecuación M.4:

$$R_h^{\text{Prot.}} = R_h^{\text{diox.}} \frac{d_{\text{diox.}}}{d_{\text{Prot.}}} \quad \text{Ecuación M.4}$$

utilizando un valor de 0,212 nm para el R_h de dioxano.

Cristalogénesis

Los ensayos de cristalización inicial se realizaron en placas de 96 pocillos usando un robot Honeybee 963 (Digilab, Marlborough, MA) y soluciones de *screening* de Jena Bioscience (Jena, Alemania) y Hampton Research (Aliso Viejo, CA), utilizando el método de difusión de “*sitting drop*”. Después de 5 días de equilibrio a 20 °C, de las 480 condiciones probadas, 22 y 10 condiciones para BA42 y C24, respectivamente, mostraron cristales prometedores. Posteriormente los cristales se optimizaron mediante el método de “*hanging drop*” mezclando 1 µl de las proteínas (15 mg/ml en Tris 10 mM, pH 7,5) con el mismo volumen de solución de cristalización. La siguiente condición produjo los mejores cristales para BA42: PEG 2000 MME 24% (m/v), acetato de sodio 0,1 M, MES 0,1 M, pH 6,5. Para C24 la mejor condición para la obtención de los cristales fue: 5000 PEG 8000 12% (m/v), citrato de sodio 0,1 M y sulfato amónico 0,1 M. Las muestras fueron crioprotegidas en licor madre añadiendo PEG 400 10% (m/v) y luego congeladas en nitrógeno líquido en bucles Hampton Research (Aliso Viejo, CA).

Ensayos de retardo en la movilidad electroforética (EMSA)

Se ensayaron diferentes fragmentos de ADN de 200 pb. Entre ellos, un fragmento genérico, la región promotora del operón BA42 y de un probable homólogo de la proteína CsgD. Estos fragmentos se obtuvieron por PCR empleando los primers detallados en la tabla M.2 La reacción de unión se llevó a cabo en un volumen final de 20 µl en un buffer 20 mM HEPES, pH 7.9 conteniendo 100 mM KCl, 5 % glycerol y 200 mg/ml seroalbúmina bovina (BSA). Las reacciones fueron realizadas incubando 1,5 pmoles de ADN con concentraciones crecientes de BA40 por 30 min a 20 °C. Posteriormente, las muestras fueron sometidas a electroforesis en geles nativos de poliacrilamida al 6% durante aproximadamente 2hs, manteniendo fija la corriente en 5 mA y bajo refrigeración. Finalmente los geles fueron teñidos con bromuro de etido para visualizar el ADN.

<i>ID</i>	<i>Primer 5'</i>	<i>Primer 3'</i>
<i>Promotor Operon BA42</i>	TGGCTAAATTTACTATTTTTGCAG	TAAATGAAAATGCTTTGTAAGCTTATG
<i>Promotor CsgD_Bizionia</i>	TTGTAAATAGCCTCATTGGAATG	TTATGAATGACTTACAACAGAATC
<i>Fragmento genérico</i>	ATGAAACGATTAGACCCTAC	CAACTTCGCCTATTTTCATC

Tabla M.2 Primers utilizados para la amplificación de los diferentes fragmentos de 200 pb.

SELEX

Se añadieron un total de 100 μ l de una suspensión al 50% de resina Ni-NTA a un tubo de centrífuga con filtro de 1,5 ml y se lavó tres veces con 400 μ l de agua ultra-pura. La proteína BA40 conteniendo en su extremo N-terminal la región de hexa-histidinas fue aplicada en la resina a una concentración de 2 μ M. La fracción no unida de BA40 se eliminó por centrifugación a 100 g y 4 °C durante 1 min. Antes de unirse al ADN, la proteína BA40 inmovilizada se lavó dos veces con solución salina (PBS, pH 7,4). A continuación, la columna Ni-NTA unida a BA40 se equilibró con buffer de unión [10 mM Tris-HCl, 75 mM NaCl, glicerol al 6%, 1 mg/ml de BSA, 30 μ g/ml de esperma de salmón, pH 7,5,].

Se preparó una biblioteca de ADN aleatoria sintetizando químicamente el ADN de una sola hebra de *GAGAGGATCCAGTCAGCATG-N18-CTCAGCCTCGAGAATTCGACA* (Integrated DNA Technologies, Coralville, EE.UU.), en el que N18 es un fragmento de ADN aleatorio de 18 bases. Estas moléculas fueron luego amplificadas para generar una biblioteca de ADN doble cadena. Los productos de extensión se extrajeron adicionalmente con fenol/cloroformo/alcohol isoamílico (25/24/1) y se precipitaron con 2,5 volúmenes de etanol y 1/10 de un volumen de acetato de sodio 3 M.

La biblioteca de ADN aleatorio se aplicó a una concentración de 50 ng/ μ l sobre la resina de Ni-NTA previamente equilibrada con PBS y con la proteína BA40 unida, incubándose a 4 °C durante 30 min. A continuación, 400 μ l de buffer de lavado (15 mM Tris-HCl, 500 mM NaCl, 0,15% Tritón-X-100, 1 % BSA, pH 7,5) se aplicaron tres veces a la columna unida al complejo BA40-ADN. Posteriormente, se realizaron lavados con imidazol a concentraciones escalonadas de 100, 200, 300 y 400 mM. Finalmente, el lavado de 300 y 400 mM de imidazol se calentó a 100 °C durante 10 minutos para eluir cualquier ADN unido por la proteína BA40 inmovilizada. La solución de ADN eluida y calentada se retiró y se sometió a la amplificación por PCR. Los productos de PCR (15 μ l) se sometieron a un segundo round de selección por SELEX. Este procedimiento se repitió seis veces. EL ADN final fue clonado en el vector pGEM®-T Easy (Promega), para luego transformar células XL-1 Blue (Stratagene) y secuenciar los clones positivos. Las secuencias de todos los ADN seleccionados por SELEX fueron sometidas a la búsqueda de motivos mediante el programa MEME suite [106] a fin de encontrar una secuencia consenso.

