

ANÁLISIS DE LA VIRULENCIA DE PROTEÍNAS MULTIFUNCIONALES MEDIANTE BIOINFORMÁTICA

Araceli Nájar García

Máster universitario en Bioinformática y bioestadística UOC-UB
Área 2

Luis Franco Serrano
Carles Ventura Royo

05/06/2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Análisis de la virulencia de proteínas multifuncionales mediante bioinformática</i>
Nombre del autor:	<i>Araceli Nájar García</i>
Nombre del consultor/a:	<i>Luis Franco Serrano</i>
Nombre del PRA:	<i>Carles Ventura Royo</i>
Fecha de entrega (mm/aaaa):	06/2019
Titulación:::	<i>Máster universitario en Bioinformática y bioestadística UOC-UB</i>
Área del Trabajo Final:	<i>Área 2</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Moonlighting; Virulencia; Proteómica</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>Las proteínas moonlighting o multifuncionales son aquellas capaces de realizar más de una función. Un 25% de las proteínas moonlighting descritas están implicadas en la virulencia de patógenos. Sin embargo, actualmente no se ha determinado la existencia de motifs o patrones comunes entre estas proteínas que puedan relacionarse con su papel en la virulencia. Por ello, el objetivo principal de este trabajo ha sido analizar bioinformáticamente las proteínas moonlighting implicadas en virulencia. Concretamente, la metodología empleada ha consistido en la búsqueda de motifs comunes, en el análisis de los partners de unión de estas proteínas y en el estudio de las regiones de las mismas implicadas en la interacción con el plasminógeno. Uno de los resultados obtenidos ha sido un listado de 240 proteínas moonlighting implicadas en virulencia, compuesto fundamentalmente por enzimas del metabolismo primario de diversos tipos de microorganismos y con la función moonlighting mayoritaria de unión a proteínas del huésped. También se han obtenido motifs comunes en algunas de estas proteínas de microorganismos patógenos, un programa para la búsqueda de los mismos, un análisis de la localización de estos en cada secuencia proteica y un listado de aminoácidos implicados en la interacción de ciertas proteínas moonlighting con el plasminógeno. Con este trabajo se puede concluir que algunas enolasas y GAPDH multifuncionales presentan motifs comunes, en algunos casos cercanos o involucrados en las zonas de interacción de estas proteínas con el plasminógeno. Además, gran parte de estas proteínas moonlighting suele interaccionar con los mismos residuos del plasminógeno.</p>	

Abstract (in English, 250 words or less):

Moonlighting or multifunctional proteins are able to execute more than one function. Nearly 25% of the moonlighting proteins described in databases are involved in the virulence of pathogens. However, the existence of common motifs or patterns among these virulence-related proteins has not been determined yet. Therefore, the main goal of this project has been to analyze bioinformatically the moonlighting proteins involved in virulence. Specifically, the methodology of this work consisted in searching common motifs, analyzing binding partners of these proteins and studying regions involved in the interaction with plasminogen. One of the main results has been a list of 240 moonlighting proteins related to virulence, mostly enzymes from the primary metabolism of several microorganisms which main moonlighting function is binding to host proteins. The results also include common motifs in some of these proteins, a program to search them, an analysis of the location of these in each protein sequence and a list of amino acids involved in the interaction of certain moonlighting proteins with plasminogen. To summarize, the results of this work suggest that some enolases and multifunctional GAPDH have common motifs, sometimes close to or involved in the areas of interaction between these proteins and plasminogen. In addition, some of these moonlighting proteins usually interact with the same plasminogen residues.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	1
A) Objetivos generales.....	1
B) Objetivos específicos.....	2
1.3 Enfoque y método seguido.....	2
1.4 Planificación del Trabajo.....	3
1.5 Breve resumen de productos obtenidos.....	3
1.6 Breve descripción de los otros capítulos de la memoria.....	4
2. Resto de capítulos.....	5
2.1 INTRODUCCIÓN.....	5
A) Proteínas moonlighting: concepto e introducción.....	5
B) Historia de las proteínas moonlighting.....	6
C) Características de las proteínas moonlighting.....	7
D) Ejemplos de las principales proteínas moonlighting.....	10
E) Aspectos evolutivos de las proteínas moonlighting.....	11
F) Proteínas moonlighting y enfermedades humanas.....	12
G) Proteínas moonlighting y virulencia.....	13
H) Identificación y predicción de proteínas moonlighting.....	16
2.2 MÉTODOS.....	18
A) Bases de datos, servidores y herramientas bioinformáticas utilizados.....	18
B) Selección de un listado de proteínas moonlighting asociadas a la virulencia.....	20
C) Búsqueda y análisis de motivos en las proteínas moonlighting seleccionadas.....	21
D) Creación de un programa de búsqueda de motivos en proteínas.....	23
E) Análisis de interacción de proteínas moonlighting con sus partners de unión.....	24
F) Dockings estructurales entre proteínas moonlighting y el plasminógeno.....	24
2.3 RESULTADOS.....	26
A) Búsqueda y análisis de proteínas moonlighting implicadas en virulencia.....	26
B) Estudio de motivos en proteínas moonlighting implicadas en virulencia.....	31
C) Programa de búsqueda de motivos en secuencias proteicas.....	40
D) Análisis interactómico de proteínas moonlighting implicadas en virulencia.....	42
E) Unión de proteínas moonlighting al plasminógeno.....	45
2.4 DISCUSIÓN.....	50
3. Conclusiones.....	55
4. Glosario.....	56
5. Bibliografía.....	57
6. Anexos.....	60

Lista de tablas

Tabla 1. Principales funciones canónicas de las proteínas moonlighting del listado y porcentaje de estas proteínas que presentan dicha función.....	27
Tabla 2. Principales proteínas a las que se unen las proteínas moonlighting del listado y número de proteínas moonlighting que tiene dicha proteína de unión (según la información sobre sus respectivas funciones moonlighting).....	28
Tabla 3. Motifs que se encuentran exclusiva o mayoritariamente en microorganismos patógenos.....	33
Tabla 4. Posición de los 6 motifs encontrados en GAPDH dentro de la secuencia de aminoácidos de cada proteína.	36
Tabla 5. Posición de los 11 motifs encontrados en enolasas dentro de la secuencia de aminoácidos de cada proteína.	37
Tabla 6. Enolasas con residuos de los motifs analizados con anotaciones en el servidor InterPro.....	39
Tabla 7. Estrategias de búsqueda de motifs en el programa desarrollado.....	41
Tabla 8. Proteínas moonlighting con motifs exclusivos/mayoritarios de patógenos que interactúan con el plasminógeno. Se indica el tipo de proteína, el microorganismo en el que se encuentra y el nombre de la base de datos en la que está registrada su interacción con el plasminógeno.....	43
Tabla 9. Listado de proteínas moonlighting enolasas y GAPDH que se unen a determinados residuos del plasminógeno.	46
Tabla 10. Aminoácidos implicados según el modelado con ClusPro 2.0 y el análisis con PyMol en la interacción plasminógeno-enolasas.....	47
Tabla 11. Aminoácidos implicados según el modelado con ClusPro 2.0 y el análisis con PyMol en la interacción plasminógeno-GAPDH.	48

Lista de figuras

Figura 1. Principales causas de la actividad moonlighting (modificado de 1).	8
Figura 2. Esquema del sistema del plasminógeno y sus funciones (36).....	15
Figura 3. Ejemplo de la interacción entre un microorganismo patógeno y el sistema fibrinolítico (36).	16
Figura 4. Esquema del flujo de trabajo seguido en el programa de búsqueda de motifs.	24
Figura 5. Representación gráfica de los principales tipos de proteínas que componen el listado de proteínas moonlighting.....	27
Figura 6. Representación esquemática de las principales proteínas de unión que comparte una proteína moonlighting del listado. El número entre paréntesis indica el número de proteínas moonlighting que se unen a cada tipo de proteína y el número en negrita representa el número de proteínas moonlighting que se unen a ambas proteínas.	29
Figura 7. Representación gráfica de los organismos en los que se encuentran las proteínas moonlighting del listado.....	30
Figura 8. Dominios encontrados con InterPro para las proteínas GAPDH.....	36
Figura 9. Dominios encontrados con InterPro para las proteínas enolasas.	36

Figura 10. Anotación de residuos para enolasas en InterPro. En rojo, residuos del motif 'IEDG'. En verde, residuos del motif 'QLLRI'. En azul, residuos del motif 'RTDR'.	39
Figura 11. Captura de pantalla del resultado de la búsqueda de la proteína P75189 en el servidor APID. La flecha en rojo señala el registro de la interacción de esta proteína con el plasminógeno humano.	43
Figura 12. A la izquierda, ejemplo del modelado del docking estructural del plasminógeno y la proteína P75189 obtenido con ClusPro 2.0. A la derecha, análisis con PyMol de una interacción del mismo modelado indicándose los aminoácidos involucrados tanto del plasminógeno (molécula verde) como la proteína P75189 (en azul).....	46

1. Introducción

1.1 Contexto y justificación del Trabajo

En este Trabajo Fin de Máster, titulado “Análisis de la virulencia de proteínas multifuncionales mediante bioinformática”, se va a realizar un análisis bioinformático de proteínas moonlighting o multifuncionales, enfocado a estudiar patrones comunes de diversas proteínas multifuncionales asociadas a la actividad virulenta de microorganismos patógenos y a la unión patógeno-huésped que sucede durante el proceso de infección. El propósito final de este trabajo es ampliar el conocimiento sobre la acción infecciosa de estos microorganismos y, a su vez, descubrir posibles regiones o secuencias que sirvan como dianas vacunales.

Las proteínas moonlighting, también denominadas multifuncionales, son aquellas capaces de realizar más de una función. Existen proteínas moonlighting tanto en microorganismos patógenos y no patógenos aunque, curiosamente, se ha descubierto que un 25% de las proteínas multifuncionales descritas hasta ahora se encuentran implicadas en los mecanismos de virulencia de los microorganismos patógenos. Actualmente no se sabe si existen motivos exclusivos de proteínas multifuncionales de patógenos que no se encuentren en no patógenos (1,2). Tampoco se ha confirmado si las proteínas multifuncionales asociadas a virulencia comparten patrones comunes que puedan servir como indicios para clasificar a una proteína como multifuncional y vinculada a virulencia.

Todo esto pone de manifiesto la importancia de disponer de herramientas bioinformáticas que permitan la predicción, búsqueda y análisis de proteínas multifuncionales, no solo para comprender mejor el proceso infeccioso y la virulencia de algunos patógenos, sino también para crear vacunas e identificar dianas farmacológicas.

1.2 Objetivos del Trabajo

A) Objetivos generales

El **objetivo general** del Trabajo Fin de Máster es el siguiente:

- 1) Analizar bioinformáticamente patrones comunes en proteínas multifuncionales ligadas a la virulencia y a la unión de microorganismos patógenos al huésped para ampliar la comprensión sobre el proceso infeccioso de estos microorganismos y descubrir posibles dianas de vacunas.

B) Objetivos específicos

Para conseguir este objetivo general del Trabajo Fin de Máster, se concretan una serie de objetivos específicos:

- 1) Identificar patrones comunes entre las diferentes proteínas multifuncionales relacionadas con virulencia y unión patógeno-huésped.
- 2) Analizar proteínas multifuncionales que se encuentren tanto en organismos patógenos como no patógenos.
- 3) Seleccionar motifs en las proteínas multifuncionales que posean únicamente los organismos patógenos.
- 4) Descubrir si las proteínas multifuncionales se unen siempre a las mismas regiones de las proteínas del huésped.
- 5) Contrastar la relevancia de estos motifs en la unión del patógeno al huésped y en la regulación del sistema inmune.
- 6) Analizar mediante interactómica los partners de unión al plasminógeno.
- 7) Identificar proteínas multifuncionales que interaccionen con el plasminógeno y analizar con qué regiones del mismo interaccionan.
- 8) Descubrir partners de interacción compartidos por las proteínas multifuncionales seleccionadas.

1.3 Enfoque y método seguido

Para cumplir los objetivos establecidos en el apartado anterior, se lleva a cabo una estrategia que consiste, a grandes rasgos, en seleccionar un conjunto de proteínas multifuncionales (seleccionadas de la base de datos http://wallace.uab.es/multitaskII/proteins_list.php) y estudiar la virulencia de las mismas haciendo uso de una serie de herramientas bioinformáticas.

Por un lado, se busca obtener un listado de proteínas moonlighting que presenten funciones moonlighting implicadas en virulencia y en la unión a proteínas del huésped. Tras esto se pretende analizar motifs comunes de aquellas proteínas moonlighting que se encuentran en microorganismos no patógenos y patógenos, respectivamente, así como estudiar regiones comunes entre las proteínas seleccionadas. Además, se quiere encontrar motifs que se den en proteínas de organismos patógenos de manera exclusiva o de manera mayoritaria, para conocer si existen algunos motifs relacionados con las funciones de virulencia de algunas proteínas moonlighting. De manera adicional, con el objetivo de comparar si un motif se encuentra mayoritariamente en proteínas de patógenos, se va a elaborar un programa que permita obtener un listado de proteínas que presenten un motif determinado y a qué organismo pertenecen cada una.

En este trabajo también se va a estudiar la unión de las proteínas multifuncionales y las proteínas con las que interaccionan, como por ejemplo el plasminógeno. Además, para alcanzar algunos de los objetivos planteados se van a estudiar partners de unión utilizando bases de datos de interactómica y realizando dockings estructurales.

1.4 Planificación del Trabajo

El trabajo se ha planificado dividiendo cada objetivo en una serie de tareas. La elaboración y cumplimiento de estas tareas se ha repartido fundamentalmente en dos fases del proyecto (PEC2 y PEC3). La primera fase del proyecto ha consistido en la búsqueda de artículos e información bibliográfica para recopilar datos y aumentar el conocimiento sobre el ámbito de las proteínas moonlighting. Además, se ha llevado a cabo la elaboración del listado de proteínas moonlighting y la descripción del mismo, junto con la búsqueda de motifs y el análisis de las proteínas en los que se han encontrado estos motifs.

En la segunda parte del proyecto se han realizado los análisis correspondientes al estudio interactómico, a la elaboración del programa de búsqueda de motifs, al modelado de dockings estructurales y al análisis de interacciones entre proteínas moonlighting y el plasminógeno. Esta fase también se ha compaginado con la redacción de la memoria.

La planificación a modo de diagrama se muestra en el Material Suplementario 0 (MS0).

1.5 Breve resumen de productos obtenidos

Los principales resultados obtenidos con la realización del trabajo son los siguientes:

- Listado y análisis de un conjunto de proteínas moonlighting asociadas a la virulencia.
- Análisis de las funciones moonlighting de estas proteínas y clasificación de las mismas según a las proteínas del huésped a las que se unen y a los microorganismos en los que se encuentran.
- Búsqueda y obtención de motifs comunes entre algunas de estas proteínas moonlighting, así como localización de los mismos en las secuencias proteicas.
- Elaboración de un programa en Python para la búsqueda de motifs en proteínas.
- Análisis de la interactómica de algunas proteínas moonlighting de patógenos.
- Estudio de la interacción de proteínas moonlighting con el plasminógeno e identificación de los aminoácidos implicados.

1.6 Breve descripción de los otros capítulos de la memoria

La primera parte de los capítulos del trabajo consiste en una introducción a las proteínas moonlighting, en la que se definen, se narra parte de la historia de estas proteínas y se explican sus principales características. Además, se exponen los principales criterios para considerar que una proteína es moonlighting y qué aspectos hacen que no se trate de un verdadero caso de multifuncionalidad. También se mencionan algunos ejemplos de proteínas moonlighting y se pone en relación con este tipo de proteínas dos ámbitos de interés: las enfermedades humanas y la virulencia de patógenos. Para finalizar este capítulo, se habla de las estrategias y retos de la identificación y predicción de proteínas moonlighting.

En segundo lugar, se exponen los métodos que han sido empleados para la realización de este trabajo, tanto las bases de datos y herramientas utilizadas como el proceso seguido en cada uno de los análisis del proyecto.

En el tercer capítulo se muestran y detallan todos los resultados obtenidos en la realización del trabajo, exponiéndose los gráficos, figuras y tablas correspondientes a cada uno de los análisis efectuados.

Por último, se realiza una discusión general para comentar e interpretar los diferentes resultados del trabajo, tratándose también las principales conclusiones que se obtienen del proyecto, algunas de las limitaciones o mejoras que se consideran interesantes y las perspectivas futuras en relación a este ámbito.

2. Resto de capítulos

2.1 INTRODUCCIÓN

A) Proteínas moonlighting: concepto e introducción

Las proteínas moonlighting, también denominadas multifuncionales, son proteínas que realizan más de una función biológica en diferentes compartimentos celulares del organismo (1–6). En este tipo de proteínas, una única cadena polipeptídica realiza dos o más funciones bioquímicas o biofísicas relevantes a nivel fisiológico y no relacionadas entre sí (4,7–10). Las funciones moonlighting no se encuentran necesariamente separadas en dos dominios proteicos distintos (1,11). Esta multifuncionalidad no se debe a la fusión de genes, ni a la generación de variantes de proteínas por splicing alternativo o fragmentos de una proteína que expresen diferentes funciones tras la proteólisis (1,2,8,10). Tampoco se consideran proteínas moonlighting aquellas homólogas pero no idénticas (proteínas con un mismo origen evolutivo que realizan funciones distintas), proteínas que sean factores de transcripción con capacidad de unión a diferentes promotores, ni enzimas que participen en diferentes procesos metabólicos o con una diversa gama de sustratos (1,3,5,6,9).

Una proteína moonlighting posee una función canónica y una o varias funciones moonlighting. La función canónica es la primera función descrita, mientras que las funciones moonlighting son aquellas descubiertas con posterioridad. Erróneamente, se suele considerar que la función canónica es la principal mientras que las funciones moonlighting son accesorias o secundarias, ya que la función canónica suele ser una función más primaria y más vital para la célula y, además, se descubre antes que las funciones moonlighting. Sin embargo, esta clasificación de las funciones de una proteína moonlighting no es indicativa de la relevancia funcional de las mismas, sino que se debe simplemente al orden histórico de descubrimiento (1). Por lo general, las funciones canónicas de las proteínas multifuncionales suelen estar involucradas en procesos celulares esenciales como, por ejemplo, metabolismo, estabilidad de los ácidos nucleicos, síntesis de proteínas o actividad chaperona; mientras que las funciones moonlighting se corresponden con funciones más complejas y de adquisición posterior a nivel evolutivo, como unión a células fagocíticas, a elementos del citoesqueleto, a mucinas, a proteínas del sistema inmune o hemostático, etc. (1,2,6,10). El hecho de que una proteína adquiera una segunda función no implica la pérdida de la función canónica (1).

A pesar de que las proteínas moonlighting se han estudiado fundamentalmente en bacterias (se han descrito numerosas proteínas moonlighting en bacterias patógenas, no patógenas y comensales, así como bacterias Gram + y Gram -), también se ha descrito este tipo de proteínas en mamíferos, levaduras, helmintos, bacterias, plantas, virus, arqueas y otros muchos organismos (2,6,8). A día de hoy existen cientos de proteínas moonlighting identificadas, concretamente un total de 694 proteínas registradas

en la base de datos MultitaskProtDB-II, aunque el número va aumentando cada día (3,6,11).

B) Historia de las proteínas moonlighting

Las proteínas moonlighting son de reciente descubrimiento, ya que hasta finales de los años 80 no se describieron las primeras. Por lo general, el descubrimiento de este tipo de proteínas es por casualidad, al identificar que una proteína que posee una función determinada es también una proteína anteriormente conocida que realiza otra función, que suele ser ancestral y más básica. Lo más común es que esta función ancestral se encuentre relacionada con el metabolismo primario (1,3,7).

La primera proteína moonlighting identificada como tal fue la aldehído deshidrogenasa, en el año 1987. Este hallazgo se llevó a cabo por Wistow y Piatigorsky, quienes mostraron que esta proteína no solo realiza una función estructural en la lente del ojo, sino que también posee actividad enzimática (2,7). Al igual que esta proteína, se han identificado otras proteínas moonlighting presentes en el cristalino de algunos organismos como proteínas estructurales y que a su vez son enzimas citosólicas, como la lactato deshidrogenasa, arginosuccinato liasa o α -enolasa (5,9,11).