Ensayos de anisotropía de fluorescencia

La anisotropía de fluorescencia se midió usando un espectrofluorímetro Jasco FP - 6500 (Jasco Inc., Easton, MD, EE.UU.) arreglado en geometría “L” equipado con un controlador peltier. Las mediciones se realizaron monitoreando la emisión del 6-FAM excitando a 495 nm, usando un monocromador de detección ajustado a 520 nm. Se detallan a continuación los oligonucleótidos utilizados unidos a 6-FAM en el extremo 5': Oligo. Consenso SELEX (5'-6-FAM-TGGGAGTGACATCGGTAC-3'), Oligo. Aleatorio (5'-6-FAM-CGGCCCCCTTCGGGGCCT-3').

Las mediciones se realizaron en una micro cubeta de fluorescencia de 3 mm x 3 mm. Las muestras se prepararon en un volumen de 250 μ l en buffer 20 mM HEPES, pH 7,2, 50 NaCl. Las curvas de titulación se realizaron a concentraciones fijas de los oligonucleotidos (50, 100 y 500 nM) agregando pequeñas cantidades de las proteínas BA40 y BA40- Δ H5. En cada medición, tras el agregado de la proteína, se dejó equilibrar la muestra por dos minutos antes de medir. En ningún caso los oligonucleótidos marcados se diluyeron en la titulación superando un 10% del volumen total. En este sentido, los datos de intensidad de fluorescencia se corrigieron linealmente según la dilución. La temperatura se mantuvo constante en 20 °C. Los datos se ajustaron mediante una regresión no lineal utilizando la ecuación de una curva alostérica sigmoidea mediante el programa GraphPad Prism. Los datos de cada serie se ajustaron de forma independiente.

Dinámica Molecular

Esta fue realizada por los Doctores Raúl E. Ithuralde y Adrián G. Turjanski. Los archivos de estructura inicial se obtuvieron a partir de la estructura de BA42 del Banco de Datos de Proteínas (PDB ID: 4OA3) [Aran]. Se utilizó el campo de fuerza AMBER99SB [118] y el programa tleap para crear los archivos de topología y coordenadas. Amber se utilizó para todas las simulaciones de dinámica molecular (MD) [119]. Las moléculas de agua para las condiciones de solvente explícito se representaron utilizando el modelo TIP3P [120]. El sistema se equilibró en el ensamble NVT mediante una simulación de MD de 25 ps de duración utilizando el termostato de Berendsen [121]. Luego, la temperatura se elevó lentamente a 300K llevando a cabo simultáneamente otra simulación de 25 ps. Durante estos procesos los átomos de CA fueron restringidos utilizando un potencial armónico con una constante de 20 kcal/mol para la termalización. Esto fue seguido por un proceso de equilibración de 50 ps en pasos de 1 fs usando la temperatura de Berendsen con una constante de acoplamiento de 0,1 ps para mantener la temperatura constante a 300 K y el baróstato de Berendsen para mantener la presión constante a 1 atm [121]. Durante los ciclos de

200 ns, la temperatura se mantuvo constante mediante el algoritmo del termostato de Berendsen ajustado a 300 K con una constante de acoplamiento de 0,1 ps y la presión se mantuvo constante a 1 atm por el baróstato de Brendsen [121]. El algoritmo SHAKE se utilizó para parametrizar los enlaces que contenían un átomo de H [122] y las interacciones electrostáticas se calcularon por el algoritmo de Particle Mesh Ewald [123]. Se utilizó un intervalo de tiempo de 2 fs. El programa VMD se utilizó para la visualización y los cálculos de RMSD y RMSF [124].

Co-inmunoprecipitación

Una muestra de *B. argentinensis* crecida en LB marino fue lisada en un buffer conteniendo 50 mM Tris-HCl (pH 7,5), 100 mM NaCl y 1mM PMSF, utilizando ultrasonido con los parámetros de tiempo y potencia descritos anteriormente. Posteriormente, se incubaron 10 μ l del antisuero preparado en ratón contra la proteína BA42 por 3 horas a 4 °C. A continuación, se agregaron 40 μ l de una solución de 0,5 mg/ml de proteína A en PBS y se incubó por 30 min a 4 °C. Se centrifugó a 2000 g por 5 min y el precipitado se lavó 3 veces con PBS y otras 3 veces con 10 mM HEPES (7,5). Finalmente, se agregó el cracking buffer, se calentó 5 minutos a 100 °C y se realizó un SDS-PAGE no reductor. Al gel se le realizó una tinción con Coomassie Blue Coloidal. Las bandas diferenciales con respecto al control sin anticuerpo anti-BA42 fueron enviadas para su identificación por espectrometría de masas MALDI-TOF/TOF en el CEQUIBIEM.

Cultivos de *Bizionia argentinensis*

A partir de colonias aisladas de *B. argentinensis* en agar marino se preparó una suspensión para inocular un cultivo en medio LB marino (5 g/l de peptona, 5 g/l de extracto de levadura y agua de mar). Los cultivos se incubaron a 20 °C y 200 rpm hasta fase exponencial tardía (aproximadamente $DO_{600\text{ nm}} = 0,8$), en la cual se aplicaron los diferentes tratamientos.

En el caso particular de los cultivos para proceder a la extracción de RNA los mismos fueron realizados por la Dra. Susana Vazquez. Se tomaron muestras a la hora y a las 7 hs de exposición a los tratamientos, que fueron conservadas con RNAlater a -80 °C para la posterior extracción del RNA, síntesis de cDNA y detección de la presencia de BA40. Para esto, se colocaron 100 μ l de RNAlater en tubos eppendorf nuevos libres de DNAsas y RNAsas, en hielo. Se agregó 1,5 ml de los cultivos, se mezcló y centrifugó, descartando el sobrenadante y conservando el pellet a -80 °C hasta su procesamiento.

Cromatografía de exclusión molecular (SEC)

Los tamaños moleculares de las proteínas BA40 y BA40- Δ H5 se determinaron por cromatografía de exclusión molecular. La corrida cromatografía se realizó en columna en una columna Superdex 200 analítica (GE Healthcare Life Sciences) utilizando un buffer Tris-HCl 25 mM pH 7,5, NaCl 250 mM. Para la calibración de la columna se utilizaron los siguientes estándares de tamaño: Ovalbúmina (43 kDa), quimiotripsina A (25 kDa) y RNasa A (13,7 kDa).

Experimentos de entrecruzamiento

Los experimentos de entrecruzamiento con las proteínas BA40 y BA40- Δ H5 solas y con el agregado de ADN se realizaron en un buffer fosfato de sodio 25 mM pH 7,4, 100 mM NaCl, 20 % glicerol. Las proteínas se ensayaron a una concentración de 0.1 μ g/ μ l en un volumen final de 20 μ l. El ADN se utilizó en una concentración de 50 ng/ μ l. Como agente de entrecruzamiento se utilizó glutaraldehído a una concentración final de 0,1 % durante 5 min a temperatura ambiente. La reacción se detuvo mediante la adición de 100 mM Tris, pH 7,4. Posteriormente la reacción se separó mediante un gel SDS-PAGE al 12%.

Anexo

Number of experimental restraints	2650
Distance restraints from NOEs	1560
Intraresidue	437
Sequential	498
Medium range	273
Long range	352
Hydrogen bond distance restraints	172
Dihedral angle restraints	254
Residual dipolar coupling constants	332
H—N	116
H _α —C _α	113
C _α —CO	103
Average number of restraints per residue	19
XPLOR energies (kcal mol ⁻¹)	
<i>E</i> _{total}	613.2 ± 19.8
<i>E</i> _{bond}	23.0 ± 3.7
<i>E</i> _{angle}	282.2 ± 7.8
<i>E</i> _{improper}	58.1 ± 3.0
<i>E</i> _{vdw}	26.9 ± 8.1
<i>E</i> _{dih}	0.3 ± 0.2
<i>E</i> _{noe}	25.6 ± 2.7
<i>E</i> _{scni}	197.1 ± 3.0
Rms deviations from experimental restraints	
Average distance restraints violation (Å)	0.021 ± 0.001
Average dihedral angle restraints violation (°)	0.13 ± 0.06
Average H—N RDC constants violation (Hz)	0.68 ± 0.03 ^a
Average H _α —C _α RDC constants violation (Hz)	1.06 ± 0.03 ^a
Average C _α —CO RDC constants violation (Hz)	3.13 ± 0.05 ^a
<i>R</i> _{rms} (%)	
H—N RDC	30.8
H _α —C _α RDC	26.8
C _α -CORDC	25.9
Rms deviations from idealized covalent geometry	
Bond (Å)	0.0030 ± 0.0001
Angle (°)	0.660 ± 0.007
Improper (°)	0.57 ± 0.01
MolProbity analysis	
Clashscore ^b	0
Ramachandran favored (%)	95.8 ± 1.2
Ramachandran outliers (%)	0.39 ± 0.48
MolProbity score ^c	0.99 ± 0.22
What if quality indicators ^{d,e}	
First generation packing quality	1.9 ± 0.5
Second generation packing quality	5.1 ± 1.5
Ramachandran plot appearance	1.7 ± 0.2
Chi-1/chi-2 rotamer normality	3.5 ± 0.5
Backbone conformation	0.5 ± 0.2
Coordinates precision	
	Backbone (Å) All heavy atoms (Å)
	0.27 ± 0.04 0.84 ± 0.06

Tabla A.1. Restricciones experimentales y estadística estructural para las 20 estructuras de energía más baja de la proteína BA42.

^a Normalizado con respecto a valor de H-N.

^b Número de superposiciones estéricas (> 0,4 Å) por 1000 átomos.

^c Combina las evaluaciones de “clashscore”, rotameros y Ramachandran en un solo valor, normalizado para estar en la misma escala que la resolución de rayos X.

^d Z-score

^e Residuos 7-145

		BA42
Data collection		
Synchrotron source		SOLEIL
Beamline		PROXIMA 1
Number of frames		600
Oscillation step (°)		0.2
Detector distance (mm)		196
Wavelength (Å)		0.97857
Exposure per frame (s)		0.2
Indexing and scaling		
Cell parameters	<i>a</i> (Å)	31.92
	<i>b</i> (Å)	39.62
	<i>c</i> (Å)	106.97
	$\alpha = \beta = \gamma$ (°)	90
Spacegroup		<i>P</i> 2 ₁ 2 ₁ 2 ₁
Resolution limit (Å)		1.39
Number of total reflections		118,799
Number of unique reflections		27,591
Average multiplicity ^a		4.3 (4.4)
$\langle I/\sigma(I) \rangle$		16.0 (2.0)
<i>R</i> _{merge}		0.049 (0.695)
Completeness (%)		98.1 (96.8)
Molecules per asymmetric unit		1
Solvent content (%)		40
<i>B</i> -factor (Wilson plot, Å ²)		20
Refinement		
Resolution range (Å)		30.6–1.39
Number of protein atoms		1109
Number of ligand atoms		2
Number of water molecules		169
<i>R</i>		0.174
<i>R</i> _{free}		0.205
Rms deviations from ideal values		
Bond lengths (Å)		0.01
Bond angles (°)		1.1
<i>B</i> -factor (average, Å ²)		24
MolProbity validation^b		
Clashscore		0.88
Poor rotamers (%)		1.6
Ramachandran plot		
Favored (%)		98.6
Allowed (%)		1.4
Disallowed (%)		—

Tabla A.2. Estadística estructural y datos experimentales de la estructura cristalina de BA42.