En 1988, se estableció el carácter moonlighting de la neuroleucina, previamente definida como una citoquina, al descubrirse su actividad enzimática como glucosa-6-fosfato isomerasa (1,2). En bacterias, la primera proteína moonlighting identificada fue la gliceraldehído-3-fosfato deshidrogenasa (GAPDH). Esta es una enzima glicolítica que se encuentra tanto en el citoplasma como en la superficie celular de los *Streptococcus* del grupo A (5,12).

A lo largo de los años se han propuesto diversos nombres para las proteínas moonlighting. Piatigorsky introdujo el concepto *gene sharing* o gen compartido (7,13). Este término no es adecuado, ya que tanto en las proteínas moonlighting como en las diferentes proteínas obtenidas por splicing alternativo se comparte un mismo gen y, sin embargo, son situaciones diferentes. Con el fenómeno de splicing se obtienen diferentes polipéptidos con sus funciones correspondientes, mientras que una proteína moonlighting es un único polipéptido con varias funciones (1).

Otro término que puede suscitar confusión es el de *proteínas promiscuas*, introducido por Nobeli (14). Una proteína promiscua es aquella que realiza una función molecular que interviene en muchas funciones biológicas. Un ejemplo de este concepto sería una proteína quinasa, cuya función es fosforilar diferentes proteínas diana, implicadas a su vez en distintas rutas metabólicas. Sin embargo este tipo de proteínas no pueden considerarse proteínas moonlighting, ya que realizan una única función (1,3).

Un término más restrictivo, acuñado por Jeffery, es el de *moonlighting* (15), en el que se excluye la fusión génica como causa de la segunda función de una proteína y el hecho de que se presenten dos funciones moleculares en el

mismo centro activo (1,4,7). Esta definición de proteínas moonlighting es tan estricta y limitativa que muy pocos investigadores la aceptan o aplican. Además se trata de un concepto que presenta gran complejidad a la hora de ser determinado, pues habría que conocer la historia evolutiva de la proteína en cuestión para saber si proviene de mutaciones o de una fusión de genes más o menos cercana en el tiempo (1,3). La historia evolutiva de una proteína, sea una fusión génica o no, no cambia el hecho de que haya desembocado en una proteína con la capacidad de realizar varias funciones (4).

De manera adicional, Jeffery y los autores más restrictivos diferencian entre los conceptos “*moonlighting*” y “*multitasking*” (proteínas multitarea o multifuncionales). Jeffery propone que tan solo las proteínas que cumplan la definición estricta de *moonlighting* se definan como tal, mientras que las proteínas fruto de una fusión genes o con dos actividades enzimáticas en un centro activo sean denominadas multifuncionales (1). En este trabajo se empleará el concepto más amplio de proteínas moonlighting, aquel que engloba ambos casos (tanto moonlighting como multitasking).

Algunos autores sugieren que las proteínas moonlighting son las denominadas proteínas intrínsecamente desestructuradas/desordenadas (*intrinsically disordered proteins, IDPs*), proteínas que interactúan con muchos partners distintos y que, por tanto, dan lugar a múltiples funciones (1,7,16). Sin embargo, se ha demostrado que la mayoría de proteínas multifuncionales no pertenecen al tipo de proteínas IDPs (17).

C) Características de las proteínas moonlighting

La multifuncionalidad suele estar asociada a algunas características. Se han descubierto varias causas de la actividad moonlighting de algunas proteínas, entre las que destacan:

- Localización de la proteína en más de un compartimento celular o presencia de la proteína en una región de la célula en la que no realiza su función canónica (Figura 1a). Por ejemplo, la proteína PutA de *E. coli* actúa como enzima asociada a la membrana plasmática y como factor de transcripción en el citoplasma (1,6,9).
- Una proteína se encuentra secretada cuando su función canónica es claramente intracelular (Figura 1b). Un ejemplo de esto es la neuroleucina mencionada en el apartado anterior, que es una enzima de la glucólisis en el citoplasma y una citoquina cuando es secretada (1,6,7).
- Expresión diferencial de una proteína, realizando distintas funciones según el factor que la haya activado o según el tipo de célula que la exprese (Figura 1c). Por ejemplo la proteína neuropilina puede estimular la producción de células sanguíneas o promover el crecimiento correcto del axón en función del factor que la activa (1,9).
- Diferente estado de oligomerización de una proteína o formación de complejos, que causan que una misma proteína tenga distintas funciones (Figura 1d). Este es el caso de la piruvato quinasa, que actúa como

- quinasa al encontrarse como un tetrámero y como factor de unión a hormonas tiroideas en estado monomérico (1,7,9).
- Realización de diferentes funciones según la concentración de algún metabolito (Figura 1e). Por ejemplo, la aconitasa interviene en el ciclo de Krebs, pero ante una concentración elevada de hierro es una proteína *Iron-Responsive Element-Binding* (1,5,9).
 - Distintos sitios de unión o partners de unión inesperados (Figura 1f). Un ejemplo es la proteína Arg5, que se une a varias regiones del ADN y es tanto una quinasa como un factor de transcripción (2,5,11).
 - Multifuncionalidad ocasionada por la modificación postraduccion de la proteína (Figura 1g). La enzima glucosa-6-fosfato isomerasa sin fosforilar actúa como enzima, mientras que tras ser fosforilada es un factor autocrino de motilidad (2,7).
 - Proteínas que, de manera ocasional, presentan una función alternativa al anclarse a la membrana y unirse a un ácido graso (Figura 1h) (7).

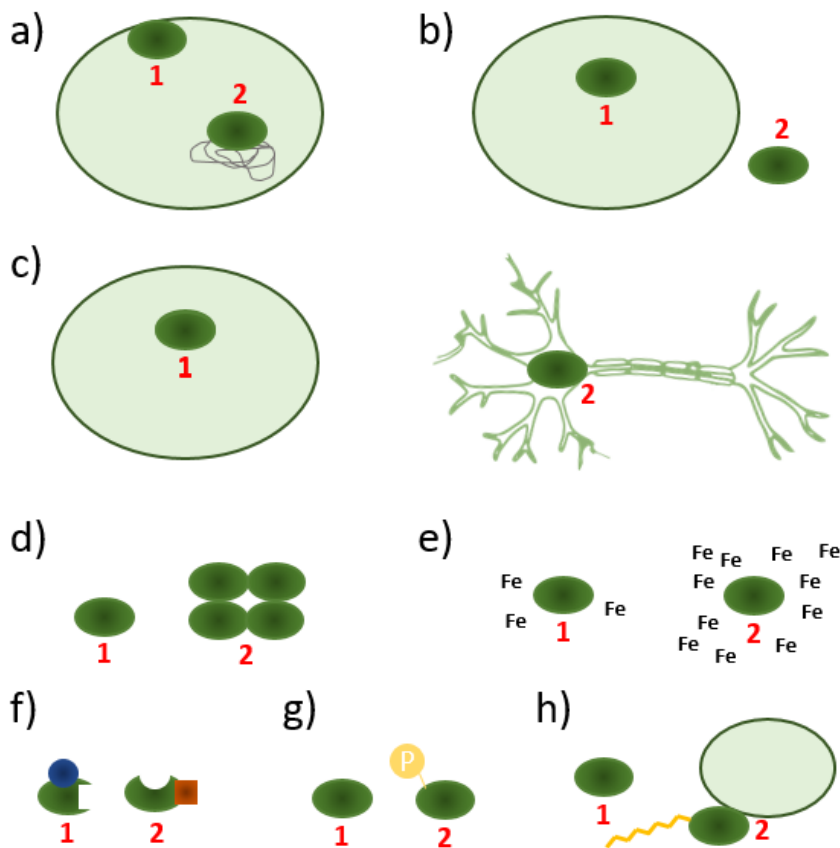


Figura 1. Principales causas de la actividad moonlighting (modificado de 1).

Partiendo de la definición de proteína moonlighting que se ha establecido anteriormente, existe una serie de casos en los que una proteína **no puede considerarse como moonlighting**:

- Variantes de una proteína producidas por splicing alternativo, ya que son polipéptidos diferentes (8,10).

- Proteínas con diferentes funciones biológicas producto de la proteólisis de un proteína original (1,10).
- Enzimas con una amplia gama de sustratos (11).
- Factores de transcripción que se unen a diferentes promotores (1,3).
- Enzimas que intervienen en diversas rutas metabólicas mediante la misma función molecular (7).
- Una proteína producto de la fusión de dos genes (únicamente para algunos autores, como Jeffery).
- Proteínas con dos actividades enzimáticas en un mismo centro activo (tan solo para algunos autores más restrictivos).

Las proteínas moonlighting tienen un papel relevante en la bioquímica de proteínas y en investigación. Por ejemplo, la multifuncionalidad de algunas proteínas genera imprecisiones o insuficiencias en la anotación genómica funcional. El fenómeno moonlighting puede afectar al análisis y predicción de metabolitos y rutas metabólicas. Asimismo, puede aumentar la complejidad a la hora de interpretar experimentos de delección o silenciamiento de genes (ya que los efectos pueden deberse a más de una función) o redes de interactómica (1,2).

Por otro lado, otra característica importante de las proteínas moonlighting es el papel de muchas de ellas en microorganismos patógenos como factor de virulencia o su relación con enfermedades humanas (2,6).

- **Clases funcionales de las proteínas moonlighting**

En cuanto a las clases funcionales que presentan las proteínas moonlighting, pese a ser variadas por lo general la mayoría suelen ser enzimas y proteínas de unión a ácidos nucleicos (1,11,18). De hecho, en una de las bases de datos más completa de registro de proteínas moonlighting, estas dos funciones mencionadas son la combinación de funciones más abundantes entre todas las enzimas incluidas en la base de datos (19). El segundo conjunto de funciones más habitual es una enzima que a su vez está implicada en procesos de adhesión celular. No obstante, estas no son las únicas categorías funcionales que pueden desempeñar las proteínas moonlighting, sino que también pueden ser, por ejemplo, chaperonas, citoquinas, proteínas del citoesqueleto, canales, proteínas ribosómicas, proteínas de andamiaje, receptores, reguladores transcripcionales, proteínas de unión a receptores, etc. (8,9,17).

Las proteínas moonlighting se caracterizan por poder desarrollar estas funciones en diferentes compartimentos de la célula, en diferentes estadios del ciclo celular o incluso pueden llevar a cabo las distintas funciones al mismo tiempo (2,3).

Por otro lado, un bajo número de las proteínas multifuncionales identificadas a día de hoy son proteínas de membrana. Esto puede deberse a complicaciones en el proceso de plegamiento al presentarse algunos dominios transmembrana. Sin embargo, en microorganismos patógenos se ha observado el curioso fenómeno de que algunas proteínas moonlighting, implicadas en procesos de adhesión del patógeno al huésped y situadas inicialmente en el

citósol o en el núcleo, se transportan a la membrana celular del microorganismo sin presentar péptido señal para participar en esta función de adhesión (1,2). Las secuencias de estas proteínas moonlighting no contienen ningún motivo específico conocido de translocación, exportación, importación o anclaje (6). La manera en la que las proteínas moonlighting se transportan al exterior de la célula no ha sido resuelta aún. Algunas de las opciones que se consideran son que las proteínas sean liberadas de células dañadas y se unan a la superficie de células vecinas, o bien que sean secretadas a la superficie celular a través de algún mecanismo que todavía no se ha descubierto (2). Pese a que este mecanismo aún no se conoce en profundidad, sí se han identificado algunos ejemplos de esto. Por ejemplo, se sabe que muchos microorganismos utilizan la enolasa para unirse al plasminógeno del huésped o la histona H1 para unirse a la tiroglobulina (1).

Además, las proteínas moonlighting pueden jugar un papel importante a la hora de coordinar diferentes actividades celulares. Pueden conseguir esta coordinación, por ejemplo, facilitando la conexión entre rutas metabólicas, regulándolas, generando cambios o interacciones entre unas vías u otras, etc. De manera adicional, otra característica de las proteínas moonlighting es que un gran número de ellas presentan regulación constitutiva (*housekeeping genes*), debido posiblemente a ser proteínas del metabolismo primario (11).

D) Ejemplos de las principales proteínas moonlighting

Actualmente, la enzima GAPDH es la proteína moonlighting con mayor número de funciones identificadas, un total de 18 funciones en el conjunto de organismos en los que se encuentra (1). En bacterias es una de las proteínas moonlighting mejor caracterizadas, junto con la enolasa. Dentro de las diversas funciones que posee, una de ellas es la adhesión a varios componentes del huésped, como antígenos séricos, proteínas del citoesqueleto, albúmina, fibrinógeno, plasminógeno, proteínas de la membrana extracelular o mucinas (2). La GAPDH también está involucrada en la regulación de la fosforilación de proteínas celulares, en la patogénesis de algunos microorganismos, en la evasión del sistema inmune del huésped y en la comunicación celular entre el patógeno y el huésped.

De igual forma, la enolasa también presenta funciones moonlighting muy importantes. Se caracteriza por, fundamentalmente en bacterias, actuar como un receptor del plasminógeno del huésped, unirse a células epiteliales y facilitar la evasión del sistema inmune del huésped (2,3).

Otra de las proteínas moonlighting con más funciones es la proteína p53, la que no solo presenta diferentes funciones descritas, sino también numerosas interacciones con proteínas reguladoras (11). Más proteínas moonlighting con muchas funciones descritas son, por ejemplo: aconitasa, factor de elongación Tu, ubiquitina, Cpn10, dihidropoliámidasa deshidrogenasa, Hsp70, Hsp90 y HMGB1 (11,20).

También se han descrito proteínas moonlighting en virus, especialmente en virus con un genoma corto, estando relacionadas principalmente con la evasión del sistema inmune del huésped. Algunos ejemplos de estos casos son las proteínas E5 y E6 del virus del papiloma humano (1).

Un aspecto de gran interés es que, sorprendentemente, se han descrito proteínas de diferentes especies (incluso genéticamente lejanas) que pueden llevar a cabo la misma función moonlighting. Esto sucede con proteínas involucradas en la adhesión entre patógeno y huésped como, por ejemplo: proteínas del cristalino, DnaK, factor de elongación Tu, enolasa, GAPDH, peroxirredoxina, fosfoglicerato quinasa, fosfoglucomutasa, etc. (21).

Por otro lado, otro grupo importante de proteínas moonlighting son aquellas que, siendo proteínas intracelulares, se encuentran secretadas sin poseer motifs para ello. Este tipo de proteínas multifuncionales llevan a cabo múltiples funciones que influyen en diversas vías metabólicas y procesos celulares, como por ejemplo los procesos de inflamación del organismo (22). Algunas de estas proteínas son la timosina β -4 (T β 4), el factor de crecimiento de fibroblastos básico (bFGF) o la metaloproteínasa de la matriz 12 (MMP-12) (6).

E) Aspectos evolutivos de las proteínas moonlighting

Las proteínas moonlighting nos demuestran que la naturaleza reutiliza secuencias y estructuras de proteínas, para así no tener que diseñarlas de nuevo. A pesar de que el proceso de adquisición de la multifuncionalidad es todavía bastante desconocido, se ha observado que no se requieren largos períodos de tiempo para obtener una segunda función ya que, en muchos casos, la adquisición de una función solo se debe a un cambio en el patrón de expresión, sin ser necesario cambio en la secuencia (1). También se han descrito muchos ejemplos que muestran que a veces se requieren muy pocos cambios en una proteína para que esta consiga sus funciones moonlighting, como sucede con la chaperona GroEL de *Enterobacter aerogenes*, donde el cambio en 4 aminoácidos supone una segunda actividad como toxina (23).

La historia evolutiva de cada proteína multifuncional es compleja y difícil de determinar, ya que cada caso concreto suscita muchos interrogantes aún por determinar. La probabilidad de que una proteína adquiriera una función moonlighting depende de muchos factores. En términos evolutivos, una función moonlighting surge como resultado de una interacción beneficiosa entre una proteína y un nuevo partner de unión que supone un aumento de la eficacia biológica (*fitness*) del organismo. Si la nueva interacción es beneficiosa, la selección natural favorecerá las mutaciones o modificaciones postraduccionales, bien en la proteína o en el partner, que mejoren la afinidad de dicha interacción. Esta interacción puede suponer dos resultados. Por un lado, el gen que codifica esta proteína que ha adquirido una nueva interacción/función puede duplicarse y codificar dos proteínas con una función cada una. Por otro lado, pueden darse cambios genéticos que minimicen el conflicto adaptativo entre las dos funciones o que mejoren el control en el tiempo y lugar en los que cada función se va a llevar a cabo, originándose una proteína moonlighting (21,24). Estos cambios

genéticos que favorecen una función moonlighting pueden suceder en el gen que codifica la proteína multifuncional, en un gen que controle la estructura de su nuevo partner de unión o en un gen que codifique un factor de transcripción que controle la expresión de dicho partner (21,24).

Un interrogante de gran importancia es conocer si la multifuncionalidad de una proteína se conserva entre diferentes especies. Por lo general se considera que el hecho de que una proteína tenga funciones moonlighting en un organismo no implica que también las tenga en especies cercanas, aunque esta suposición no se encuentra confirmada (1).

Por otro lado, se sabe que una proteína puede ser multifuncional en varias especies diferentes, pero teniendo diferentes funciones moonlighting en unas y otras. Esto sucede, por ejemplo, en el caso de la aconitasa, que tiene una misma función canónica entre especies pero sus funciones moonlighting varían (21).

F) Proteínas moonlighting y enfermedades humanas

Como ya se ha mencionado con anterioridad, las proteínas moonlighting pueden estar relacionadas con patologías humanas. Por lo general, esta relación se debe a alguna mutación (25). Un ejemplo de esto es la GAPDH, que participa en la neurodegeneración y en el Alzheimer (26). Otras proteínas se encuentran involucradas en el cáncer, como la transglutaminasa 2, la proteína p53, la glucosa-6-fosfato isomerasa o la chaperona Hsp90 (3,11,27).

Hay muchos ejemplos de proteínas multifuncionales involucradas en enfermedades humanas cuando la función moonlighting no se considera normal para esa proteína, sino que sucede por una ganancia de función tóxica. En estos casos se suele emplear el término función neomórfica (*neomorphic function*), acuñado por Jeffery (1,15). Una proteína moonlighting con función neomórfica no presenta las mismas funciones en condiciones de salud o de enfermedad, mostrándose la función moonlighting tan solo durante el proceso patológico (25,28).

Análisis realizados con la información contenida en la base de datos MultitaskProtDB-II y otras bases de datos como OMIM (28), Therapeutic Target Database (29) y DrugBank (30) indican que el 78% de las proteínas moonlighting humanas conocidas están actualmente involucradas en patologías conocidas y que el 48% de estas proteínas se corresponden con dianas de fármacos existentes (7).

Todo esto hace que el análisis de las dianas farmacológicas y la toxicidad de los fármacos sea más complejo. Aun así, en algunos casos puede suponer una ventaja farmacocinética, ya que un fármaco que solo afecte a una función moonlighting concreta puede tener menos efectos secundarios (si la función moonlighting es extracelular no tendría que penetrar en la células, o si es una función de interacción con un partner concreto sería más específico) (11).

G) Proteínas moonlighting y virulencia

Muchas proteínas moonlighting de diversos microorganismos tienen una función de virulencia, sobre todo aquellas proteínas del metabolismo primario. El 25% de las proteínas moonlighting descritas hasta ahora se encuentra implicado en los mecanismos de virulencia de microorganismos patógenos (31). Algunas de estas enzimas, como la fosfoglicerato quinasa o la GAPDH, son enzimas citosólicas con un segundo papel en la virulencia de patógenos al actuar como receptores de elementos del huésped como el colágeno, la fibronectina o el plasminógeno. En general, estas proteínas moonlighting no solo pueden ser importantes en procesos infecciosos o en la respuesta inmune, sino que también algunas podrían ser posibles dianas de vacunas (1,8).

En algunas ocasiones sucede el llamado moonlighting forzoso (*forced moonlighting*), que supone el secuestro de proteínas del huésped para forzarlas a que realicen una segunda función que facilita la actividad del patógeno. Habitualmente, los virus usan proteínas del huésped para llevar a cabo aquellas funciones del proceso infeccioso que no son capaces de realizar por sí solos. Otro ejemplo es el caso de *E. coli*, que recluta la actina del huésped para facilitar su unión al epitelio intestinal de este.

Las proteínas del metabolismo primario, probablemente por su ancestralidad, suelen presentar funciones moonlighting. Es más, estas proteínas multifuncionales del metabolismo primario tienen la misma función moonlighting asociada a la función de virulencia. A pesar de no compartir secuencias, sí que compartirían algún motif o alguna característica conformacional que en la mayoría de los casos no está identificada aún (1).

Aun no se ha llegado a una conclusión clara sobre este tema. Algunos estudios han identificado algunos motifs en *Streptococcus* implicados en la unión al plasminógeno humano. Sin embargo, estos motifs no son exclusivos de microorganismos patógenos, e incluso están presentes en microorganismos no patógenos sin actividad probada de unión al plasminógeno.

Una de las proteínas del huésped de mayor relevancia en este tema es el plasminógeno. Muchas de las proteínas moonlighting asociadas a la virulencia de patógenos interactúan con el plasminógeno humano (32,33). Algunas proteínas moonlighting que se unen al plasminógeno son la enolasa, la fosfoglicerato mutasa o la glicerol-3-fosfato deshidrogenasa (10,12). El por qué estas proteínas moonlighting se unen principalmente al plasminógeno es un asunto sin resolver actualmente. No se conoce qué características estructurales o a nivel de secuencia posee el plasminógeno para que esto suceda así. Aun así, sí es un hecho curioso y de gran interés que tantas proteínas multifuncionales de patógenos realicen sus funciones moonlighting asociadas a virulencia mediante la interacción y unión con el plasminógeno del huésped.

- **Sistema del plasminógeno**

El sistema del plasminógeno juega un relevante papel homeostático en la degradación de coágulos de fibrina, matrices extracelulares y barreras de tejido. Es importante para la migración celular, así como para promover la liberación de neurotransmisores (34).

El plasminógeno (Plg) es una glicoproteína de unos 90 kDa y 810 aminoácidos. Está compuesto por siete dominios, incluyendo un dominio N-terminal donde tienen lugar sucesos críticos de procesamiento de la molécula, 5 dominios kringle y un dominio proteasa en el extremo C-terminal (35,36). Los dominios kringle median la unión a diferentes sustratos, siendo los dominios 1 y 4 los responsables de la unión a fibrina y otros receptores al contener regiones de unión a lisina (37). El dominio C-terminal solamente es accesible en la forma activa, la plasmina, y contiene un sitio catalítico compuesto por los aminoácidos HIS603, ASP646 and SER741 (38).

En mamíferos, el plasminógeno circula en dos formas principales: GLU-Plg y LYS-Plg. La forma nativa circulante es GLU-Plg, enzimáticamente inactiva (39). Esta proenzima inactiva se une a proteínas ricas en lisina u otros residuos de carga positiva, como la fibrina. Tras esta unión, ocurre la activación del plasminógeno, a través de la ruptura por parte de proteasas (los activadores del plasminógeno tPA y uPA) de la ruptura de un enlace peptídico entre los aminoácidos ARG560 y VAL561, convirtiéndolo en la plasmina, una serin proteasa activa (35). La activación del plasminógeno se encuentra regulada por el balance existente entre sus activadores y los inhibidores de estos activadores (PAIs) (40,41).

La actividad proteolítica de la plasmina degrada los coágulos de fibrina, por lo que es un elemento clave en la homeostasis del sistema fibrinolítico. Además, esta proteína también degrada otros sustratos como algunos componentes de la membrana extracelular o proteínas de unión (Figura 2). Esta capacidad es crucial para la migración celular y para la eliminación de células apoptóticas (38,42).

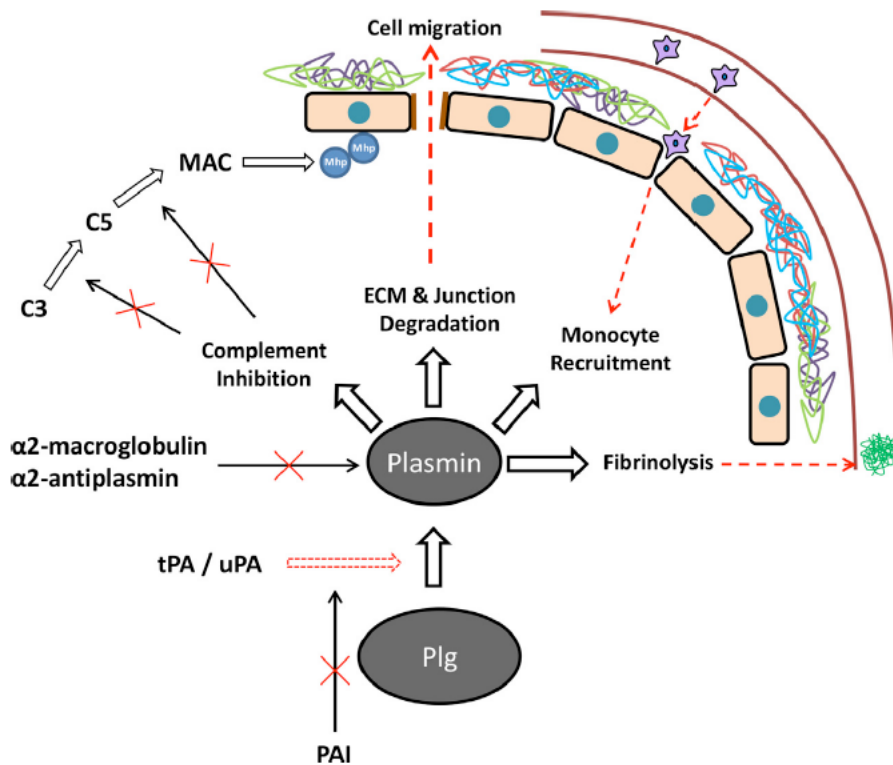


Figura 2. Esquema del sistema del plasminógeno y sus funciones (36).

Se han descrito varios procesos patológicos en los que el plasminógeno está implicado, como inflamación, autoinmunidad o infección microbiana (38). Los microorganismos patógenos utilizan el sistema fibrinolítico y el plasminógeno para facilitar su adhesión a células del huésped y promover el proceso infeccioso (43). Estos microorganismos capturan el plasminógeno, al unirse este a adhesinas y otras proteínas de la superficie celular del patógeno, y promueven su conversión a plasmina. La unión del plasminógeno a estos receptores de la superficie celular del patógeno induce un cambio conformacional en el mismo que permite que los activadores del plasminógeno tengan un mayor acceso a los sitios de activación (40). A su vez, los microorganismos utilizan la actividad proteolítica de la plasmina para: a) atacar el sistema fibrinolítico del huésped y degradar coágulos de fibrina (34,38), b) degradar la matriz extracelular para facilitar la diseminación y para generar fragmentos bioactivos de esta matriz que alteran algunas vías de señalización (35,44), c) activar metaloproteasas de la matriz que ayudan a la destrucción de barreras de tejido y promueven la metástasis microbiana (45) y d) destruir moléculas efectoras del sistema inmune (Figura 3) (42,46).

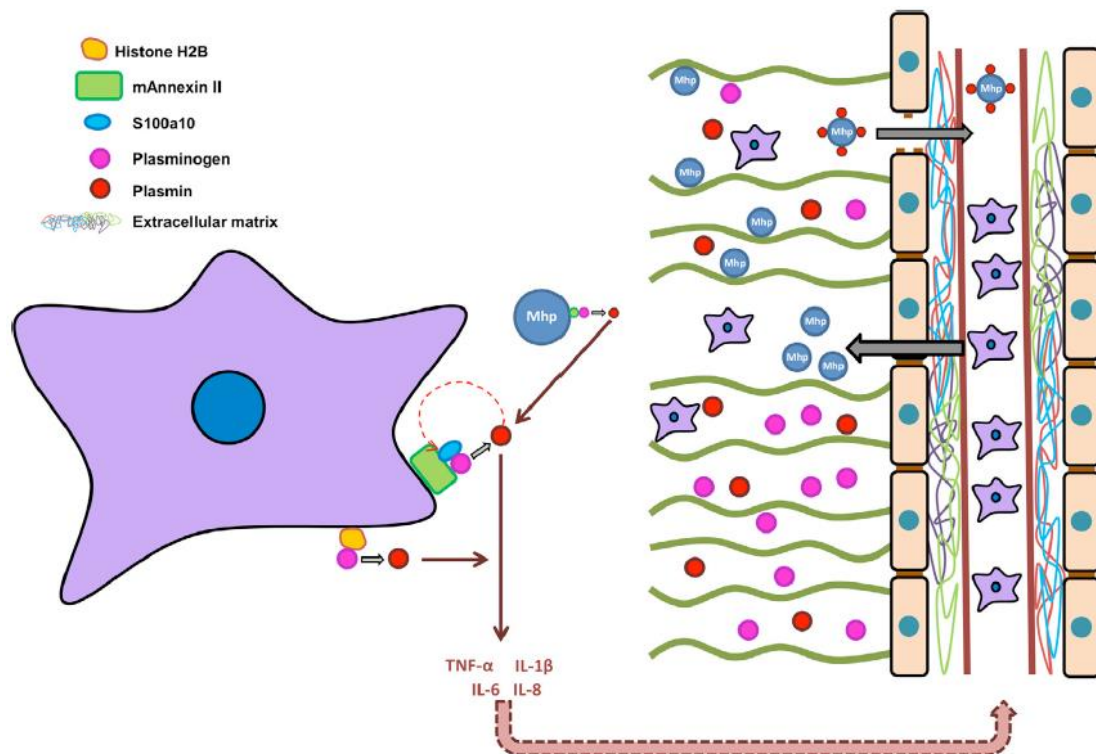


Figura 3. Ejemplo de la interacción entre un microorganismo patógeno y el sistema fibrinolítico (36).

H) Identificación y predicción de proteínas moonlighting

A la hora de identificar la multifuncionalidad de las proteínas, existe una serie de indicadores que pueden indicar que una proteína es moonlighting.

Uno de ellos es el hecho de encontrar una enzima en un compartimento celular que no es el esperado para esa proteína por su función conocida (por ejemplo, como sucede con la lactato deshidrogenasa, la GAPDH o la aldolasa). También puede ser indicativo de multifuncionalidad encontrar proteínas secretada o exportadas a la membrana a pesar de carecer de motivos para dicha secreción (lo que ocurre con la enolasa de algunos organismos patógenos) (5–7).

Otro posible indicio de que una proteína es moonlighting es que se encuentre en mayor cantidad de la que se necesita para realizar su función canónica. Asimismo, también puede señalar multifuncionalidad el encontrar un fenotipo inesperado al deletar un gen (1,11).

Otra opción para facilitar la identificación de proteínas moonlighting es analizar por interactómica los partners de unión de una proteína, ya que el hecho de que esta interactúe con otras proteínas inesperadas puede asociarse a multifuncionalidad. El análisis interactómico también puede servir para descartar que una proteína sea promiscua o *sticky* (proteínas con muchas interacciones pero biológicamente irrelevantes) (9).

Por último, el estudio de proteínas moonlighting con herramientas bioinformáticas es también un método muy importante para la identificación de este tipo de proteínas como, por ejemplo, análisis de dominios o motivos en combinación con programas de predicción de homología. Este punto pone de manifiesto la relevancia del tema principal de este Trabajo Fin de Máster, el análisis bioinformático de la virulencia de las proteínas multifuncionales (9,21).

2.2 MÉTODOS

A) Bases de datos, servidores y herramientas bioinformáticas utilizados

MultitaskProtDB-II (<http://wallace.uab.es/multitaskII/>)

Esta es la primera base de datos de proteínas moonlighting creada. Actualmente cuenta con un total de 694 proteínas moonlighting registradas, de las que se puede obtener diversa información, como: número de acceso de Uniprot, nombre de la proteína, función canónica, funciones moonlighting, términos GO asociados a la función canónica, términos GO relacionados con las funciones moonlighting, organismo en el que se encuentra o su estructura tridimensional, entre otros. Otros datos que se pueden obtener de las proteínas de esta base de datos son si se encuentran asociadas a enfermedades humanas o si actualmente son dianas conocidas de algún medicamento, así como un enlace a la referencia bibliográfica en que se describe a cada proteína como multifuncional.

Todos los análisis realizados en este trabajo se han hecho en proteínas moonlighting seleccionadas de esta base de datos, así como utilizando la información que provee de cada proteína.

UniProt (<https://www.uniprot.org/>)

UniProt es una base de datos gratuita que engloba información sobre la secuencia, estructura y funciones de un gran número de proteínas (a día de hoy, más de 156 millones de proteínas, de las cuales unas 560.000 están revisadas e identificadas). En esta base de datos se aúnan los datos sobre proteínas procedentes de otras bases de datos como Swiss-Prot, TrEMBL y PIR-PSD.

MinimotifMiner (http://cse-mnm.engr.uconn.edu:8080/_MNM/SMS_SearchServlet)

Es una base de datos online de acceso gratuito diseñada para identificar minimotifs en proteínas. Los minimotifs son secuencias peptídicas cortas (de una longitud menor de 15 aminoácidos) que realizan una función en al menos una proteína. MinimotifMiner realiza una búsqueda de motifs cortos en una proteína determinada que suelen estar relacionados con zonas de interacción o a modificaciones postraduccionales.

MEME (<http://meme-suite.org/tools/meme>)

Multiple EM for Motif Elicitation (MEME) es una herramienta de descubrimiento de motifs que forma parte de la colección de herramientas para descubrir y analizar secuencias de motifs MEME Suite. MEME analiza grupos de secuencias de proteínas y obtiene como resultado aquellos motifs más significativos estadísticamente que se encuentran en todas o en algunas de las proteínas del conjunto.

DREME (<http://meme-suite.org/tools/dreme>)

Discriminative Regular Expression Motif Elicitation (DREME) también forma parte del conjunto de herramientas de MEME Suite. Se trata de una herramienta para descubrir *de novo* motifs de corta longitud que se repitan de manera significativa en un amplio grupo de proteínas.

InterPro (<http://www.ebi.ac.uk/interpro/>)

InterPro es un servidor de dominio público que permite realizar un análisis funcional de secuencias proteicas e identificar motifs, sitios funcionales o dominios, así como agruparlas por familias de proteínas. En esta plataforma, las características identificables que se encuentran en proteínas ya conocidas y descritas pueden ser aplicadas a nuevas secuencias proteicas. Está formada por las bases de datos PROSITE, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, GENE3D y PANTHER. La nueva versión de InterPro (74.0, mayo de 2019) incluye 10637 dominios, 913 sitios funciones y 24847 familias.

BioGrid (<https://thebiogrid.org/>)

BioGrid es un repositorio de interacciones biológicas que recopila información sobre interacciones genéticas, interacciones químicas y modificaciones postraduccionales. Actualmente, en esta base de datos de libre acceso se puede encontrar información sobre unas 1670000 interacciones genéticas y entre proteínas en humanos y los principales organismos modelo, procedentes de publicaciones, estudios individuales y conjuntos de datos de alto rendimiento (*high throughput datasets*).

APID (<http://cicblade.dep.usal.es:8080/APID/init.action>)

Agile Protein Interactomes DataServer (APID) es una base de datos curada que alberga una colección de los interactomas proteicos de más de 1100 organismos, incluyendo unas 30 especies con más de 500 interacciones registradas. Toda la información contenida en este servidor deriva de la integración de interacciones proteína-proteína (PPIs) determinadas experimentalmente, y gran parte de estos datos procede de otras bases de datos de interacción entre proteínas, como por ejemplo BioGrid, BIND, DIP, IntAct, HPRD o MINT.

PDB (<https://www.wwpdb.org/>)

Protein Data Bank (PDB) es una base de datos de dominio público que contiene las estructuras tridimensionales de proteínas y ácidos nucleicos. Estos datos se obtienen, por lo general, a través de resonancia magnética nuclear o cristalografía de rayos X, y pueden proceder de grupos de investigación de todo el mundo. Actualmente supera las 150000 estructuras de proteínas, siendo estas de diversos organismos. PDB está gestionada por el *Research Collaboratory for Structural Bioinformatics* (RCSB, <http://www.rcsb.org/>).

Autodock Vina (<http://vina.scripps.edu/>)

Es un programa de libre acceso que sirve para realizar dockings moleculares, es decir, intentar predecir estructuras de complejos intermoleculares (ligando-receptor) buscando y evaluando configuraciones de unión favorables. Este software procede del conjunto de herramientas para realizar dockings denominado AutoDock. De los programas de AutoDock, AutoDock Vina se caracteriza por facilidad de uso (ya que no requiere conocimientos excesivamente avanzados sobre la interacción entre proteínas para su manejo, tan solo es necesario disponer de las estructuras de las proteínas a analizar y especificar un espacio de búsqueda que incluya el sitio de unión entre ellas). Sin embargo, presenta la limitación del tamaño de las moléculas a analizar, ya que este programa es más adecuado para estudiar la unión entre una proteína y un ligando que entre dos proteínas. Por ello, se ha utilizado una herramienta alternativa para poder realizar un mejor análisis de la interacción entre proteínas, ClusPro 2.0.

AutoDock Tools (<http://autodock.scripps.edu/>)

AutoDock Tools (ADT) es un programa desarrollado por los creadores de AutoDock que permite modificar, analizar y preparar las moléculas o proteínas con las que se quieren analizar los dockings estructurales. Por ejemplo, se pueden observar las estructuras tridimensionales de estas proteínas y añadir hidrógenos, asignar cargas atómicas, establecer los enlaces rotables de los ligandos, etc. También sirve para configurar las moléculas y los dockings.

ClusPro 2.0 (<https://cluspro.bu.edu/>)

Este servidor es una herramienta web para realizar dockings estructurales entre proteínas, partiendo de la estructura tridimensional de ambas en formato PDB.

PyMol (<https://pymol.org/2/>)

PyMol es un visor molecular de código abierto que permite visualizar y producir imágenes 3D de alta calidad tanto de moléculas pequeñas como macromoléculas. También sirve para manipular diversas estructuras moleculares de manera independiente, visualizar trayectorias moleculares y las superficies cristalográficas u orbitales.

B) Selección de un listado de proteínas moonlighting asociadas a la virulencia

El primer paso del trabajo ha consistido en la elaboración de un listado de proteínas moonlighting relacionadas con la virulencia de microorganismos patógenos con las que realizar todos los análisis posteriores. Para ello, se ha utilizado la base de datos MultitaskProtDB-II, ya que actualmente es la base de datos de proteínas moonlighting más completa y con el mayor número de

proteínas moonlighting registradas. El criterio a seguir para la inclusión de estas proteínas moonlighting en el listado definitivo ha sido que presenten una función relacionada con virulencia, bien sea su función canónica o sus funciones moonlighting.

Partiendo de este requisito, se ha realizado una búsqueda entre las 694 proteínas que posee la base de datos, buscando en 4 de los registros de información que esta web presenta de cada proteína, concretamente: función canónica, términos GO asociados a la función canónica, funciones moonlighting y términos GO de las funciones moonlighting. Algunas de las funciones que se han buscado son, por ejemplo, invasión a células del huésped (*host cell invasion*), colonización de organismos (*colonization of the organism*), unión al huésped (*host binding, host adherence, adhesion*), o bien un papel general en la virulencia y en el proceso infeccioso (*a role in virulence*). De igual manera, se han seleccionado aquellas proteínas moonlighting implicadas en la unión a proteínas del huésped (*binding to laminin, collagen binding, mucine binding, etc.*), prestando especial atención a la unión a proteínas del sistema fibrinolítico (*plasminogen binding, plasmin binding, fibronectin binding, etc.*).

Una vez finalizada la búsqueda y selección de las proteínas moonlighting de interés para este trabajo, se ha elaborado un listado de todas las proteínas escogidas junto con algunos datos sobre las mismas. La información de cada proteína también ha sido tomada de MultitaskProtDB-II. Así, para cada proteína se ha incluido en el listado: código identificativo de UniProt, nombre de la proteína moonlighting, función canónica, término GO de la función canónica, funciones moonlighting, términos GO de las funciones moonlighting, organismo en el que se encuentra, enlace o similitud con estructuras de proteínas en PDB y referencia bibliográfica correspondiente.

C) Búsqueda y análisis de motifs en las proteínas moonlighting seleccionadas

Parte de los objetivos de este trabajo ha consistido en identificar si existen patrones comunes entre las proteínas moonlighting asociadas a virulencia y si algunos de estos patrones se encuentran exclusivamente en proteínas de organismos patógenos. Para alcanzar estos objetivos, se ha realizado una búsqueda de dominios y motifs en todas las proteínas moonlighting del listado que están relacionadas con la virulencia de microorganismos patógenos.