^a Los valores entre paréntesis corresponden a la capa de máxima resolución: 1.47-1.39 Å.

^b Chen et al. [125]

Bibliografía

1. Kanehisa, M. & Bork, P. (2003) Bioinformatics in the post-sequence era, *Nature genetics*. **33 Suppl**, 305-10.
2. Levitt, M. (2009) Nature of the protein universe, *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 11079-84.
3. Khafizov, K., Madrid-Aliste, C., Almo, S. C. & Fiser, A. (2014) Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative, *Proceedings of the National Academy of Sciences of the United States of America*. **111**, 3733-8.
4. Dessailly, B. H., Nair, R., Jaroszewski, L., Fajardo, J. E., Kouranov, A., Lee, D., Fiser, A., Godzik, A., Rost, B. & Orengo, C. (2009) PSI-2: structural genomics to cover protein domain family space, *Structure*. **17**, 869-81.
5. Jaroszewski, L., Li, Z., Krishna, S. S., Bakolitsa, C., Wooley, J., Deacon, A. M., Wilson, I. A. & Godzik, A. (2009) Exploration of uncharted regions of the protein universe, *PLoS biology*. **7**, e1000205.
6. Chim, N., Habel, J. E., Johnston, J. M., Krieger, I., Miallau, L., Sankaranarayanan, R., Morse, R. P., Bruning, J., Swanson, S., Kim, H., Kim, C. Y., Li, H., Bulloch, E. M., Payne, R. J., Manos-Turvey, A., Hung, L. W., Baker, E. N., Lott, J. S., James, M. N., Terwilliger, T. C., Eisenberg, D. S., Sacchettini, J. C. & Goulding, C. W. (2011) The TB Structural Genomics Consortium: a decade of progress, *Tuberculosis (Edinb)*. **91**, 155-72.
7. Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreuzsch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L. S., Miller, M. D., McPhillips, T. M., Miller, M. A., Scheibe, D., Canaves, J. M., Guda, C., Jaroszewski, L., Selby, T. L., Elsliger, M. A., Wooley, J., Taylor, S. S., Hodgson, K. O., Wilson, I. A., Schultz, P. G. & Stevens, R. C. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, *Proceedings of the National Academy of Sciences of the United States of America*. **99**, 11664-9.
8. Li, L., Nan, J., Li, D., Brostromer, E., Wang, Z., Liu, C., Hou, Q., Fan, X., Ye, Z. & Su, X. D. (2014) Structural genomics studies of human caries pathogen *Streptococcus mutans*, *Journal of structural and functional genomics*. **15**, 91-9.
9. Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J. & Davis, S. J. (2007) Glycoprotein structural genomics: solving the glycosylation problem, *Structure*. **15**, 267-73.
10. Perera, R. & Kuhn, R. J. (2008) Structural proteomics of dengue virus, *Current opinion in microbiology*. **11**, 369-77.
11. Leroy, B., Gillet, L., Vanderplasschen, A. & Wattiez, R. (2016) Structural Proteomics of Herpesviruses, *Viruses*. **8**.

12. Grabowski, M., Niedzialkowska, E., Zimmerman, M. D. & Minor, W. (2016) The impact of structural genomics: the first quinquennial, *Journal of structural and functional genomics*. **17**, 1-16.
13. Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G. & Joachimiak, A. (2011) High-throughput protein purification and quality assessment for crystallization, *Methods*. **55**, 12-28.
14. Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., Wooley, J., Wuthrich, K. & Wilson, I. A. (2010) The JCSG high-throughput structural biology pipeline, *Acta crystallographica Section F, Structural biology and crystallization communications*. **66**, 1137-42.
15. DiDonato, M., Deacon, A. M., Klock, H. E., McMullan, D. & Lesley, S. A. (2004) A scalable and integrated crystallization pipeline applied to mining the *Thermotoga maritima* proteome, *Journal of structural and functional genomics*. **5**, 133-46.
16. Shin, J. & Lee, W. (2008) Structural proteomics by NMR spectroscopy, *Expert review of proteomics*. **5**, 589-601.
17. Banci, L., Bertini, I., Luchinat, C. & Mori, M. (2010) NMR in structural proteomics and beyond, *Progress in nuclear magnetic resonance spectroscopy*. **56**, 247-66.
18. Serrano, P., Dutta, S. K., Proudfoot, A., Mohanty, B., Susac, L., Martin, B., Geralt, M., Jaroszewski, L., Godzik, A., Elsliger, M., Wilson, I. A. & Wuthrich, K. (2016) NMR in structural genomics to increase structural coverage of the protein universe: Delivered by Prof. Kurt Wuthrich on 7 July 2013 at the 38th FEBS Congress in St. Petersburg, Russia, *The FEBS journal*. **283**, 3870-3881.
19. Galperin, M. Y. & Koonin, E. V. (2004) "Conserved hypothetical" proteins: prioritization of targets for experimental study, *Nucleic acids research*. **32**, 5452-63.
20. Suits, M. D., Matte, A., Jia, Z. & Cygler, M. (2008) The Contribution of Structural Proteomics to Understanding the Function of Hypothetical Proteins in *Structural Proteomics and Its Impact on the Life Sciences* (Joel L Sussman, I. S., ed), World Scientific.
21. Mistry, J., Kloppmann, E., Rost, B. & Punta, M. (2013) An estimated 5% of new protein structures solved today represent a new Pfam family, *Acta crystallographica Section D, Biological crystallography*. **69**, 2186-93.
22. Chandonia, J. M. & Brenner, S. E. (2006) The impact of structural genomics: expectations and outcomes, *Science*. **311**, 347-51.
23. Manjasetty, B. A., Büsow, K., Panjekar, S. & Turnbull, A. P. (2011) Current methods in structural proteomics and its applications in biological sciences, *3 Biotech*. **2**, 89-113.