En primer lugar, se ha escogido la base de datos MinimotifMiner para realizar una búsqueda de los motifs cortos existentes y conocidos en cada una de las proteínas moonlighting. La idea principal de este análisis era obtener motifs de función conocida en todas las proteínas y analizar si algunas de ellas compartían estos motifs. Para este análisis, se han descargado de la base de datos UniProt las secuencias de todas las proteínas moonlighting.

Sin embargo, la página web de acceso a esta base de datos se ha encontrado inhabilitada y sin funcionamiento durante prácticamente la totalidad del periodo de realización del trabajo. Este inconveniente inesperado ha hecho

imposible tanto realizar el rastreo de minimotifs como obtener resultados con esta herramienta. Para solventar esta situación, se han escogido otras herramientas alternativas de búsqueda de motifs y minimotifs ya existentes y, además, *de novo*.

Dos de las herramientas escogidas han sido MEME y DREME, procedentes del conjunto de herramientas bioinformáticas de MEME-Suite. La principal diferencia entre la búsqueda de motifs con esta herramientas es que MEME descubre motifs compartidos por un grupo de proteínas, mientras que DREME encuentra nuevos motifs de corta longitud o minimotifs. Se han utilizado estas dos herramientas para poder estudiar estos dos tipos de motifs.

Para el análisis con ambas herramientas se han descargado las secuencias proteicas en formato FASTA desde Uniprot. Tras esto, se han elaborado diversos grupos de proteínas para buscar motifs compartidos por diferentes tipos de proteínas. Los grupos en los que se han dividido las proteínas moonlighting del listado son los siguientes:

- 1) **Todas las proteínas moonlighting del listado.** Así, se han buscado todos los motifs que se encuentren en general en todas las proteínas moonlighting, independientemente del organismo en el que se encuentren o del tipo de proteína que sea.
- 2) **Todas las proteínas moonlighting que se encuentran en microorganismos patógenos.** De esta manera, la búsqueda se ha delimitado a encontrar motifs que compartan organismos patógenos (para, posteriormente, analizar si estos también se encuentran en organismos no patógenos o bien si son exclusivos de patógenos).
- 3) **Todas las proteínas moonlighting de un mismo organismo.** Por cada organismo diferente que se ha encontrado en el listado se ha creado un conjunto con todas las proteínas multifuncionales asociadas a virulencia.
- 4) **Todas las proteínas moonlighting del mismo tipo.** Es decir, en este caso se han agrupado las proteínas por ser del mismo tipo (por ejemplo, todas las enolasas, todas las GAPDH, todas las DnaK, etc.), independientemente del organismo en el que se encuentran.

Una vez ejecutados estos dos tipos de búsqueda de motifs, se han identificado manualmente las proteínas que los contienen, ya que ni DREME ni MEME especifican qué proteínas los presentan, sino el número de proteínas del grupo analizado que contiene un motif concreto. Se ha determinado la identidad de estas proteínas buscando cada uno de los motifs en los ficheros FASTA descargados de Uniprot que contienen las secuencias de cada uno de los grupos mencionados anteriormente.

Tras esto, se ha llevado a cabo un análisis de en qué microorganismos se encuentran estas proteínas, para descartar los motifs compartidos por proteínas de patógenos y no patógenos y también para seleccionar proteínas que muestren estos motifs únicamente en microorganismos patógenos.

De manera adicional, se ha llevado a cabo la búsqueda de motifs y dominios con el servidor InterPro. Con este se han examinado todas las proteínas moonlighting del listado de manera individual para obtener los dominios con funciones conocidas que poseen y evaluar si algunas de estas proteínas comparten dominios o motifs.

D) Creación de un programa de búsqueda de motifs en proteínas

Tras obtenerse algunos motifs exclusivos de patógenos y compartidos por un mismo tipo de proteínas según el listado de proteínas moonlighting, se ha diseñado un programa para realizar una rápida búsqueda de dicho motif en todas las proteínas del mismo tipo registradas en la base de datos UniProt. De esta manera, lo que se ha pretendido ha sido realizar una valoración rápida de si ese motif efectivamente se encuentra únicamente o mayormente en patógenos, para poder descartar o seleccionar este motif y las proteínas que lo contienen para continuar con los análisis posteriores.

El programa se ha desarrollado en el lenguaje Python (<https://www.python.org/>), un lenguaje de programación interpretado multiplataforma, utilizando el *Notebook* de IPython (<https://ipython.org/>).

El programa toma como *input* o información de partida dos elementos: un motif y el nombre de un tipo de proteína. Después, se conecta con la base de datos Uniprot y se realiza una búsqueda de todas las proteínas con el mismo nombre que se encuentre revisadas en esta base de datos. Así se obtiene un listado de las proteínas y su secuencia correspondiente en formato FASTA que engloba todas las proteínas resultantes de la búsqueda. Tras esto, se pasa a la identificación del motif establecido en todas las proteínas obtenidas. Cada vez que se encuentra el motif en una de las secuencias, se genera como *output* el ID identificativo de Uniprot de dicha proteína y el microorganismo en el que se encuentra. Finalmente, el resultado de la ejecución del programa es un listado con todos los IDs y microorganismos de las proteínas que poseen el motif de partida (Figura 4).

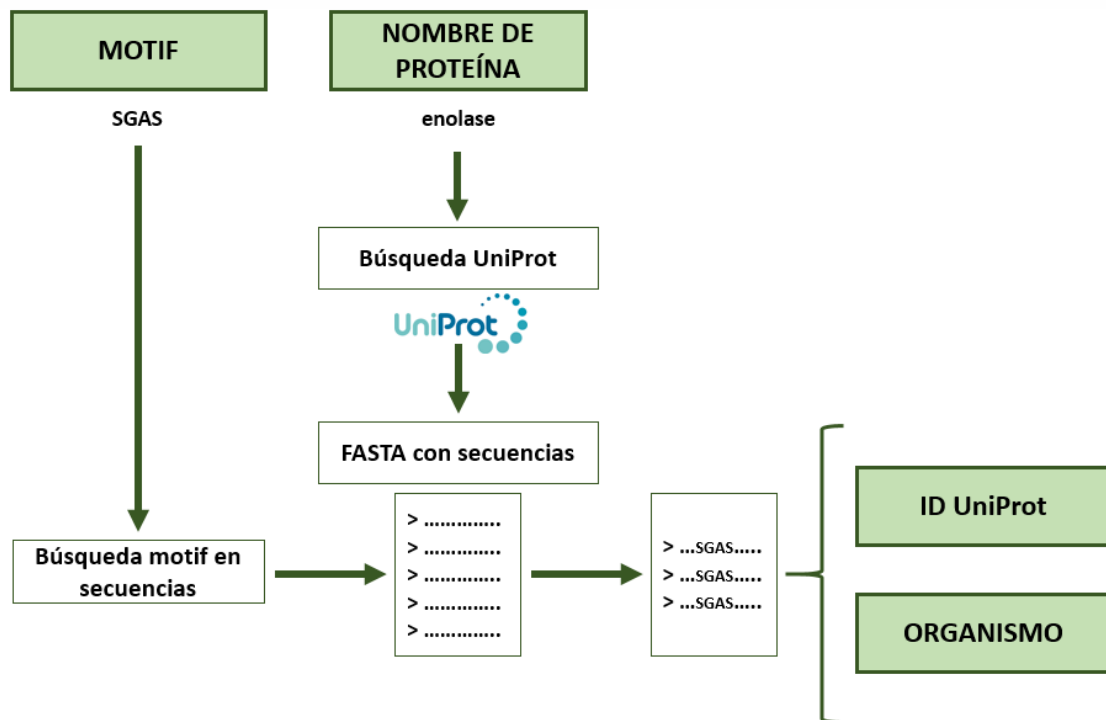


Figura 4. Esquema del flujo de trabajo seguido en el programa de búsqueda de motifs.

E) Análisis de interactómica de proteínas moonlighting con sus partners de unión

Por un lado, partiendo del listado de proteínas moonlighting, se ha efectuado un análisis manual de las proteínas a las que se unen las proteínas moonlighting, con el objetivo de examinar las principales proteínas de unión que presentan y si comparten algunas de ellas, sobre todo proteínas del sistema fibrinolítico. También se ha evaluado si algunas de las proteínas moonlighting que interactúan con elementos como, por ejemplo, el plasminógeno también interactúan con otros del sistema fibrinolítico como la fibronectina.

Por otra parte, se han utilizado las bases de datos BioGrid y APID para los partners de unión de algunas de las proteínas moonlighting. El objetivo de esta búsqueda ha sido evaluar con qué proteínas interactúan estas proteínas moonlighting y, fundamentalmente, si interactúan con el plasminógeno. Concretamente se han analizado aquellas proteínas moonlighting del mismo tipo en las que se encontraron motifs comunes exclusivos de patógenos y con relación con el plasminógeno según la base de datos MultitaskProtDB-II.

F) Dockings estructurales entre proteínas moonlighting y el plasminógeno

Otra tarea del trabajo se ha basado en realizar dockings estructurales para analizar las uniones entre proteínas moonlighting y el plasminógeno con la finalidad de estudiar si estas proteínas moonlighting comparten regiones de unión al plasminógeno.

AutoDock Vina ha sido una de las herramientas escogidas inicialmente para esta tarea. Sin embargo, el tamaño reducido que permite el programa para las moléculas analizadas (ya que es más adecuado para analizar la interacción de una proteína con un ligando que para analizar la interacción entre dos proteínas) y el gran tamaño de las proteínas que se han analizado han limitado la obtención de resultados con este programa.

Como herramienta alternativa para realizar los dockings estructurales se ha empleado el servidor ClusPro 2.0. Para ello se han obtenido las estructuras tridimensionales de las proteínas moonlighting a analizar de la base de datos PDB en el formato con el mismo nombre. Con la herramienta ClusPro 2.0 se han analizado las regiones de unión que se establecen entre el plasminógeno y las mismas proteínas moonlighting evaluadas con los métodos de la sección E, con el objetivo de comparar si estas regiones coinciden con algunos de los motifs de unión a proteínas del huésped identificados en este trabajo.

2.3 RESULTADOS

A) Búsqueda y análisis de proteínas moonlighting implicadas en virulencia

En primer lugar, se ha realizado una búsqueda en la base de datos MultitaskProtDB-II para seleccionar un listado de proteínas moonlighting que puedan estar implicadas en la virulencia y el proceso infeccioso de organismos patógenos.

Del total de proteínas moonlighting registradas en la base de datos MultitaskProtDB-II, que cuenta con 694 proteínas moonlighting, se han escogido 240 para formar parte del listado. Fundamentalmente se han seleccionado proteínas moonlighting con funciones relacionadas con la invasión, colonización o adhesión al huésped, involucradas en el proceso infeccioso en general, con funciones de unión a proteínas del huésped o adhesinas, entre otros. Asimismo, debido a que uno de los principales mecanismos de virulencia de muchos patógenos se basa en la unión al plasminógeno y otras proteínas del sistema fibrinolítico, también se han buscado proteínas moonlighting cuya función canónica o moonlighting se encuentre en relación con la unión al plasminógeno, a la fibronectina y, en general, a otras proteínas del organismo a invadir. En el Material Suplementario 1 (MS1) se puede encontrar el listado completo de estas 240 proteínas moonlighting.

Este listado presenta tipos variados de proteínas, siendo las más abundantes las enolasas (43 proteínas), las GAPDH (28 proteínas), las chaperoninas de 60 kDa (19 proteínas), las chaperonas DnaK (15 proteínas) y los factores de elongación (11 proteínas). Las dos proteínas que se encuentran en mayor número son las dos proteínas moonlighting más conocidas y mejor descritas, de las que se han determinado un gran número de funciones moonlighting. En la Figura 5 se pueden observar los principales tipos de proteínas moonlighting englobadas en el listado.

En cuanto a las funciones canónicas de estas proteínas, más de la mitad de ellas son enzimas y están relacionadas con procesos y rutas metabólicas. Otra parte considerable de estas son chaperonas, mientras que el resto de las proteínas moonlighting del listado se agrupan en proteínas estructurales, receptores y factores de virulencia (Tabla 1).

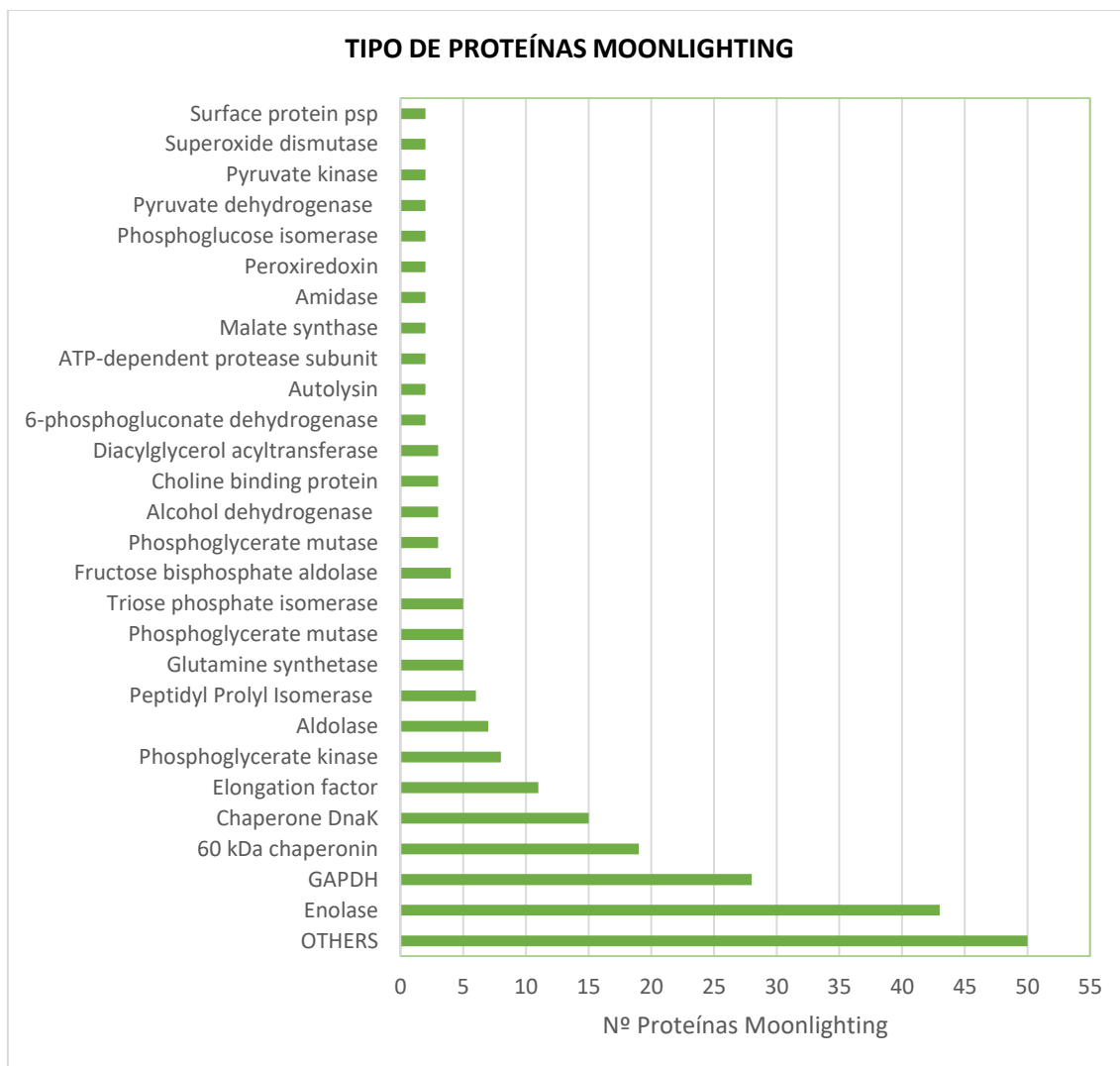


Figura 5. Representación gráfica de los principales tipos de proteínas que componen el listado de proteínas moonlighting.

Tabla 1. Principales funciones canónicas de las proteínas moonlighting del listado y porcentaje de estas proteínas que presentan dicha función.

FUNCIÓN CANÓNICA	% PROTEÍNAS MOONLIGHTING
Enzima	75,42%
Chaperona	15,41%
Proteína estructural	4,16%
Factor de virulencia	2,92%
Receptor	2,09%

Las funciones moonlighting de las proteínas del listado son más variadas. Además, muchas de estas proteínas poseen más de una función moonlighting descrita. Entre toda esta diversidad de funciones, claramente destacan las siguientes funciones: unión a proteínas, adhesinas, factor de virulencia, receptores, proteínas de membrana e inmunomodulación.

Centrándonos en las funciones de estas proteínas moonlighting que están relacionadas con la unión a proteínas del huésped, se han analizado cuáles son las proteínas a las que se unen mayoritariamente las proteínas moonlighting del estudio. Como es posible observar en la Tabla 2, el plasminógeno es la proteína de unión mayoritaria, ya que casi un 40% del total de proteínas moonlighting estudiadas se unen a esta molécula. Otras proteínas con las que interaccionan un número considerable de proteínas moonlighting son: la fibronectina, las células epiteliales, la mucina, el colágeno y la laminina. En el Material Suplementario 2 (MS2) están expuestas todas las proteínas moonlighting que se unen a estas proteínas mencionadas.

Tabla 2. Principales proteínas a las que se unen las proteínas moonlighting del listado y número de proteínas moonlighting que tiene dicha proteína de unión (según la información sobre sus respectivas funciones moonlighting).

PROTEÍNA DE UNIÓN	Nº PROTEÍNAS MOONLIGHTING
Plasminógeno	94
Fibronectina	26
Células epiteliales	15
Mucina	13
Colágeno	9
Laminina	8
Otras	12

Un aspecto curioso que se observó durante el análisis de estas proteínas moonlighting fue el hecho de que algunas de ellas presentan funciones moonlighting de unión a varias de estas proteínas. Es decir, no se unen exclusivamente a una proteína, sino que algunas interaccionan con más de una proteína de unión. También se da el caso de que muchas tienen otras funciones moonlighting adicionales a la unión a proteínas del huésped. En la Figura 6, presentada a continuación, se ha realizado una representación esquemática de los principales pares de proteínas de unión que comparten las proteínas moonlighting del listado.

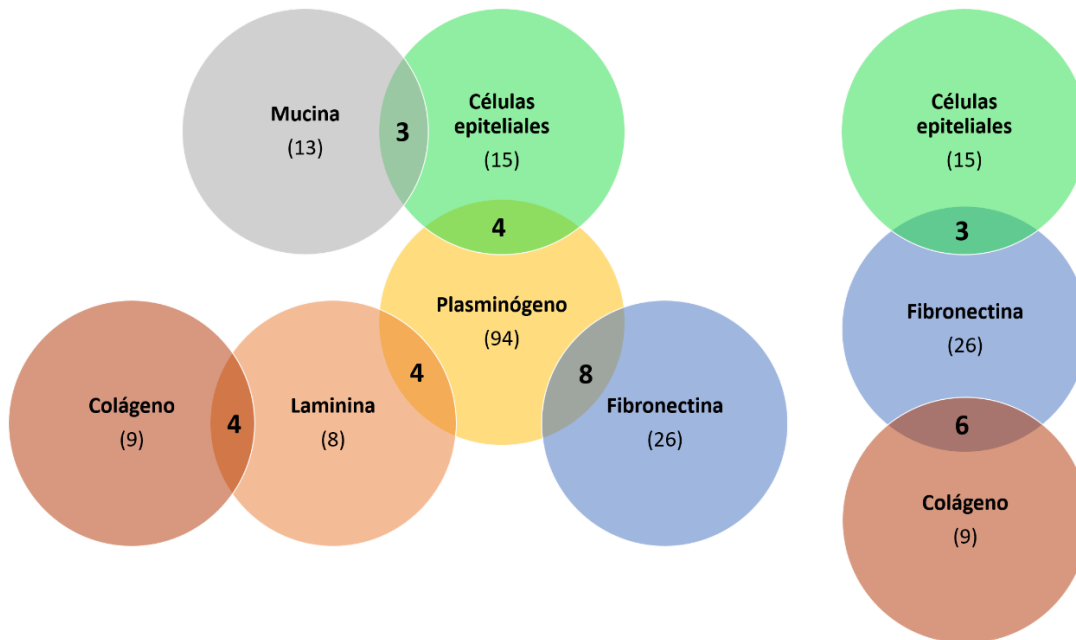


Figura 6. Representación esquemática de las principales proteínas de unión que comparte una proteína moonlighting del listado. El número entre paréntesis indica el número de proteínas moonlighting que se unen a cada tipo de proteína y el número en negrita representa el número de proteínas moonlighting que se unen a ambas proteínas.

En el Material Suplementario 3 (MS3) se ha elaborado un listado con información ampliada sobre estas proteínas moonlighting que comparten las proteínas de unión representadas en la figura anterior. Además, se han incluido otras parejas de proteínas de unión compartidas por un menor número de proteínas moonlighting.

Además de la gran variedad de tipos de proteínas, las proteínas moonlighting que componen el listado se encuentran en una gran diversidad de organismos, tal y como se muestra en la Figura 7. Más de la cuarta parte del listado está formada por distintas proteínas moonlighting de la especie *Streptococcus*. Otros de los microorganismos con mayor representación en este listado son *Mycobacterium*, *Lactobacillus*, *Bifidobacterium*, *Neisseria*, *Candida albicans* y *Staphylococcus*. Se trata de organismos tanto patógenos como no patógenos, ya que no solo se ha tomado el criterio de escoger funciones asociadas a la virulencia, sino unión a proteínas como el plasminógeno. Es interesante tener en consideración proteínas moonlighting de organismos no patógenos para poder analizar las diferencias que puedan presentar en comparación con las proteínas moonlighting de patógenos. De esta manera, se pueden descartar aspectos que presenten en común los dos tipos de proteínas y valorar aquellas características que sean exclusivas de patógenos como factores importantes de la virulencia de estos. Para completar este análisis, se incluye un listado de las proteínas moonlighting ordenadas y clasificadas por organismos en el Material Suplementario 4 (MS4).

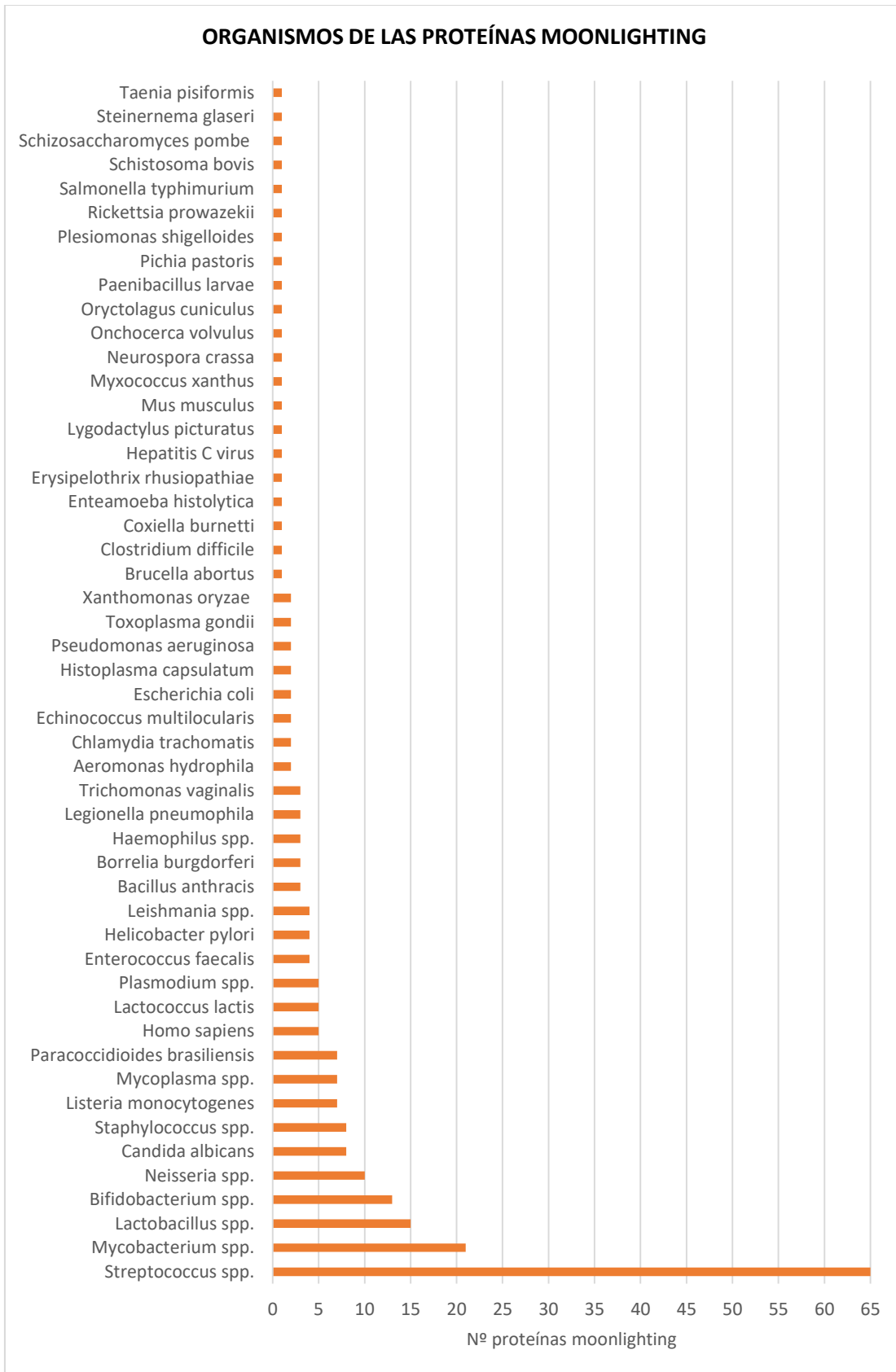


Figura 7. Representación gráfica de los organismos en los que se encuentran las proteínas moonlighting del listado.

B) Estudio de motifs en proteínas moonlighting implicadas en virulencia

Una vez se ha dispuesto de un conjunto de proteínas moonlighting implicadas en virulencia, uno de los objetivos planteados en este proyecto ha sido estudiar si entre las secuencias de algunas de estas proteínas moonlighting existen patrones comunes.

Por tanto, otro de los análisis realizados ha consistido en la búsqueda de motifs en estas proteínas moonlighting seleccionadas.

Primeramente, se escogió la base de datos MinimotifMiner para descubrir minimotifs de función conocida en cada una de las proteínas y comparar los resultados en busca de coincidencias entre algunas de las secuencias proteicas analizadas. Desafortunadamente, tal y como se ha explicado en la sección C de métodos, la página de acceso a esta base de datos se ha encontrado inhabilitada y se ha tenido que buscar herramientas alternativas de análisis de motifs.

Una de las herramientas elegidas fue MEME, para la búsqueda de nuevos motifs en proteínas, a la que se sumó también DREME, destinada a buscar motifs en secuencias proteicas. Con estos dos servidores se procedió a analizar conjuntos de proteínas agrupadas según 4 criterios: 1) las 240 proteínas moonlighting del listado, b) todas las proteínas moonlighting de patógenos, c) todas las proteínas moonlighting de un mismo organismo patógeno y d) todas las proteínas moonlighting de un mismo tipo.

La búsqueda de motifs con MEME no obtuvo ningún motif que fuera significativo estadísticamente (p -valor $< 0,1$), por lo que no se tendrá en cuenta ninguno de los motifs obtenidos con esta herramienta.

Por otro lado, la búsqueda de motifs con DREME sí produjo resultados significativos con e -valores por debajo de 0,05. A continuación se van a comentar los diferentes datos resultantes del análisis de los distintos grupos de proteínas creados:

- Grupo A: se obtuvieron 36 motifs diferentes compartidos por conjuntos de 50-25 proteínas diferentes del total de 240 proteínas moonlighting analizadas.
- Grupo B: se analizaron 124 proteínas moonlighting de organismos patógenos. De entre todas estas, se identificaron 8 motifs comunes, por lo general compartidos cada uno por unas 20-28 proteínas, respectivamente.
- Grupo C: se analizaron todos los diferentes organismos patógenos de las 240 proteínas moonlighting del listado inicial y se crearon grupos de organismo patógenos para los que había más de una proteína moonlighting en el listado. En total la búsqueda de motifs se efectuó sobre 25 grupos de organismos patógenos distintos. Solo se obtuvieron 3 motifs con valores significativos en las proteínas de moonlighting de la especie *Streptococcus*.

- Grupo D: en este caso las proteínas se agruparon según pertenecieran a un mismo tipo de proteína. La búsqueda de motifs se llevó a cabo en 27 grupos de proteínas distintas, aunque solo se encontraron motifs en los siguientes tipos: chaperonina 60kDa (13 motifs), chaperona DnaK (21 motifs), enolasa (53 motifs) y GAPDH (19 motifs).

Para acceder al análisis completo de DREME y a sus resultados explicados en detalle se puede consultar el Material Suplementario 5 (MS5).

Tras poseer estos motifs relevantes se continuó con el análisis, concretamente profundizando en qué proteínas comparten estos motifs para evaluar si algunos de estos motifs están compartidos de manera exclusiva por proteínas de patógenos. Por consiguiente, el siguiente paso del análisis fue seleccionar motifs exclusivos de proteínas de organismos patógenos.

La evaluación en mayor profundidad de la identidad de las proteínas que poseen los motifs de los 4 grupos mencionados anteriormente se hizo buscando los motifs, individualmente, en las 240 proteínas moonlighting y descartó un gran número de ellos. Por ejemplo, los 36 motifs obtenidos del análisis de las 240 proteínas moonlighting se descartaron porque los poseían un alto número de proteínas de organismos no patógenos.

En el caso de los motifs identificados dentro de todas las proteínas de patógenos y todas las proteínas de *Streptococcus* también se descartaron porque, a pesar de ser comunes en estos grupos de patógenos, se encontraron en otras de las proteínas moonlighting de no patógenos del listado inicial. La misma situación se dio con los motifs compartidos por las chaperoninas 60kDa y las chaperonas DnaK.

Definitivamente, se seleccionaron 17 motifs exclusivos de patógenos o mayoritariamente de patógenos (solo compartidos por 1 proteína de organismos no patógenos). Los dos únicos tipos de proteínas con motifs que cumplían los requisitos para ser seleccionados fueron enolasas y GAPDH, 54 proteínas en total. En la Tabla 3 se muestran los 17 motifs seleccionados, el número de proteínas moonlighting de patógenos y no patógenos que los contienen, el código UniProt de estas proteínas y los tipos de microorganismos patógenos a los que pertenecen.

Tabla 3. Motifs que se encuentran exclusiva o mayoritariamente en microorganismos patógenos.

Motif	Proteína	Nº Patógenos	Nº No Patógenos	ID UniProt Proteínas Patógenos	Microorganismos Patógenos
TAVV	Enolasa	16	0	A0A1F1EC06, A4W2T1, A8AY46, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O69174, P0DM31, P69949, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2	<i>Streptococcus</i> spp, <i>Bacillus anthracis</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> ,
TLTET	Enolasa	21	1	A0A1F1EC06, A4W2T1, A8AY46, C1KY94, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O51312, O69174, P0DM31, P69949, P75189, P9WNL1, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Borrelia burgdorferi</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> , <i>Mycoplasma pneumoniae</i> , <i>Mycobacterium tuberculosis</i> , <i>Paenibacillus larvae</i>
IADI	Enolasa	18	0	A0A1F1EC06, A4W2T1, A8AY46, C1KY94, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O69174, P69949, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, T1WGD6, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Staphylococcus aureus</i> , <i>Paenibacillus larvae</i>
IEDG	Enolasa	22	0	A0A1F1EC06, A4W2T1, A8AY46, A8VHK4, C1KY94, C3LDI0, C6GGT7, C6SKU1, F2QDA9, M4ZD78, O51312, O69174, P0DM31, P69949, P75189, Q27727, Q845Q5, Q8GE63, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Mycoplasma</i> spp, <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Neisseria meningitidis</i> , <i>Borrelia burgdorferi</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> , <i>Plasmodium falciparum</i> , <i>Aeromonas hydrophila</i> , <i>Paenibacillus larvae</i>
MMNI	Enolasa	20	0	A0A1F1EC06, A4W2T1, A8AY46, C1KY94, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O69174, P0DM31, P69949, P9WNL1, Q845Q5, Q8DTS9, Q8GE63, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Enterococcus faecalis</i> , <i>Mycobacterium tuberculosis</i> , <i>Aeromonas hydrophila</i> , <i>Paenibacillus larvae</i>
ATNA	Enolasa	17	1	A0A1F1EC06, A4W2T1, A8AY46, C1KY94, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O69174, P0DM31, P69949, Q845Q5, Q8DTS9, Q97QS2, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp , <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Enterococcus faecalis</i> , <i>Paenibacillus larvae</i>
ESGA	Enolasa	17	1	A0A1F1EC06, A4W2T1, A8AY46, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O69174, P0DM31, P69949, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp , <i>Bacillus anthracis</i> , <i>Enterococcus faecalis</i> , <i>Paenibacillus larvae</i>
FHAL	Enolasa	16	1	A4W2T1, A8AY46, C3LDI0, C6GGT7, C6SKU1, F2QDA9, M4ZD78, P0DM31, P69949, P75189, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2	<i>Streptococcus</i> spp, <i>Bacillus anthracis</i> , <i>Neisseria meningitidis</i> , <i>Enterococcus faecalis</i> , <i>Mycoplasma pneumoniae</i>

QLLRI	Enolasa	24	1	A0A1F1EC06, A4W2T1, A8AY46, B8Q027, C1KY94, C3LDI0, C6GGT7, C6SKU1, F2QDA9, M4ZD78, O69174, P0DM31, P69949, P9WNL1, Q27727, Q3HL75, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, T1WGD6, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Steinernema glaseri</i> , <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Neisseria meningitidis</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> , <i>Mycobacterium tuberculosis</i> , <i>Plasmodium falciparum</i> , <i>Leishmania mexicana</i> , <i>Taenia pisiformis</i> , <i>Paenibacillus larvae</i>
RTDR	Enolasa	20	1	A0A1F1EC06, A4W2T1, A8AY46, A8VHK40, C1KY94, C3LDI0, C6GGT7, F2QDA9, M4ZD78, O51312, O69174, P0DM31, P69949, Q845Q5, Q8DTS9, Q97QS2, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Streptococcus</i> spp, <i>Mycoplasma fermentans</i> , <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Borrelia burgdorferi</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> , <i>Leishmania mexicana</i> , <i>Paenibacillus larvae</i>
YPII	Enolasa	22	1	A0A113S8Y8, A0A1F1EC06, A4W2T1, A8AY46, A8VHK4, C1KY94, C6GGT7, F2QDA9, M4ZD78, O51312, O69174, P0DM31, P69949, P75189, Q845Q5, Q8DTS9, Q97QS2, Q9UAE6, Q9XDS7, T1TFA0, U2XJZ2, V9W1Z1	<i>Plasmodium</i> spp, <i>Streptococcus</i> spp, <i>Steinernema glaseri</i> , <i>Listeria monocytogenes</i> , <i>Bacillus anthracis</i> , <i>Neisseria meningitidis</i> , <i>Staphylococcus aureus</i> , <i>Enterococcus faecalis</i> , <i>Mycoplasma pneumoniae</i> , <i>Leishmania mexicana</i> , <i>Toxoplasma gondii</i> , <i>Taenia pisiformis</i> , <i>Paenibacillus larvae</i>
GKLD	GAPDH	17	0	A0A1D2ITN5, C6S993, C7VDQ3, F2QGM8, J7M6M6, P94915, P99136, P9WN83, Q05G10, Q1J8I3, Q3Y454, Q59906, Q8CPY5, Q8DXS8, Q9AJT5, U2Z425, W8S2K4	<i>Listeria monocytogenes</i> , <i>Neisseria meningitidis</i> , <i>Enterococcus faecalis</i> , <i>Streptococcus</i> spp, <i>Mycobacterium</i> spp, <i>Staphylococcus</i> spp, <i>Mycoplasma suis</i> , <i>Erysipelothrix rhusiopathiae</i>
CLAP	GAPDH	19	1	A0A0J1HQ93, A0A1D2ITN5, B2LXU3, C7VDQ3, C8UPL3, F2QGM8, J7M6M6, P47543, P94915, P9WN83, Q1J8I3, Q3Y454, Q59906, Q8DXS8, Q8X1X3, Q92211, Q9AJT5, U2Z425, W8S2K4	<i>Bacillus anthracis</i> , <i>Listeria monocytogenes</i> , <i>Schistosoma bovis</i> , <i>Enterococcus faecalis</i> , <i>Escherichia coli</i> , <i>Streptococcus</i> spp, <i>Mycoplasma genitalium</i> , <i>Mycobacterium</i> spp, <i>Staphylococcus aureus</i> , <i>Paracoccidioides brasiliensis</i> , <i>Candida albicans</i> , <i>Erysipelothrix rhusiopathiae</i>
GAKK	GAPDH	18	1	A0A1D2ITN5, A2DHT2, C6S993, C8UPL3, F2QGM8, J7M6M6, P94915, P99136, P9WN83, Q1J8I3, Q3Y454, Q59906, Q8CPY5, Q8DXS8, Q8X1X3, Q92211, Q9AJT5, U2Z425	<i>Listeria monocytogenes</i> , <i>Trichomonas vaginalis</i> , <i>Neisseria meningitidis</i> , <i>Escherichia coli</i> , <i>Streptococcus</i> spp, <i>Mycobacterium</i> spp, <i>Staphylococcus epidermidis</i> , <i>Paracoccidioides brasiliensis</i> , <i>Candida albicans</i>
RAR	GAPDH	19	1	A0A0J1HQ93, A0A1D2ITN5, A2DHT2, C6S993, C7VDQ3, F2QGM8, J7M6M6, P47543, P94915, P99136, P9WN83, Q05G10, Q1J8I3, Q3Y454, Q59906, Q8CPY5, Q8DXS8, Q9AJT5, U2Z425	<i>Bacillus anthracis</i> , <i>Listeria monocytogenes</i> , <i>Trichomonas vaginalis</i> , <i>Neisseria meningitidis</i> , <i>Enterococcus faecalis</i> , <i>Streptococcus</i> spp, <i>Mycoplasma</i> spp, <i>Mycobacterium</i> spp, <i>Staphylococcus</i> spp,

RVP	GAPDH	24	1	A0A0J1HQ93, A0A1D2ITN5, A2DHT2, B2LXU3, C6S993, C7VDQ3, C8UPL3, F2QGM8, J7M6M6, P47543, P94915, P99136, P9WN83, Q05G10, Q1J8I3, Q3Y454, Q59906, Q8CPY5, Q8DXS8, Q8X1X3, Q92211, Q9AJT5, U2Z425, W8S2K4	<i>Bacillus anthracis</i> , <i>Listeria monocytogenes</i> , <i>Trichomonas vaginalis</i> , <i>Schistosoma bovis</i> , <i>Neisseria meningitidis</i> , <i>Enterococcus faecalis</i> , <i>Escherichia coli</i> , <i>Streptococcus</i> spp, <i>Mycoplasma</i> spp, <i>Mycobacterium</i> spp, <i>Staphylococcus</i> spp, <i>Paracoccidioides brasiliensis</i> , <i>Candida albicans</i> , <i>Erysipelothrix rhusiopathiae</i>
TGAACA	GAPDH	21	1	A0A0J1HQ93, A0A1D2ITN5, B2LXU3, C6S993, C7VDQ3, C8UPL3, F2QGM8, J7M6M6, P47543, P94915, P99136, P9WN83, Q1J8I3, Q3Y454, Q8CPY5, Q8DXS8, Q8X1X3, Q92211, Q9AJT5, U2Z425, W8S2K4	<i>Bacillus anthracis</i> , <i>Listeria monocytogenes</i> , <i>Schistosoma bovis</i> , <i>Neisseria meningitidis</i> , <i>Enterococcus faecalis</i> , <i>Escherichia coli</i> , <i>Streptococcus</i> spp, <i>Mycoplasma genitalium</i> , <i>Mycobacterium</i> spp, <i>Staphylococcus</i> spp, <i>Paracoccidioides brasiliensis</i> , <i>Candida albicans</i> , <i>Erysipelothrix rhusiopathiae</i>

Con el objetivo de obtener más información sobre estos motivos, su localización y su posible función dentro de las proteínas en las que se encuentran se utilizó InterPro. Con este servidor se realizó una búsqueda independiente de cada una de las 54 proteínas expuestas en la Tabla 3 para realizar un análisis funcional de estas secuencias proteicas y analizar dominios o motivos que se encuentren registrados en esta base de datos. Los dominios que muestra esta herramienta son mucho más largos que los motivos identificados previamente, y no se obtuvieron motivos que pudieran estar relacionados con la unión a proteínas del huésped o virulencia. Por lo general este programa muestra dominios relacionados con la función canónica de estas proteínas y, además, el número de motivos conocidos es limitado, por lo que en definitiva este programa no proporciona información adicional sobre la función de los motivos de interés.