24. Manjasetty, B. A., Turnbull, A. P. & Panjekar, S. (2009) The impact of Structural Proteomics on Biotechnology, *Biotechnology and Genetic Engineering Reviews*. **26**, 353-370.
25. Owens, R. (2011) Methods in structural proteomics, *Methods*. **55**, 1-2.
26. Overton, I. M. & Barton, G. J. (2011) Computational approaches to selecting and optimising targets for structural biology, *Methods*. **55**, 3-11.
27. Bryan, C. M., Bhandari, J., Napuli, A. J., Leibly, D. J., Choi, R., Kelley, A., Van Voorhis, W. C., Edwards, T. E. & Stewart, L. J. (2011) High-throughput protein production and purification at the Seattle Structural Genomics Center for Infectious Disease, *Acta crystallographica Section F, Structural biology and crystallization communications*. **67**, 1010-4.
28. Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H. W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S. H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C. Y., Hung, L. W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K. & Gunsalus, K. C. (2008) Protein production and purification, *Nature methods*. **5**, 135-46.
29. Page, R., Peti, W., Wilson, I. A., Stevens, R. C. & Wuthrich, K. (2005) NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline, *Proceedings of the National Academy of Sciences of the United States of America*. **102**, 1901-5.
30. Newman, J. (2011) One plate, two plates, a thousand plates. How crystallisation changes with large numbers of samples, *Methods*. **55**, 73-80.
31. Abola, E. E. & Stevens, R. C. (2008) The Importance of Target Selection Strategies in Structural Biology in *Structural Proteomics and Its Impact on the Life Sciences* (Joel L Sussman, I. S., ed), World Scientific.
32. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004) The Pfam protein families database, *Nucleic acids research*. **32**, D138-41.
33. Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future, *Nucleic acids research*. **44**, D279-85.

34. Casanueva, A., Tuffin, M., Cary, C. & Cowan, D. A. (2010) Molecular adaptations to psychrophily: the impact of 'omic' technologies, *Trends in microbiology*. **18**, 374-81.
35. D'Amico, S., Collins, T., Marx, J. C., Feller, G. & Gerday, C. (2006) Psychrophilic microorganisms: challenges for life, *EMBO reports*. **7**, 385-9.
36. De Maayer, P., Anderson, D., Cary, C. & Cowan, D. A. (2014) Some like it cold: understanding the survival strategies of psychrophiles, *EMBO reports*. **15**, 508-17.
37. Siddiqui, K. S., Bokhari, S. A., Afzal, A. J. & Singh, S. (2004) A novel thermodynamic relationship based on Kramers Theory for studying enzyme kinetics under high viscosity, *IUBMB life*. **56**, 403-7.
38. Feller, G. (2007) Life at low temperatures: is disorder the driving force?, *Extremophiles : life under extreme conditions*. **11**, 211-6.
39. Struvay, C. & Feller, G. (2012) Optimization to low temperature activity in psychrophilic enzymes, *International journal of molecular sciences*. **13**, 11643-65.
40. Feller, G. (2013) Psychrophilic enzymes: from folding to function and biotechnology, *Scientifica*. **2013**, 512840.
41. Chiuri, R., Maiorano, G., Rizzello, A., del Mercato, L. L., Cingolani, R., Rinaldi, R., Maffia, M. & Pompa, P. P. (2009) Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases, *Biophysical journal*. **96**, 1586-96.
42. Kirchman, D. L. (2002) The ecology of Cytophaga-Flavobacteria in aquatic environments, *FEMS microbiology ecology*, 91-100.
43. Fernandez-Gomez, B., Richter, M., Schuler, M., Pinhassi, J., Acinas, S. G., Gonzalez, J. M. & Pedros-Alio, C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach, *The ISME journal*. **7**, 1026-37.
44. Amaral-Zettler, L., Artigas, L. F., Baross, J., Bharathi, L., Boetius, A., Chandramohan, D., Herndl, G., Kogure, K., Neal, P., Pedrós-Alió, C., Ramette, A., Schouten, S., Stal, L., Thessen, A., Leeuw, J. d. & Sogin, M. (2010) A global census of marine microbes in *Life in the World's Oceans: Diversity, Distribution, and Abundance* (McIntyre, A. D., ed), Blackwell
45. Bowman, J. P. (2006) The Marine Clade of the Family Flavobacteriaceae: The Genera Aequorivita, Arenibacter, Cellulophaga, Croceibacter, Formosa, Gelidibacter, Gillisia, Maribacter, Mesonia, Muricauda, Polaribacter, Psychroflexus, Psychroserpens, Robiginitalea, Salegentibacter, Tenacibaculum, Ulvibacter, Vitellibacter and Zobellia, 677-694.

46. Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K. H., Ludwig, W., Glockner, F. O. & Rossello-Mora, R. (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains, *Systematic and applied microbiology*. **31**, 241-50.
47. Gonzalez, J. M., Fernandez-Gomez, B., Fernandez-Guerra, A., Gomez-Consarnau, L., Sanchez, O., Coll-Llado, M., Del Campo, J., Escudero, L., Rodriguez-Martinez, R., Alonso-Saez, L., Latasa, M., Paulsen, I., Nedashkovskaya, O., Lekunberri, I., Pinhassi, J. & Pedros-Alio, C. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria), *Proceedings of the National Academy of Sciences of the United States of America*. **105**, 8724-9.
48. Matthew T. Cottrell & David L. Kirchman (2000) Natural Assemblages of Marine Proteobacteria and Members of the *Cytophaga-Flavobacter* Cluster Consuming Low- and High-Molecular-Weight Dissolved Organic Matter, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*. **6**, 1692–1697.
49. McBride, M. J., Xie, G., Martens, E. C., Lapidus, A., Henrissat, B., Rhodes, R. G., Goltsman, E., Wang, W., Xu, J., Hunnicutt, D. W., Staroscik, A. M., Hoover, T. R., Cheng, Y. Q. & Stein, J. L. (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis, *Appl Environ Microbiol*. **75**, 6864-75.
50. DeLong, E. F., Franks, D. G. & Alldredge, A. L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages *Limnol Oceanogr*. **38** 924-934
51. Pinhassi, J., Sala, M. M., Havskum, H., Peters, F., Guadayol, O., Malits, A. & Marrase, C. (2004) Changes in bacterioplankton composition under different phytoplankton regimens, *Appl Environ Microbiol*. **70**, 6753-66.
52. Shrivastava, A. & Berg, H. C. (2015) Towards a model for *Flavobacterium* gliding, *Current opinion in microbiology*. **28**, 93-7.
53. McBride, M. J. & Zhu, Y. (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes, *Journal of bacteriology*. **195**, 270-8.
54. J., C. M. (2008) Resource control of bacterial dynamics in the sea in *Microbial Ecology of the Oceans* (L, K. D., ed) pp. 335–382, John Wiley & Sons.
55. Bercovich, A., Vazquez, S. C., Yankilevich, P., Coria, S. H., Foti, M., Hernandez, E., Vidal, A., Ruberto, L., Melo, C., Marensi, S., Criscuolo, M., Memoli, M., Arguelles, M. & Mac Cormack, W. P. (2008) *Bizionia argentinensis* sp. nov., isolated from surface marine water in Antarctica, *International journal of systematic and evolutionary microbiology*. **58**, 2363-7.
56. Lanzarotti, E., Pellizza, L., Bercovich, A., Foti, M., Coria, S. H., Vazquez, S. C., Ruberto, L., Hernandez, E. A., Dias, R. L., Mac Cormack, W. P., Cicero, D. O., Smal, C., Nicolas, M. F.,

- Vasconcelos, A. T., Marti, M. A. & Turjanski, A. G. (2011) Draft genome sequence of *Bizionia argentinensis*, isolated from Antarctic surface water, *Journal of bacteriology*. **193**, 6797-8.
57. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database, *Nucleic acids research*. **43**, D261-9.
58. Skolnick, J., Fetrow, J. S. & Kolinski, A. (2000) Structural genomics and its importance for gene function analysis, *Nature biotechnology*. **18**, 283-7.
59. Zhang, M. & Leong, H. W. (2010) Bidirectional best hit r-window gene clusters, *BMC bioinformatics*. **11 Suppl 1**, S63.
60. Galvão-Botton, L. M. P., Katsuyama, Â. M., Guzzo, C. R., Almeida, F. C. L., Farah, C. S. & Valente, A. P. (2003) High-throughput screening of structural proteomics targets using NMR, *FEBS Letters*. **552**, 207-213.
61. Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C. S., Semesi, A., Guido, V., Edwards, A. M. & Arrowsmith, C. H. (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins, *Journal of the American Chemical Society*. **127**, 16512-7.
62. Dessau, M. A. & Modis, Y. (2011) Protein crystallization for X-ray crystallography, *Journal of visualized experiments : JoVE*.
63. Aran, M., Smal, C., Pellizza, L., Gallo, M., Otero, L. H., Klinke, S., Goldbaum, F. A., Ithurralde, E. R., Bercovich, A., Mac Cormack, W. P., Turjanski, A. G. & Cicero, D. O. (2014) Solution and crystal structure of BA42, a protein from the Antarctic bacterium *Bizionia argentinensis* comprised of a stand-alone TPM domain, *Proteins*. **82**, 3062-78.
64. Jenney, F. E., Jr. & Adams, M. W. (2008) The impact of extremophiles on structural genomics (and vice versa), *Extremophiles : life under extreme conditions*. **12**, 39-50.
65. Wegener, K. M., Bennewitz, S., Oelmuller, R. & Pakrasi, H. B. (2011) The Psb32 protein aids in repairing photodamaged photosystem II in the cyanobacterium *Synechocystis* 6803, *Molecular plant*. **4**, 1052-61.
66. Wu, H. Y., Liu, M. S., Lin, T. P. & Cheng, Y. S. (2011) Structural and functional assays of AtTLP18.3 identify its novel acid phosphatase activity in thylakoid lumen, *Plant physiology*. **157**, 1015-25.
67. Eletsky, A., Acton, T. B., Xiao, R., Everett, J. K., Montelione, G. T. & Szyperski, T. (2012) Solution NMR structures reveal a distinct architecture and provide first structures for protein domain family PF04536, *Journal of structural and functional genomics*. **13**, 9-14.