Sin embargo, al efectuar una búsqueda de estos motivos en cada una de las secuencias, ha sido posible determinar la localización de estos motivos en cada una de las proteínas y, en concreto, dentro de determinados dominios especificados por InterPro.

De los 6 motivos identificados en las GAPDH, todos se han encontrado en el dominio catalítico de las proteínas, a excepción del motivo 'GAKK', que se localiza en el dominio de unión a NAD(P) en el extremo N-terminal (Figura 8). Respecto a los 11 motivos encontrados en las enolasas, se han localizado todos en el dominio C-terminal de estas proteínas, salvo el motivo 'ESGA', situado en el dominio N-terminal (Figura 9). Las Tablas 4 y 5, presentadas a continuación, resumen la localización de cada motivo en las GAPDH o las enolasas en las que se han encontrado, respectivamente.

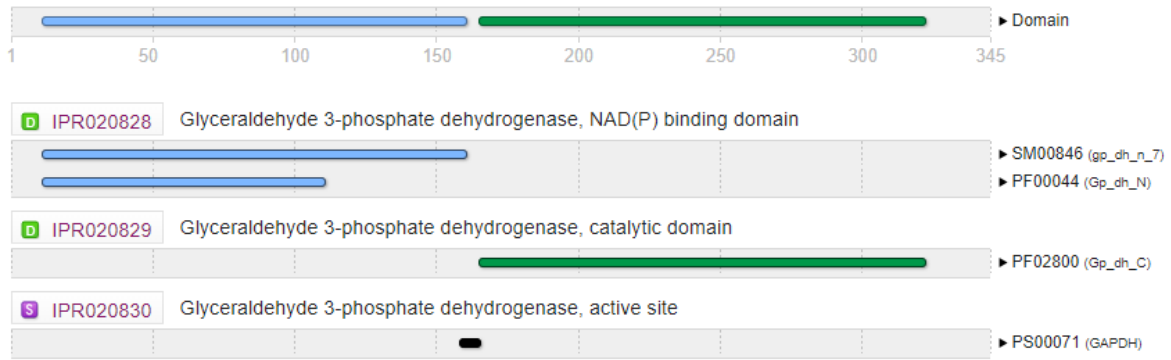


Figura 8. Dominios encontrados con InterPro para las proteínas GAPDH.

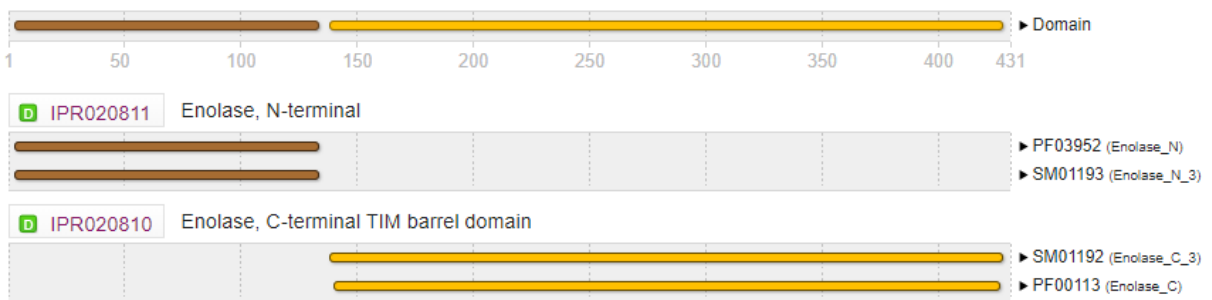


Figura 9. Dominios encontrados con InterPro para las proteínas enolasas.

Tabla 4. Posición de los 6 motivos encontrados en GAPDH dentro de la secuencia de aminoácidos de cada proteína.

	GKLD	CLAP	GAKK	RAR	RVP	TGAAKA
A0A0J1HQ93	NE	155-158	NE	196-198	232-234	209-214
A0A1D2ITN5	226-229	155-158	113-116	198-200	234-236	211-216
A2DHT2	NE	NE	133-136	217-219	254-256	NE
B2LXU3	NE	157-160	NE	NE	235-237	212-217
C6S993	228-231	NE	116-119	200-202	236-238	213-218
C7VDQ3	226-229	155-158	NE	198-200	234-236	211-216
C8UPL3	NE	154-157	113-116	NE	232-234	209-214
F2QGM8	226-229	155-158	113-116	198-200	234-236	211-216
J7M6M6	236-239	165-168	123-126	208-210	244-246	221-226
P47543	NE	161-164	NE	202-204	238-240	215-220
P94915	231-234	162-165	NE	203-205	NE	216-221
P99136	226-229	NE	113-116	198-200	234-236	211-216
P9WN83	231-234	162-165	NE	203-205	NE	216-221
Q05G10	232-235	NE	NE	204-206	NE	NE
Q1J8I3	236-239	165-168	123-126	208-210	244-246	221-226
Q3Y454	227-230	156-159	114-117	199-201	235-237	212-217
Q59906	227-230	156-159	114-117	199-201	235-237	NE
Q8CPY5	226-229	NE	113-116	198-200	234-236	211-216
Q8DXS8	227-230	156-159	114-117	199-201	235-237	212-217

Q8X1X3	NE	156-159	115-118	NE	234-236	211-216
Q92211	NE	155-158	114-117	NE	233-235	210-215
Q9AJT5	218-221	147-150	105-108	190-192	226-228	203-208
U2Z425	226-229	155-158	112-115	198-200	234-236	211-216
W8S2K4	226-229	155-158	NE	NE	234-236	211-216

Los números representan la posición de los aminoácidos que forman los motivos dentro de cada una de las secuencias. NE = No Encontrado. En azul, motivos situados en el dominio de unión a NAD(P). En verde, motivos situados en el dominio catalítico.

Tabla 5. Posición de los 11 motivos encontrados en enolasas dentro de la secuencia de aminoácidos de cada proteína.

	TAVV	TLTET	IADI	IEDG	MMNI	ATNA	ESGA	FHAL	QLLRI	RTDR	YPII
A0A113S8Y8	NE	NE	381-384	NE	NE	NE	NE	NE	NE	NE	NE
A0A1F1EC06	366-369	348-352	NE	289-292	146-149	387-390	28-31	NE	409-413	400-403	285-288
A4W2T1	366-369	349-353	382-385	290-293	146-149	388-391	28-31	186-189	410-414	401-404	286-289
A8AY46	366-369	NE	381-384	289-292	146-149	387-390	28-31	186-189	409-413	400-403	285-288
A8VHK4	NE	NE	NE	311-314	NE	NE	NE	NE	NE	421-424	306-309
B8Q027	NE	NE	NE	NE	NE	NE	NE	NE	412-416	NE	NE
C1KY94	NE	344-348	377-380	285-288	146-149	383-386	NE	NE	405-409	396-399	281-284
C3LDI0	363-366	345-349	378-381	286-289	146-149	384-387	31-34	186-189	406-410	397-400	NE
C6GGT7	367-370	348-352	382-385	290-293	146-149	388-391	28-31	186-189	410-414	401-404	286-289
C6SKU1	NE	NE	NE	283-286	NE	NE	NE	186-189	403-407	NE	NE
F2QDA9	366-369	348-352	381-384	289-292	146-149	387-390	28-31	186-189	409-413	400-403	285-288
M4ZD78	367-370	349-353	382-385	290-293	146-149	388-391	28-31	186-189	410-414	401-404	286-289
O51312	NE	346-350	NE	287-290	NE	NE	NE	NE	NE	398-401	283-286
O69174	366-369	348-352	381-384	289-292	148-151	387-390	28-31	NE	409-413	400-403	285-288
P0DM31	363-366	345-349	NE	286-289	146-149	384-387	28-31	186-189	406-410	397-400	282-285
P69949	367-370	349-353	382-385	290-293	146-149	388-391	28-31	186-189	410-414	401-404	286-289
P75189	NE	368-372	NE	309-312	NE	NE	NE	200-203	NE	NE	304-307

P9WNL1	NE	340-344	NE	NE	145-148	NE	NE	NE	401-405	NE	NE
Q27727	366-369	NE	NE	NE	NE	NE	NE	NE	422-426	NE	NE
Q3HL75	NE	NE	NE	NE	NE	NE	NE	NE	409-413	NE	NE
Q845Q5	365-368	347-350	380-383	213-216	146-149	386-389	28-31	186-189	408-412	399-402	284-287
Q8DTS9	364-367	346-350	382-385	NE	146-149	385-388	28-31	186-189	407-411	398-401	283-286
Q8GE63	NE	NE	NE	289-292	150-153	NE	NE	NE	NE	NE	NE
Q97QS2	NE	348-352	381-384	289-292	146-149	387-390	31-34	186-189	409-413	400-403	285-288
Q9UAE6	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	297-300
Q9XDS7	366-369	348-352	381-384	289-292	146-149	NE	28-31	186-189	409-413	400-403	285-288
T1TFA0	367-370	349-353	382-385	290-293	146-149	388-391	28-31	186-189	410-414	401-404	286-289
T1WGD6	NE	NE	382-385	NE	NE	NE	NE	NE	410-414	NE	NE
U2XJZ2	366-369	348-352	381-384	289-292	146-149	387-390	28-31	186-189	409-413	400-403	285-288
V9W1Z1	NE	342-346	375-378	283-286	146-149	381-384	28-31	NE	403-407	394-397	279-282

Los números representan la posición de los aminoácidos que forman los motifs dentro de cada una de las secuencias. NE = No Encontrado. En marrón, motifs localizados en el extremo N-terminal. En amarillo, motifs localizados en el extremos C-terminal.

De acuerdo con la anotación de residuos que muestra el servidor InterPro (Tabla 6), algunos de los aminoácidos que forman el motif 'IEDG' (encontrado en 22 de las 30 enolasas analizadas) están identificados como regiones de unión a metales y a ligandos. Por otro lado, los motifs 'QLLRI' y 'RTDR' (encontrados en 24 y 20 de las 30 enolasas analizadas, respectivamente) presentan aminoácidos anotados como regiones de unión a polipéptidos (Figura 10). Es posible que estos motifs estén implicados en las funciones de virulencia de las proteínas analizadas, siendo zonas de unión a proteínas del huésped.

Residue annotation

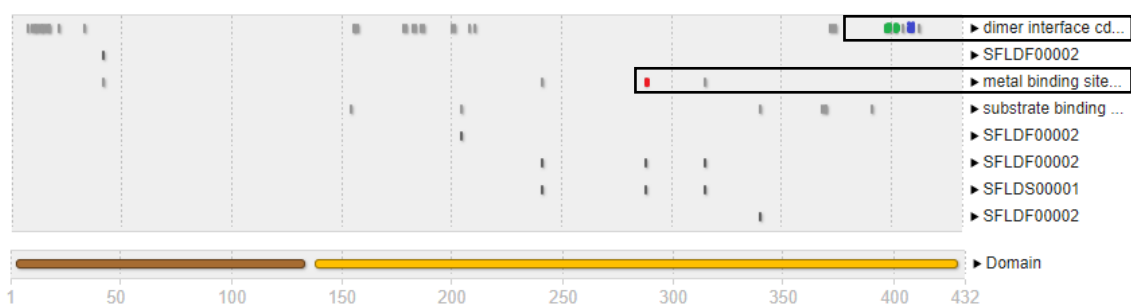


Figura 10. Anotación de residuos para enolasas en InterPro. En rojo, residuos del motif 'IEDG'. En verde, residuos del motif 'QLLRI'. En azul, residuos del motif 'RTDR'.

Tabla 6. Enolasas con residuos de los motifs analizados con anotaciones en el servidor InterPro.

Proteína	Residuo	Anotación	Motif
A0A113S8Y8	384, 385	substrate binding pocket	NE
A0A1F1EC06	291	membrane binding site	IEDG
A4W2T1	187, 188, 401, 403, 412	dimer interface	QLLRI, RTDR
	292	membrane binding site	IEDG
A8AY46	187, 188, 401, 403, 412	dimer interface	QLLRI, RTDR
	291	membrane binding site	IEDG
A8VHK4	421-424	dimer interface	RTDR
	312	membrane binding site	IEDG
B8Q027	415	dimer interface	QLLRI
C1KY94	396-400, 407, 408	dimer interface	QLLRI, RTDR
	287	membrane binding site	IEDG
C3LDI0	34, 187, 188, 397, 398, 401, 408,	dimer interface	QLLRI, RTDR
	288	membrane binding ligand	IEDG
C6GGT7	187, 188, 401, 403, 412	dimer interface	QLLRI, RTDR
	292	membrane binding ligand	IEDG
C6SKU1	187, 188, 405	dimer interface	QLLRI
	285	membrane binding ligand	IEDG
F2QDA9	400-402, 404, 412	dimer interface	QLLRI, RTDR
	291	membrane binding ligand	IEDG
M4ZD78	187, 188	dimer interface	NE
	292	membrane binding ligand	IEDG
O51312	398	dimer interface	RTDR

	289	<i>membrane binding ligand</i>	IEDG
O69174	400-402, 404, 412	<i>dimer interface</i>	QLLRI, RTDR
	291	<i>membrane binding ligand</i>	IEDG
P0DM31	187, 188, 397-401, 408	<i>dimer interface</i>	QLLRI, RTDR
	288	<i>membrane binding ligand</i>	IEDG
P69949	187, 188, 401, 403, 412	<i>dimer interface</i>	QLLRI, RTDR
	292	<i>membrane binding site</i>	IEDG
P75189	202	<i>dimer interface</i>	NE
	310	<i>membrane binding site</i>	IEDG
P9WNL1	403	<i>dimer interface</i>	QLLRI
Q27727	424	<i>dimer interface</i>	QLLRI
Q3HL75	411	<i>dimer interface</i>	QLLRI
Q845Q5	187, 188, 400-402, 410, 411	<i>dimer interface</i>	QLLRI, RTDR
Q8DTS9	187, 188, 398-402, 409, 410	<i>dimer interface</i>	QLLRI, RTDR
Q8GE63	291	<i>membrane binding site</i>	IEDG
Q97QS2	34, 187, 188, 400-402, 404, 412	<i>dimer interface</i>	QLLRI, RTDR
	291	<i>membrane binding site</i>	IEDG
Q9XDS7	187, 188, 400-402, 411, 412	<i>dimer interface</i>	QLLRI, RTDR
	291	<i>membrane binding site</i>	IEDG
T1TFA0	187, 188, 401, 403, 412	<i>dimer interface</i>	QLLRI, RTDR
	292	<i>membrane binding site</i>	IEDG
T1WGD6	412	<i>dimer interface</i>	QLLRI
U2XJZ2	400-402, 404, 412	<i>dimer interface</i>	QLLRI, RTDR
	291	<i>membrane binding ligand</i>	IEDG
V9W1Z1	394, 396, 398, 405	<i>dimer interface</i>	QLLRI, RTDR
	285	<i>membrane binding ligand</i>	IEDG

Los números representan la posición de los aminoácidos en cada una de las secuencias proteicas. NE = No Encontrado.

C) Programa de búsqueda de motifs en secuencias proteicas

Según los resultados obtenidos en la sección anterior, 5 motifs están presentes en proteínas moonlighting del listado que pertenecen exclusivamente a patógenos (4 de estos motifs en enolasas y 1 en GAPDH). Además, 12 motifs se encuentran casi por completo en proteínas de patógenos, con la excepción de que se han hallado en 1 proteína de microorganismos no patógenos (7 motifs en enolasas y 5 motifs en GAPDH).

Este análisis de motifs se limita al conjunto de proteínas moonlighting seleccionadas en el listado. Es posible que estos motifs se encuentren en muchas más proteínas del mismo tipo y que no se hayan integrado en este estudio, bien por no ser proteínas moonlighting, por no ser proteínas moonlighting confirmadas aun o por no tener una función implicada en virulencia conocida o descrita todavía.

Sin embargo, conocer a grandes rasgos en qué otras proteínas no moonlighting se pueden encontrar algunos de estos motifs y, sobre todo, de qué organismos forman parte estas proteínas puede ser interesante. El principal interés de esto radica en poder hacer una valoración de si estos motifs son exclusivos o mayoritariamente de microorganismos patógenos más allá de encontrarse en proteínas moonlighting, así como conocer cuál es la función de los mismos.

Para ello, se ha desarrollado un programa escrito en Python que permite de manera rápida realizar un cribado de todas las proteínas de un mismo tipo registradas en la base de datos Uniprot para seleccionar aquellas que tengan uno de estos motifs analizados. A su vez, el programa muestra información sobre estas proteínas, concretamente sus códigos identificativos de UniProt y el organismo al que pertenecen, respectivamente. El código completo de este programa, junto con comentarios que explican detalladamente su funcionamiento se encuentran en el Material Suplementario 6 (MS6).

Se ha utilizado este programa para buscar todos los motifs mencionados anteriormente, centrándonos en el tipo de proteína en el que se dan según se había visto en los resultados del apartado anterior. De esta manera, las búsquedas realizadas en el programa han sido las siguientes (Tabla 7):

Tabla 7. Estrategias de búsqueda de motifs en el programa desarrollado.

motif =	name =	
'ATNA'	'enolase'	
'ESGA'		
'FHAL'		
'IADI'		
'IEDG'		
'MMNI'		
'QLLRI'		
'RTDR'		
'TAVV'		
'TLET'		
'YPII'		
'CLAP'		'glyceraldehyde+3+phosphate+dehydrogenase'
'GAKK'		
'GKLD'		
'RAR'		
'RVP'		
'TGAAK'		

Para cada una de las búsquedas, el programa da como resultado un amplio listado de proteínas presentes en una amplia variedad de organismos. Todos estos listados están incluidos en el Material Suplementario 7 (MS7).

Al observar los organismos mostrados por el programa se puede confirmar que la amplia mayoría de ellos son organismos patógenos. Esto ha llevado a continuar con la selección de proteínas moonlighting que poseen estos motifs para analizarlas en mayor profundidad.

D) Análisis interactómico de proteínas moonlighting implicadas en virulencia

El siguiente paso que se ha llevado a cabo es un análisis de los partners de unión de las proteínas moonlighting seleccionadas en las dos secciones anteriores. El objetivo de este análisis es observar si estas proteínas interactúan con alguna proteína en común, es decir, si comparten partners de unión.

En primer lugar, se llevó a cabo una búsqueda de cada una de estas 54 proteínas moonlighting en la base de datos BioGrid. Sin embargo, ninguna búsqueda obtuvo resultados, ya que en este servidor no se han encontrado datos registrados sobre las interacciones que tienen ninguna de las proteínas. Por lo general, los análisis de interactómica suelen estar hechos en un número limitado y concreto de especies, principalmente los organismos más conocidos, por lo que para muchos organismos no suele haber estudios de este tipo.

Tras esto se repitió la misma búsqueda, pero en este caso en la base de datos APID. En 3 proteínas se encontraron datos de interacciones registradas en este servidor, dos enolasas (P75189 y Q97QS2) y una GAPDH (Q05G10). Para el resto de proteínas APID no poseía información de interacciones conocidas. Para la enolasa Q97QS7, de *Streptococcus pneumoniae*, se obtuvieron 15 partners de unión, aunque todas ellas son otras proteínas del mismo microorganismo. En el caso de la GAPDH de *Mycoplasma suis*, Q05G10, APID mostró que interactúa con 2 proteínas del cerdo (una de ellas es una actina, otra proteína estructural del huésped). Por último, la búsqueda de la enolasa P75189, de *Mycoplasma pneumoniae*, obtuvo como resultado 7 proteínas con las que interactúa, siendo 6 de ellas proteínas del propio microorganismo y destacando el plasminógeno humano (Figura 11). Las interacciones descubiertas en APID para estas tres proteínas moonlighting se encuentran detalladas en el Material Suplementario 8 (MS8).

7 interactions for ENO_MYCPN (P75189, Mycoplasma pneumoniae M129)

Enolase

Minimum number of **exp. evidences**: 1 Minimum number of **methods**: 1 Minimum number of **publications**: 1 Minimum number of **3D structures**: 1

Showing 1 to 7 of 7 entries

UniprotID	Protein	TaxId	Experimental Evidences	Methods	Publications	3D structures
P75344	DNAK_MYCPN	272634	1	1	1	0
P75183	Y612_MYCPN	272634	1	1	1	0
P75209	YBEY_MYCPN	272634	1	2	1	0
P78010	TPIS_MYCPN	272634	1	2	1	0
P75465	Y316_MYCPN	272634	1	2	1	0
P75247	CLPB_MYCPN	272634	0	1	1	0
P00747	PLMN_HUMAN	9606	1	1	1	0

Figura 11. Captura de pantalla del resultado de la búsqueda de la proteína P75189 en el servidor APID. La flecha en rojo señala el registro de la interacción de esta proteína con el plasminógeno humano.

Tras buscar en las dos bases de datos de interactómica BioGrid y APID, tan solo se ha visto la interacción de esta única enolasa con el plasminógeno. Para ampliar el listado de proteínas moonlighting que tienen al plasminógeno como uno de sus partners de unión se volvió a realizar una búsqueda manual de estas proteínas moonlighting en el listado de proteínas comentado en la sección A de los resultados, para ver cuáles se encuentran asociadas a la unión del plasminógeno. La información resultante muestra que, de acuerdo con la base de datos MultitaskProtDB-II, un total de 35 proteínas moonlighting (en concreto, 19 enolasas y 16 GAPDH) interaccionan con el plasminógeno uniéndose a este. Dos de estas proteínas son las enolasas con interacciones conocidas según APID (aunque para una de ellas APID no tiene registrada la interacción con el plasminógeno). El listado completo de estas proteínas se muestra en la Tabla 8.

Tabla 8. Proteínas moonlighting con motifs exclusivos/mayoritarios de patógenos que interaccionan con el plasminógeno. Se indica el tipo de proteína, el microorganismo en el que se encuentra y el nombre de la base de datos en la que está registrada su interacción con el plasminógeno.

ID Uniprot	Proteína	Microorganismo	Base de datos
A0A1F1EC06	Enolasa	<i>Streptococcus oralis</i>	MultitaskProtDB-II
A8VHK4	Enolasa	<i>Mycoplasma fermentans</i>	MultitaskProtDB-II
C1KY94	Enolasa	<i>Listeria monocytogenes</i>	MultitaskProtDB-II
C3LDI0	Enolasa	<i>Bacillus anthracis</i>	MultitaskProtDB-II
C6SKU1	Enolasa	<i>Neisseria meningitidis</i>	MultitaskProtDB-II
F2QDA9	Enolasa	<i>Streptococcus oralis</i>	MultitaskProtDB-II
M4ZD78	Enolasa	<i>Streptococcus dysgalactiae</i>	MultitaskProtDB-II

O51312	Enolasa	<i>Borrelia burgdorferi</i>	MultitaskProtDB-II
O69174	Enolasa	<i>Staphylococcus aureus</i>	MultitaskProtDB-II
P69949	Enolasa	<i>Streptococcus pyogenes</i>	MultitaskProtDB-II
P75189	Enolasa	<i>Mycoplasma pneumoniae</i>	MultitaskProtDB-II APID
P9WNL1	Enolasa	<i>Mycobacterium tuberculosis</i>	MultitaskProtDB-II
Q3HL75	Enolasa	<i>Leishmania mexicana</i>	MultitaskProtDB-II
Q8DTS9	Enolasa	<i>Streptococcus mutans</i>	MultitaskProtDB-II
Q8GE63	Enolasa	<i>Aeromonas hydrophila</i>	MultitaskProtDB-II
Q97QS2	Enolasa	<i>Streptococcus pneumoniae</i>	MultitaskProtDB-II
Q9XDS7	Enolasa	<i>Streptococcus anginosus</i>	MultitaskProtDB-II
T1WGD6	Enolasa	<i>Taenia pisiformis</i>	MultitaskProtDB-II
U2XJZ2	Enolasa	<i>Streptococcus anginosus</i>	MultitaskProtDB-II
A0A0J1HQ93	GAPDH	<i>Bacillus anthracis</i>	MultitaskProtDB-II
A0A1D2ITN5	GAPDH	<i>Listeria monocytogenes</i>	MultitaskProtDB-II
A2DHT2	GAPDH	<i>Trichomonas vaginalis</i>	MultitaskProtDB-II
B2LXU3	GAPDH	<i>Schistosoma bovis</i>	MultitaskProtDB-II
C8UPL3	GAPDH	<i>Escherichia coli</i>	MultitaskProtDB-II
F2QGM8	GAPDH	<i>Streptococcus oralis</i>	MultitaskProtDB-II
J7M6M6	GAPDH	<i>Streptococcus pyogenes</i>	MultitaskProtDB-II
P99136	GAPDH	<i>Staphylococcus aureus</i>	MultitaskProtDB-II
Q1J8I3	GAPDH	<i>Streptococcus pyogenes</i>	MultitaskProtDB-II
Q3Y454	GAPDH	<i>Streptococcus suis</i>	MultitaskProtDB-II
Q59906	GAPDH	<i>Streptococcus equisimilis</i>	MultitaskProtDB-II
Q8CPY5	GAPDH	<i>Staphylococcus epidermidis</i>	MultitaskProtDB-II
Q8DXS8	GAPDH	<i>Streptococcus pneumoniae</i>	MultitaskProtDB-II
Q9AJT5	GAPDH	<i>Streptococcus pneumoniae</i>	MultitaskProtDB-II
U2Z425	GAPDH	<i>Streptococcus anginosus</i>	MultitaskProtDB-II
W8S2K4	GAPDH	<i>Erysipelothrix rhusiopathiae</i>	MultitaskProtDB-II

En definitiva, 35 proteínas moonlighting de microorganismos patógenos y relacionadas con la virulencia poseen algunos motifs en común y, además, interaccionan con el plasminógeno. Una cuestión pendiente de analizar es si la unión de estas proteínas moonlighting al plasminógeno se realiza en distintas regiones de esta molécula o bien si algunas de estas proteínas comparten zonas de unión.

E) Unión de proteínas moonlighting al plasminógeno

Para estudiar las zonas de unión de estas proteínas moonlighting al plasminógeno se realizaron dockings estructurales. En un principio se intentó obtener resultados utilizando la herramienta AutoDock Vina, aunque el gran tamaño de las proteínas analizadas impidió que se pudieran analizar las interacciones entre estas moléculas.

Por ello, se escogió como alternativa el servidor ClusPro 2.0, el cual modela posibles dockings estructurales para un receptor y un ligando dados y permite que estos dos elementos sean proteínas (Figura 12). En este caso el receptor es el plasminógeno y los ligandos son las proteínas moonlighting analizadas. De las 35 proteínas seleccionadas en el apartado anterior por presentar unión al plasminógeno muchas comparten cierto porcentaje de similitud entre ellas. Debido a esto, se han seleccionado únicamente 14 proteínas (8 enolasas y 6 GAPDH), las que presentan una estructura tridimensional disponible en la base de datos PDB y resuelta experimentalmente. Así, se han analizado de manera individual los dockings estructurales del plasminógeno con cada una de las 14 proteínas moonlighting.

ClusPro 2.0 genera varios modelos posibles para cada pareja de receptor y ligando según la formación de sitios de unión o interacción con una baja energía. De cada pareja plasminógeno-proteína moonlighting se ha escogido el modelo con la energía más baja para ampliar el análisis y observar los aminoácidos involucrados en el docking.

Para esta tarea se ha utilizado el programa PyMol, en el que se puede visualizar las estructuras tridimensionales de las proteínas moonlighting y del plasminógeno, así como los modelos obtenidos con ClusPro 2.0 (Figura 12).

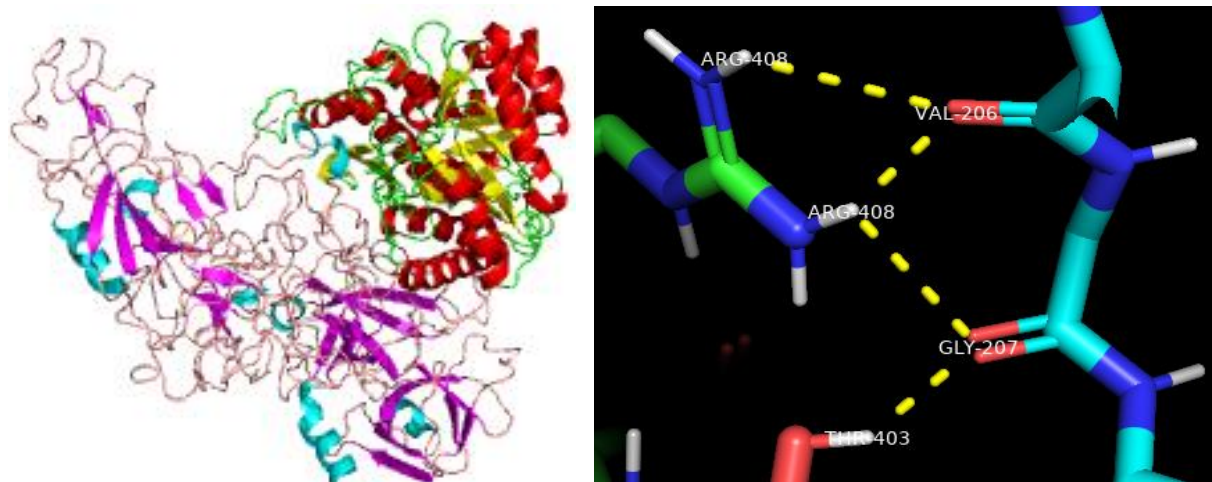


Figura 12. A la izquierda, ejemplo del modelado del docking estructural del plasminógeno y la proteína P75189 obtenido con ClusPro 2.0. A la derecha, análisis con PyMol de una interacción del mismo modelado indicándose los aminoácidos involucrados tanto del plasminógeno (molécula verde) como la proteína P75189 (en azul).

Utilizando las herramientas que proporciona PyMol es posible visualizar y señalar los contactos polares entre las zonas de interacción modeladas por ClusPro 2.0 para cada pareja de moléculas. A su vez, este programa también ofrece la posibilidad de determinar qué residuos forman esta interacción polar. El listado completo de aminoácidos del plasminógeno y aminoácidos de cada una de las proteínas moonlighting involucrados en esta interacción y que forman parte de la zona de unión modelada por ClusPro 2.0 también se puede consultar en el Material Suplementario 9 (MS9).

El análisis de aminoácidos del plasminógeno a los que se unen estas proteínas moonlighting desvela que algunas proteínas moonlighting (tanto enolasas como GAPDH) se unen a un mismo residuo del plasminógeno. Es decir, distintas proteínas comparten zonas de unión en el plasminógeno. Un residuo de interés el ARG561, que en el plasminógeno se encuentra unido a VAL562. Cuando los activadores del plasminógeno rompen esa unión, el plasminógeno se activa a plasmina. La Tabla 9 muestra los residuos del plasminógeno con el mayor número de proteínas moonlighting que interactúan con los mismos.

Tabla 9. Listado de proteínas moonlighting enolasas y GAPDH que se unen a determinados residuos del plasminógeno.

Residuos	Enolasa	GAPDH
ARG-561	P75189, Q3HL75, Q8GE63, Q97QS2, T1WGD6	P99136, Q3Y454, Q8DXS8, W8S2K4
THR-403	P69949, P75189, Q8GE63, Q97QS2	Q3Y454, Q9AJT5
ARG-367	P69949, P75189, Q3HL75, Q8GE63, T1WGD6	Q3Y454
LYS-556	P75189, Q3HL75, Q97QS2, T1WGD6	Q3Y454, Q8DXS8

ASP-735	Q3HL75	A0A0J1HQ93, P99136, Q9AJT5, W8S2K4
TYR-406	P75189, P69949, Q97QS2	Q3Y454
THR-346	P75189, Q3HL75, Q8GE63	Q3Y454
ARG-408	P69949, P75189, Q8GE63	Q3Y454
GLN-622	Q8GE63, Q97QS2, T1WGD6	A0A0J1HQ93

Si comparamos la posición de estos residuos con la localización de los motifs analizados en la sección B de resultados (Tablas 4 y 5) en las secuencias proteicas de estas proteínas moonlighting, observamos que en solo una de las enolasas, la proteína P69949, existe un aminoácido que se une al plasminógeno (ARG401) y que a su vez forma parte de uno de los motifs analizados ('RTDR'). En las demás proteínas no se encuentran estas coincidencias, aunque sí se aprecia que los residuos implicados en la unión al plasminógeno se encuentran muy próximos a los aminoácidos que componen los motifs. Por ejemplo, algunos residuos implicados en la unión al plasminógeno de la proteína Q3HL5 son LEU407 y ASN408, próximos a la localización del motif 'QLLR' (aminoácidos 409 a 413). En la Tabla 10 se muestran los aminoácidos del plasminógeno y los aminoácidos de cada enolasa moonlighting implicados en la interacción entre ellos.

Tabla 10. Aminoácidos implicados según el modelado con ClusPro 2.0 y el análisis con PyMol en la interacción plasminógeno-enolasas.

ENOLASAS	
Plasminógeno	C1KY94
ARG471, ARG504, ARG637, ARG789, ASN790, ASN790, ASN791, ASN791, ASP547, GLU785, LYS468, LYS550, LYS752, SER594, SER669, TRP597, TRP783, TRP783	ARG34, ASN72, ASP124, ASP6, ASP83, GLN127, GLN132, GLN92, GLU28, GLU73, GLU78, GLU96, GLY30, GLY81, LEU126, THR29, THR85
Plasminógeno	O69174
ARG471, ARG504, ARG789, ASP461, GLU460, GLU596, LYS468, LYS473, LYS651, PRO655	ALA98, ASP126, GLU416, GLU78, GLU81, GLU83, LYS134, LYS142, LYS94, PRO77
Plasminógeno	P69949
ARG367, ARG408, ASN399, ASP357, GLN356, GLU460, GLY437, SER365, SER436, THR369, THR403, THR456, TYR254, TYR406	ALA209, ARG401, ASP157, ASP204, ASP69, GLY197, LYS190, LYS194, LYS211, LYS224, LYS56, LYS56, SER154, SER156, THR200, VAL202
Plasminógeno	P75189
ARG367, ARG408, ARG561, ASN339, GLU395, GLY565, LYS556, LYS557, PHE692, SER383, THR346, THR403, TYR406	ALA160, ALA411, ARG21, ARG23, ARG396, ARG408, ASN404, ASP161, GLU373, GLY207, HIS159, LEU18, LYS67, SER397, THR204, VAL206
Plasminógeno	Q3HL75
ARG367, ARG561, ARG561, ASP735, GLN689, GLU395, GLU687, LYS556, SER736, THR346, THR691, VAL563	ARG130, ASN204, ASN408, GLN200, GLU377, GLY125, LEU407, LYS195, PRO398, SER177, SER184

Plasminógeno	Q8GE63
ALA568, ARG367, ARG408, ARG561, ARG677, GLN622, GLY565, LEU402, MET404, PHE692, SER383, THR346, THR403, THR691, VAL563	ALA198, ASN161, ASN214, ASP160, ASP254, ASP54, GLU157, GLU250, GLY200, GLY216, LYS197, LYS55, SER41, SN214, THR42, VAL258
Plasminógeno	Q97QS2
ALA694, ARG561, ASN625, GLN622, GLY693, LYS556, LYS557, PRO555, THR403, TYR254, TYR406	ARG85, ASP124, ASP6, ASP86, GLN333, GLU127, GLU28, LEU126, TYR125, TYR133
Plasminógeno	T1WGD6
ALA694, ARG367, ARG561, ASP357, GLN622, LYS556, LYS557	ASN82, GLU124, GLU415, GLY126, GLY126, LYS125, LYS27, LYS27, LYS419, SER418

En cuanto a las GAPDH, tampoco se encuentran coincidencias entre la localización de los residuos que interaccionan con el plasminógeno y los aminoácidos que forman los motifs encontrados. Únicamente en la proteína W8S2K4 el residuo LYS212 forma parte del motif 'TGAAKA' (que en la secuencia de esta proteína se encuentra en los aminoácidos 211 a 216). En la Tabla 11 se muestran los aminoácidos del plasminógeno y los aminoácidos de cada GAPDH moonlighting implicados en la interacción entre ellos.

Tabla 11. Aminoácidos implicados según el modelado con ClusPro 2.0 y el análisis con PyMol en la interacción plasminógeno-GAPDH.

GAPDH	
Plasminógeno	A0A0J1HQ93
ALA568, ARG561, ASN625, ASN769, ASP735, ASP740, GLN622, GLU623, GLU627, GLU687, GLU714, GLY590, GLY686, GLY732, GLY733, HIS586, LYS607, LYS770, PHE715, SER736, THR346, THR403, THR688, TYR406	ALA215, ALA294, ALA294, ASN123, ASN226, ASN256, ASN256, ASN38, ASN60, ASN62, GLU103, GLU254, GLU275, GLY61, LEU216, LYS101, LYS191, LYS212, LYS222, LYS260, TYR273, TYR273
Plasminógeno	P99136
ALA731, ARG220, ARG561, ARG677, ARG712, ASN769, ASP735, GLN255, GLU172, GLU714, GLY690, GLY732, GLY733, LEU730, PRO252, SER249, SER250, THR253, TYR 713	ARG19, ARG20, ARG326, ARG66, ASN68, ASP62, GLU23, GLU25, GLU333, GLU57, GLU57, GLU71, GLY69, LSY73, LYS46, PHE54, SER267, SER76, THR287, THR55
Plasminógeno	Q3Y454
ARG367, ARG389, ARG408, ARG561, ASN396, ASN399, ASP357, ASP357, GLN391, GLN552, GLU395, GLY437, HIS360, HIS390, LEU402, LYS556, PRO398, THR346, THR403, TYR406	ALA261, ALA262, ALA265, ALA267, ARG20, ARG3, ARG53, ASN167, ASN23, ASP269, ASP55, GLU224, GLU25, GLU329, GLU59, GLU66, GLY56, GLY7, LYS333, LYS46, MET1, SER287, THR57, TYR288
Plasminógeno	Q8DXS8
ARG561, GLN689, GLU623, GLU632, GLY584, HIS569, LYS556, THR688	ARG80, ASP39, GLU89, LEU37, LYS107, LYS127, SER104, TYR193

Plasminógeno	Q9AJT5
ARG767, ASP660, ASP735, GLU627, LYS661, PHE692, THR403	ALA223, ARG212, ARG73, ARG73, LYS66, TYR67, TYR67
Plasminógeno	W8S2K4
ALA694, ALA731, ALA731, ARG220, ARG561, ARG561, ARG677, ARG677, ARG719, ARG767, ASP735, ASP740, GLU687, GLY690, HIS729, HIS729, PHE587, SER736	ALA140, ARG77, ARG77, ARG98, ASN123, ASN62, ASN81, ASP125, ASP78, GLU76, GLU76, GLU80, GLU80, LYS191, LYS212, LYS222, THR100, THR207, THR34, VAL73

2.4 DISCUSIÓN

El número de proteínas moonlighting descritas ha ido aumentando en los últimos años y, además, la identificación de nuevas proteínas multifuncionales es un proceso en continua ampliación. Una parte considerable de todas las proteínas moonlighting conocidas se encuentra relacionada con los mecanismos de virulencia de microorganismos patógenos. Sin embargo, actualmente no se conocen qué características de estas proteínas (a nivel de secuencia o de estructura) son las causantes de que sean factores de virulencia. Analizar si existen regiones comunes en estas proteínas multifuncionales que puedan encargarse de la actividad virulenta del microorganismo ha sido el objetivo principal de este trabajo. Este es un tema de gran importancia, no solo para poder comprender mejor el proceso de infección de algunos microorganismos patógenos, sino también para valorar el desarrollo de nuevas vacunas que vayan dirigidas a estas regiones comunes y responsables de la virulencia.

Al analizar el listado de proteínas moonlighting creado en este trabajo desde la información contenida en la base de datos MultitaskProtDB-II se observa que del número total de proteínas moonlighting descritas a día de hoy casi la mitad de ellas se encuentran implicadas en mecanismos de virulencia (ya que se han escogido 240 de un total de 694 proteínas). Además, según se comentó en la sección de Introducción, se sabe que un 25% de todas las proteínas de esta base de datos son factores de virulencia de organismos patógenos. Asimismo, este listado contiene distintos tipos de proteínas, por lo que los microorganismos no solo utilizan un gran número de proteínas moonlighting como factores de virulencia, sino que además emplean distintos tipos de proteínas (enolasas, GAPDH, chaperonas, etc.).