68. Sirpio, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N. & Aro, E. M. (2007) TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle, *The Biochemical journal*. **406**, 415-25.
69. Boulin, T., Rapti, G., Briseno-Roa, L., Stigloher, C., Richmond, J. E., Paoletti, P. & Bessereau, J. L. (2012) Positive modulation of a Cys-loop acetylcholine receptor by an auxiliary transmembrane subunit, *Nature neuroscience*. **15**, 1374-81.
70. Sinha, A., Eniyan, K., Sinha, S., Lynn, A. M. & Bajpai, U. (2015) Functional analysis of TPM domain containing Rv2345 of Mycobacterium tuberculosis identifies its phosphatase activity, *Protein expression and purification*. **111**, 23-7.
71. Stark, J. L., Copeland, J. C., Eletsky, A., Somerville, G. A., Szyperski, T. & Powers, R. (2014) Identification of low-molecular-weight compounds inhibiting growth of corynebacteria: potential lead compounds for antibiotics, *ChemMedChem*. **9**, 282-5.
72. Daragan, V. A. & Mayo, K. H. (1997) Motional model analyses of protein and peptide dynamics using ^{13}C and ^{15}N NMR relaxation, *Progress in nuclear magnetic resonance spectroscopy*. **31**, 63–105.
73. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations, *J Magn Reson*. **147**, 138-46.
74. Finkelstein, A. V. & Ptitsyn, O. B. (1987) Why Do Globular Proteins Fit The Limited Set Of Folding Patterns? , *Prog Biophys molec Biol*. **50**, 171-190.
75. Alexei V. Finkelstein, Gutun, A. M. & Badretdinov, A. Y. (1993) Why are the same protein folds used to perform different functions? , *The FEBS journal*. **325**, 23-28.
76. Hsu, F. & Mao, Y. (2015) The structure of phosphoinositide phosphatases: Insights into substrate specificity and catalysis, *Biochimica et biophysica acta*. **1851**, 698-710.
77. Seifried, A., Schultz, J. & Gohla, A. (2013) Human HAD phosphatases: structure, mechanism, and roles in health and disease, *The FEBS journal*. **280**, 549-71.
78. Holm, L. & Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic acids research*. **26**, 316–319.
79. Kawabata, T. (2003) MATRAS: a program for protein 3D structure comparison, *Nucleic acids research*. **31**, 3367-3369.

80. Pellizza, L. A., Smal, C., Ithuralde, R. E., Turjanski, A. G., Cicero, D. O. & Aran, M. (2016) Structural and functional characterization of a cold-adapted stand-alone TPM domain reveals a relationship between dynamics and phosphatase activity, *The FEBS journal*. **283**, 4370-4385.
81. G.Dougherty, W., C.Carrington, J., M.Cary, S. & Parks, T. D. (1988) Biochemical and mutational analysis of a plant virus polyprotein cleavage site, *The EMBO Journal* **7**, 1281-1287.
82. Zhao, H. L., Chen, X. L., Xie, B. B., Zhou, M. Y., Gao, X., Zhang, X. Y., Zhou, B. C., Weiss, A. S. & Zhang, Y. Z. (2012) Elastolytic mechanism of a novel M23 metalloprotease pseudoalterin from deep-sea Pseudoalteromonas sp. CF6-2: cleaving not only glycol bonds in the hydrophobic regions but also peptide bonds in the hydrophilic regions involved in cross-linking, *The Journal of biological chemistry*. **287**, 39710-20.
83. Cerutti, M. L., Otero, L. H., Smal, C., Pellizza, L., Goldbaum, F. A., Klinke, S. & Aran, M. (2016) Structural and functional characterization of a cold adapted TPM-domain with ATPase/ADPase activity, *Journal of structural biology*.
84. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. (2003) The MerR family of transcriptional regulators, *FEMS microbiology reviews*. **27**, 145-163.
85. Huang, H., Pandey, C., Liua, C., Al-Obaidid, N. F., Wang, M., Zheng, L., Sarah Toews Keatinga, Aonob, M., Loved, J. D., Evansd, B., Seideld, R. D., Hillerichd, B. S., Garforthd, S. J., Almod, S. C., Marianoa, P. S., Dunaway-Marianoa, D., Allenb, K. N. & Farellib, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling, *PNAS*, 1974–1983.
86. Khachane, A. N., Timmis, K. N. & dos Santos, V. A. (2005) Uracil content of 16S rRNA of thermophilic and psychophilic prokaryotes correlates inversely with their optimal growth temperatures, *Nucleic acids research*. **33**, 4016-22.
87. Russell, N. J. Psychophilic Bacteria-Molecular Adaptations of Membrane Lipids, *Camp Biochem Physiol Vol 118A, No 3, pp , 1997*. **118A**, 489-493.
88. Scheeff, E. & Bourne, P. (2005) Structural Evolution of the Protein Kinase-Like Superfamily, *PLoS Computational Biology*. **preprint**, e49.
89. Taylor, S. S. & Radzio-Andzelm, E. (1994) Three protein kinase structures define a common motif, *Structure*. **2**, 345–355.
90. Fagerbakke, K. M., Norland, S. & Haldal, M. (1999) The inorganic ion content of native aquatic bacteria, *Can J Microbiol*. **45**, 304-311.
91. Haldal, M., Norland, S., Erichsen, E. S., Sandaa, R. A., Larsen, A., Thingstad, F. & Bratbak, G. (2012) Mg²⁺ as an indicator of nutritional status in marine bacteria, *The ISME journal*. **6**, 524-30.

92. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond, *FEMS microbiology reviews*. **29**, 231-62.
93. Godsey, M. H., Zheleznova Heldwein, E. E. & Brennan, R. G. (2002) Structural biology of bacterial multidrug resistance gene regulators, *The Journal of biological chemistry*. **277**, 40169-72.
94. Newberry, K. J. & Brennan, R. G. (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus, *The Journal of biological chemistry*. **279**, 20356-62.
95. Heldwein, E. E. Z. & Brennan, R. G. (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug, *NATURE* **409**, 378-382.
96. Fisher, S. H. (1999) Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la différence!, *Molecular microbiology*. **32**, 223–232.
97. Hauf, K., Kayumov, A., Gloge, F. & Forchhammer, K. (2016) The Molecular Basis of TnrA Control by Glutamine Synthetase in *Bacillus subtilis*, *The Journal of biological chemistry*. **291**, 3483-95.
98. Fisher, S. H. & Wray, L. V., Jr. (2008) *Bacillus subtilis* glutamine synthetase regulates its own synthesis by acting as a chaperone to stabilize GlnR-DNA complexes, *Proceedings of the National Academy of Sciences of the United States of America*. **105**, 1014-9.
99. Wray, L. V., Jr. & Fisher, S. H. (2007) Functional analysis of the carboxy-terminal region of *Bacillus subtilis* TnrA, a MerR family protein, *Journal of bacteriology*. **189**, 20-7.
100. Ogasawara, H., Yamamoto, K. & Ishihama, A. (2010) Regulatory role of MlrA in transcription activation of csgD, the master regulator of biofilm formation in *Escherichia coli*, *FEMS microbiology letters*. **312**, 160-8.
101. Gualdi, L., Tagliabue, L. & Landini, P. (2007) Biofilm formation-gene expression relay system in *Escherichia coli*: modulation of sigmaS-dependent gene expression by the CsgD regulatory protein via sigmaS protein stabilization, *Journal of bacteriology*. **189**, 8034-43.
102. Burkhard, P., Stetefeld, J. & Strelkov, S. V. (2001) Coiled coils: a highly versatile protein folding motif, *TRENDS in Cell Biology* **11**, 82-88.
103. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999) Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques, *Biochemistry* **38**, 16424-16431.

104. Wray, L. V., Jr. & Fisher, S. H. (2008) Bacillus subtilis GlnR contains an autoinhibitory C-terminal domain required for the interaction with glutamine synthetase, *Molecular microbiology*. **68**, 277-85.
105. Tuerk, C. & Gold, L. (1990) Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase, *Science*. **249**, 505-510.
106. Bailey, T. L. & C, E. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36.
107. Strauch, M. A. (2001) Protein–DNA Complexes: Specific in *eLS*, John Wiley & Sons, Ltd.
108. Fedorova, K. P., Scharafutdinov, I. S., Turbina, E. Y., Bogachev, M. I., Ilinskaja, O. N. & Kayumov, A. R. (2013) C-terminus of transcription factor TnrA from Bacillus subtilis controls DNA-binding domain activity but is not required for dimerization, *Molecular Biology*. **47**, 293-298.
109. Hobman, J. L. (2007) MerR family transcription activators: similar designs, different specificities, *Molecular microbiology*. **63**, 1275-1278.
110. Uhlich, G. A., Chen, C. Y., Cottrell, B. J., Hofmann, C. S., Dudley, E. G., Strobaugh, T. P., Jr. & Nguyen, L. H. (2013) Phage insertion in mlrA and variations in rpoS limit curli expression and biofilm formation in Escherichia coli serotype O157: H7, *Microbiology*. **159**, 1586-96.
111. Froger, A. & Hall, J. E. (2007) Transformation of plasmid DNA into E. coli using the heat shock method, *Journal of visualized experiments : JoVE*, 253.
112. G, S. & SS, S. (1990) The "megaprimer" method of site-directed mutagenesis., *Biotechniques* 1990 Apr;8(4):404-7. **8**, 404-407.
113. Eibl, H. & Lands, W. E. M. (1968) A new, sensitive determination of phosphate, *Analytical Biochemistry*. **30**, 51-57.
114. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *Journal of biomolecular NMR*. **6**, 277-293.
115. Johnson, B. A. (2004) Using NMR View to Visualize and Analyze the NMR Spectra of Macromolecules, *Methods in Molecular Biology: Protein NMR Techniques*. **278**, 313-352.
116. Merrill, M. R. (1993) NMR Diffusion Measurements Using a Composite Gradient PGSE Sequence, *Journal of Magnetic Resonance*. **103**, 223-225.
117. SG, S. & AK, S. (1961) Determination of the effective hydrodynamic radii of small molecules by viscometry, *J Gen Physiol*. **44**, 1189-1199.

-
118. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins*. **65**, 712-25.
119. Cerutti, D. S., Swope, W. C., Rice, J. E. & Case, D. A. (2014) ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins, *Journal of chemical theory and computation*. **10**, 4515-4534.
120. Price, D. J. & Brooks, C. L., 3rd (2004) A modified TIP3P water potential for simulation with Ewald summation, *The Journal of chemical physics*. **121**, 10096-103.
121. WF, V. G. & HJ, B. (1984) Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. , *J Mol Biol* **176**, 559-564.
122. W, S. & TR, F. (1996) DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package., *J Mol Graph* **14**, 136-141.
123. TA, D. & LG, P. (1993) Molecular modeling: an experimental tool., *Environ Health Perspect* **101**, 410-412.
124. W, H., A, D. & K, S. (1996) VMD: visual molecular dynamics., *J Mol Graph*. **14**, 27-38.
125. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta crystallographica Section D, Biological crystallography*. **66**, 12-21.