Este hecho lleva a la pregunta de por qué tantas proteínas moonlighting son utilizadas por organismos patógenos como factores de virulencia. Un posible motivo es que la mayoría son enzimas del metabolismo primario, una función básica y ancestral, por lo que son proteínas altamente conservadas entre organismos procariontes y eucariotes a lo largo de la evolución. Es decir, es probable que muchas de estas proteínas moonlighting de patógenos compartan epitopos con el mismo tipo de proteína en humanos. Esto supone una ventaja para el patógeno a la hora de la invasión al huésped, ya que el sistema inmune humano no las detectaría como extrañas y no generaría anticuerpos frente a ellas, lo que en última instancia supondría una inhibición de la respuesta inmune. En relación con el diseño de vacunas, debido a este suceso no sería aconsejable emplear como elementos vacunales proteínas moonlighting cuyas secuencias tengan alto nivel de homología con proteínas del huésped (31).

Como se vio en el apartado A de Resultados, según su función canónica la gran mayoría de las proteínas moonlighting del listado son enzimas (el mayor número de ellas son enolasas y GAPDH) y chaperonas, proteínas que por lo general se encuentran en el núcleo o en el citoplasma del organismo. Sin embargo, al analizar las funciones moonlighting se observa que las principales son la unión a proteínas, adhesinas, receptores o proteínas de membrana. Es decir, estas proteínas que generalmente cumplen funciones en el interior del organismo también realizan muchas otras en la membrana. El mecanismo por el

que están proteínas, que carecen de péptido señal para su exportación, son transportadas a la membrana celular del organismo todavía se desconoce. No obstante, se han propuesto algunos mecanismos de secreción de proteínas alternativos, como el sistema independiente de péptido señal SecA2 (47) o el sistema T5SS para la exportación de proteínas bacterianas en bacterias Gram - (48).

Dentro de las funciones implicadas en virulencia que realizan estas proteínas desde la superficie del organismo, fundamentalmente destaca la unión e interacción con proteínas del huésped y componentes de la matriz extracelular del mismo. El plasminógeno es la proteína del huésped con la que interaccionan un mayor número de proteínas moonlighting (un 40% de las proteínas del listado). Aun así, también interaccionan con otras proteínas del huésped como la fibronectina, la laminina, el colágeno, la mucina o células epiteliales. Es más, algunas de estas proteínas moonlighting tienen la capacidad de interaccionar con más de una de estas proteínas del huésped. Las dos proteínas del huésped a las que unen doblemente el mayor número de proteínas moonlighting son el plasminógeno y la fibronectina. La fibronectina (al igual que la laminina o el colágeno) es uno de los principales componentes de la superficie celular y la matriz extracelular y puede favorecer la colonización de microorganismos patógenos cuando estos se adhieren a ella. Es decir, los patógenos que presentan proteínas moonlighting capaces de unirse tanto a plasminógeno como a fibronectina presentarían una doble estrategia para favorecer la invasión al huésped: por un lado, la adhesión a fibronectina favorece la colonización y el asentamiento del patógeno, a la vez que la interacción con el plasminógeno desencadena la degradación de la matriz extracelular. En definitiva, la unión a estas distintas proteínas del huésped supone una combinación de acciones para facilitar la diseminación e invasión del patógeno.

Otro aspecto de interés son los diferentes tipos de organismos de los que forman parte las proteínas moonlighting del listado, ya que se trata de organismos patógenos y no patógenos. De hecho, aunque la especie con mayor representación en el listado es el patógeno *Streptococcus*, dos de las especies con mayor número de proteínas moonlighting son *Lactobacillus* y *Bidifobacterium*, dos especies de microorganismos no patógenos que habitan el tracto gastrointestinal humano e inhiben el crecimiento de microorganismos patógenos. Es decir, proteínas de organismos no patógenos también presentan la función de unión a proteínas del huésped como el plasminógeno. En este caso, las proteínas de no patógenos compiten con las proteínas de patógenos por unirse a las mismas proteínas y tejidos del huésped para ejercer una función protectora evitando la invasión y colonización de los patógenos. Esta podría ser la diferencia entre la unión de proteínas moonlighting de patógenos y no patógenos al plasminógeno ya que mientras los patógenos interaccionan con el plasminógeno para activar la plasmina y facilitar la invasión, los no patógenos se unirían al plasminógeno para proteger al huésped sin activarlo.

En resumen, del análisis del listado de proteínas moonlighting se observa que algunos organismos (tanto patógenos como no patógenos) presentan el mismo tipo de proteína (enolasa, GAPDH, chaperonas, etc.) o incluso proteínas diferentes capaces de realizar las mismas funciones moonlighting,

fundamentalmente la unión a proteínas del huésped. Este hecho suscita la pregunta de cómo estas proteínas, que pueden ser estructuralmente diferentes o presentan distintas secuencias, interactúan con las mismas proteínas del huésped.

Para intentar dilucidar esta cuestión se llevó a cabo el análisis de motivos que pudieran ser comunes en varias proteínas moonlighting, para estudiar en qué proteínas se pueden encontrar, en qué región dentro de la secuencia proteica y qué función podrían desempeñar. Actualmente se desconoce si las proteínas moonlighting que se unen a las mismas proteínas diana comparten dominios o motivos comunes. El único caso descrito es el de 3 pequeños motivos relacionados con la unión al plasminógeno en la enolasa, aunque no se comparte con otras proteínas (49).

Como se ha comentado en la sección B de Resultados, con DREME se han encontrado pequeños motivos en distintas proteínas moonlighting del listado, sobre todo en enolasas y GAPDH, las proteínas más abundantes en este listado. La atención se centró en motivos encontrados en proteínas de patógenos de manera mayoritaria, aunque también se tuvieron en cuenta algunos motivos localizados en patógenos y no patógenos. No se ha encontrado información de la función de estos motivos en sí, aunque se ha visto que la localización de estos suele coincidir para el mismo tipo de proteína (el dominio catalítico para los motivos encontrados en GAPDH y el dominio C-terminal para los motivos encontrados en enolasas). Observando la anotación de residuos facilitada por el servidor InterPro, se ve que los aminoácidos que forman parte de algunos de los motivos son regiones de unión a polipéptidos. Es posible que estas regiones sean zonas de unión a proteínas del huésped. De hecho, la búsqueda de las proteínas que poseen estos motivos en bases de datos de interactómica indica la interacción de alguna de ellas con el plasminógeno.

Con el objetivo de ampliar el estudio de la interacción entre el plasminógeno y las proteínas moonlighting con motivos comunes y con registros de unión al plasminógeno según las bases de datos APID y MultitaskProtDB-II, se utilizó el programa ClusPro 2.0 para modelar las posibles regiones de interacción. Sin embargo, al analizar los aminoácidos de las proteínas moonlighting potencialmente involucrados en la interacción con el plasminógeno no se encuentra relación con los aminoácidos que forman los motivos comunes, exceptuando la enolasa P69949 de *Streptococcus pyogenes* y la GAPDH W8S2K4 *Erysipelothrix rhusiopathiae*. Desafortunadamente, esto no permite establecer residuos comunes entre las secuencias de los diferentes tipos de proteínas moonlighting. No obstante, un hecho de importancia es que los residuos de algunas enolasas y GAPDH que pueden formar parte de la interacción con el plasminógeno son aminoácidos cercanos o contiguos a los aminoácidos que forman algunos de los motivos analizados. Existe la posibilidad de que el plasminógeno interactúe con unos aminoácidos pero se una a otros, estando todos involucrados en el papel de adhesión y virulencia de algunas proteínas moonlighting.

Algo que sí ha mostrado el análisis de dockings estructurales ha sido que algunas de las proteínas moonlighting estudiadas interactúan con las mismas

regiones del plasminógeno. En algunos casos tanto enolasas como GAPDH comparten esta característica. Aun así, los resultados no establecen ningún indicio sobre qué hace que estas proteínas moonlighting de distintos organismos se unan al plasminógeno o si existen elementos comunes entre las proteínas moonlighting que realizan esta interacción, por lo que todavía quedan pendientes muchos análisis y estudios sobre este tema.

Es posible que aunque estas proteínas no tengan motifs o dominios comunes sí que puedan tener una función común o una estructura tridimensional conservada y que se dispongan de manera común en la interacción molecular con el plasminógeno. Por tanto, el análisis de las secuencias de proteínas moonlighting debería complementarse con el estudio de la estructura tridimensional de las mismas.

Las herramientas empleadas en el trabajo presentan algunas desventajas o limitaciones. Por ejemplo, los programas de búsqueda de dominios y motifs que están integrados en el servidor InterPro son bases de datos curadas en las que el número de dominios y motifs actualmente descritos es reducido. Además, en el proceso de curado se suelen eliminar aquellas entradas que no coinciden con la función canónica de la proteína, lo cual juega en contra de los objetivos de predicción de proteínas moonlighting. Es más, suele predominar la información sobre los motifs más comunes y las funciones canónicas de las proteínas moonlighting frente a las otras funciones. Una alternativa para continuar con este trabajo en un futuro sería valorar y usar otros servidores o programas no curados.

De igual forma, otra limitación ha sido no disponer de la herramienta MinimotifMiner, que podría haber ampliado los resultados sobre motifs comunes y las funciones de los mismos. Por otro lado, las bases de datos de interactómica proporcionan una información muy limitada sobre los partners de interacción de las proteínas moonlighting, ya que muchas de las especies que poseen estas proteínas no han sido analizadas por interactómica y, por tanto, no se encuentran partners de interacción de la gran mayoría de las proteínas moonlighting estudiadas en este trabajo.

Uno de los aspectos que aún queda por completar y que sería de interés en cuanto a la interacción entre el plasminógeno y las proteínas moonlighting es realizar el análisis de todos los modelos propuestos por ClusPro 2.0 para cada una de las parejas de proteínas e identificar los aminoácidos de ambas partes involucrados en la interacción. Es más, también sería útil emplear otras herramientas alternativas para hacer dockings estructurales y comparar los diferentes resultados obtenidos. Esto puede mostrar otros aminoácidos de interés y que tal vez coincidan con los residuos que forman los motifs encontrados con DREME.

De la misma manera, un análisis interesante pendiente de hacer es analizar las proteínas moonlighting que no solo se unen al plasminógeno, sino que también interaccionan con otras proteínas del huésped como la fibronectina o las células epiteliales y estudiar las regiones de interacción. Así sería posible

evaluar si las proteínas moonlighting presentan residuos concretos destinados a la unión de diferentes proteínas del huésped.

En relación con el objetivo de identificar los residuos proteicos que interaccionan con el plasminógeno u otras proteínas y con el programa en Python para búsqueda de residuos elaborado en este trabajo, también se podría plantear la creación de un programa con mayor complejidad que no solo mostrara en qué proteínas y en qué organismos está un motif determinado, sino que proporcionara la localización de los aminoácidos que forman el motif dentro de cada una de las secuencias proteicas en las que se encuentre.

En definitiva, la identificación de dominios, patrones o estructuras comunes en las proteínas moonlighting implicadas en la virulencia de organismos patógenos tiene aún un largo camino por recorrer. Los análisis de estas proteínas no deben limitarse a encontrar elementos comunes en las secuencias proteicas, sino también analizar con exhaustividad las estructuras tridimensionales y las interacciones que presentan estas proteínas moonlighting, así como a utilizar distintas bases de datos y herramientas que permitan disponer de la máxima información posible.

Las incógnitas que todavía se encuentran por descubrir sobre las proteínas moonlighting y su complejo papel en la virulencia de microorganismos patógenos, junto con otros aspectos aún desconocidos de estas proteínas, hacen que las proteínas moonlighting sean un ámbito de gran interés y con mucho potencial que continuará ampliándose cada vez más con el paso de los años.

3. Conclusiones

1. Casi la mitad de las proteínas moonlighting identificadas actualmente se encuentran relacionadas con la virulencia de microorganismos patógenos. La mayoría de estas proteínas son enolasas y GAPDH, enzimas del metabolismo primario, y se encuentran en un gran número de organismos patógenos y no patógenos. La principal función moonlighting de estas proteínas multifuncionales es la adhesión a proteínas del huésped, fundamentalmente el plasminógeno.
2. Algunas enolasas y GAPDH multifuncionales presentan motifs comunes, respectivamente. Sin embargo, no se dispone de información suficiente sobre la función que pueden ejercer ni se han encontrado relacionados claramente con la unión de estas proteínas al plasminógeno.
3. La localización de estos motifs suele darse en el mismo dominio para el mismo tipo de proteína y algunos de los residuos que componen los motifs están anotados como zonas de unión a polipéptidos.
4. El análisis de la interacción de estas enolasas y GAPDH multifuncionales con el plasminógeno humano mediante dockings estructurales relaciona algunos aminoácidos de los motifs encontrados en las proteínas moonlighting con los aminoácidos que forman parte de la interacción de alguna de estas proteínas con el plasminógeno. Además, en otras de las proteínas multifuncionales los aminoácidos que interaccionan con el plasminógeno se localizan en posiciones cercanas a los aminoácidos de los motifs encontrados.
5. El análisis de los dockings estructurales entre el plasminógeno y estas proteínas multifuncionales indica la presencia de residuos aminoacídicos en el plasminógeno a los que se unen varias de estas proteínas moonlighting.

4. Glosario

IDPs = intrinsically disordered proteins.

GAPDH = gliceraldehído-3-fosfato deshidrogenasa

Cpn10 = 10 kDa chaperonin

Hsp70, Hsp90 = heat shock protein 70, 90.

HMGB1 = high-mobility group protein B1

T β 4 = timosina β -4

bFGF = factor de crecimiento de fibroblastos básico

MMP-12 = metaloproteínasa de la matriz 12

Plg = plasminógeno

tPA, uPA = activadores del plasminógeno

PAIs = inhibidores de los activadores del plasminógeno

MEME = Multiple EM for Motif Elicitation

DREME = Discriminative Regular Expression Motif Elicitation

APID = Agile Protein Interactomes DataServer

PPIs = interacciones proteína-proteína (protein-protein interactions)

PDB = Protein Data Bank

RCSB = Research Collaboratory for Structural Bioinformatics

ADT = AutoDock Tools

NE = No Encontrado

5. Bibliografía

1. Franco Serrano L. Proteínas «Moonlighting»: Identificación y relación con la infección por microorganismos patógenos y clínica humana. Universitat Autònoma de Barcelona; 2018.
2. Kainulainen V, Korhonen T. Dancing to Another Tune—Adhesive Moonlighting Proteins in Bacteria. *Biology (Basel)*. 2014;3(1):178-204.
3. Henderson B, Martin A. Bacterial Moonlighting Proteins and Bacterial Virulence. *Curr Top Microbiol Immunol*. 2011.
4. Chapple CE, Brun C. Redefining protein moonlighting. *Oncotarget*. 2015;6(19):16812-3.
5. Amblee V, Jeffery CJ. Physical features of intracellular proteins that moonlight on the cell surface. *PLoS One*. 2015;10(6):1-16.
6. Yoon JH, Ryu J, Baik SJ. Moonlighting activity of secreted inflammation-regulatory proteins. *Yonsei Med J*. 2018;59(4):463-9.
7. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I, et al. Bioinformatics and Moonlighting Proteins. *Front Bioeng Biotechnol*. 2015;3(June).
8. Jeffery CJ. Why study moonlighting proteins? *Front Genet*. 2015;6(June).
9. Khan I, Chen Y, Dong T, Hong X, Takeuchi R, Mori H, et al. Genome-scale identification and characterization of moonlighting proteins. *Biol Direct*. 2014;9:30.
10. Gancedo C, Flores C, Gancedo JM. The Expanding Landscape of Moonlighting Proteins in Yeasts. *Microbiol Mol Biol Rev*. 2016;80(3):765-77.
11. Jeffery, C. Protein moonlighting: what is it, and why is it important? *Phil Trans R Soc*. 2017.
12. Parikh S, Pancholi V, Song YP, Fischetti VA, Derbise A. Role of the C-Terminal Lysine Residues of Streptococcal Surface Enolase in Glu- and Lys-Plasminogen-Binding Activities of Group A Streptococci. *Infect Immun*. 2003;72(1):94-105.
13. Piatigorsky J, Wistow GJ. Enzyme/crystallins: gene sharing as an evolutionary strategy. *Cell*. 1989;57(2):197-9.
14. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009;27(2):157-67.
15. Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci*. enero de 1999;24(1):8-11.
16. Tompa P, Szász C, Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci*. 2005;30(9):484-9.
17. Henderson B, Martin A. Bacterial Virulence in the Moonlight: Multitasking Bacterial Moonlighting Proteins Are Virulence Determinants in Infectious Disease. *Infect Immun*. 2011;79(9):3476-91.
18. Torres-Machorro AL, Aris JP, Pillus L. A moonlighting metabolic protein influences repair at DNA double-stranded breaks. *Nucleic Acids Res*. 2015;43(3):1646-58.
19. Irving HR, Cahill DM, Gehring C. Moonlighting Proteins and Their Role in the Control of Signaling Microenvironments, as Exemplified by cGMP and Phytosulfokine Receptor 1 (PSKR1). *Front Plant Sci*. 2018;9(Mar):1-8.

20. Tungteakkhun SS, Duerksen-Hughes PJ. Cellular binding partners of the human papillomavirus E6 protein. *Arch Virol.* 2008;153(3):397-408.
21. Espinosa-Cantú A, Ascencio D, Barona-Gómez F, De Luna A. Gene duplication and the evolution of moonlighting proteins. *Front Genet.* 2015;6(Jul):1-7.
22. Yoshida N, Oeda K, Watanabe E, Mikami T, Fukita Y, Nishimura K, et al. Protein function. Chaperonin turned insect toxin. *Nature.* 2001;411(6833):1-44.
23. Copley S. An Evolutionary Perspective on Protein Moonlighting Shelley. *Biochem Soc Trans.* 2014;42(6):1684-91.
24. Zanzoni A, Chapple CE, Brun C. Relationships between predicted moonlighting proteins, human diseases, and comorbidities from a network perspective. *Front Physiol.* 2015;6(JUN):1-8.
25. Sriram G, Martinez JA, McCabe ERB, Liao JC, Dipple KM. Single-Gene Disorders: What Role Could Moonlighting Enzymes Play? *Am J Hum Genet.* 2005;76:911-24.
26. Min K-W, Lee S-H, Baik SJ. Moonlighting proteins in cancer. *Cancer Lett.* 2016;370(1):108-16.
27. Jeffery CJ. Proteins with neomorphic moonlighting functions in disease. *IUBMB Life.* 2011;63(7):489-94.
28. Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA. Online Mendelian Inheritance In Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2002;30(1):52-5.
29. Qin C, Zhang C, Zhu F, Xu F, Chen SY, Zhang P, et al. Therapeutic target database update 2014: A resource for targeted therapeutics. *Nucleic Acids Res.* 2014;42(1):1118-23.
30. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(1):1074-82.
31. Franco-Serrano L, Cedano J, Perez-Pons JA, Mozo-Villarias A, Piñol J, Amela I, et al. A hypothesis explaining why so many pathogen virulence proteins are moonlighting proteins. *Pathog Dis.* 2018;76(5):1-4.
32. Fulde M, Bergmann S. Impact of Streptococcal enolase in virulence. *Moonlighting Proteins Nov Virulence Factors Bact Infect.* 2017;3(Nov):1-8.
33. Backert S, Feller SM, Wessler S. Emerging roles of Abl family tyrosine kinases in microbial pathogenesis. *Trends Biochem Sci.* 2008;33(2):80-90.
34. Fernandes LG, Siqueira GH, Teixeira ARF, Silva LP, Figueredo JM, Cosate MR, et al. Veterinary Immunology and Immunopathology *Leptospira* spp.: Novel insights into host – pathogen interactions. *Vet Immunol Immunopathol.* 2016;176:50-7.
35. Raymond BBA, Djordjevic S. Exploitation of plasmin(ogen) by bacterial pathogens of veterinary significance. *Vet Microbiol.* 2015;178(1-2):1-13.
36. Verhamme IM, Panizzi PR, Bock PE. Pathogen Activators of Plasminogen. *J Thromb Haemost.* 2015;13(1):106-14.
37. Rios-Steiner JL, Schenone M, Mochalkin I, Tulinsky A, Castellino FJ. Structure and binding determinants of the recombinant kringle-2 domain of human plasminogen to an internal peptide from a group A Streptococcal surface protein. *J Mol Biol.* 2001;308(4):705-19.
38. Lähteenmäki K, Kuusela P, Korhonen TK. Bacterial plasminogen activators and receptors. *FEMS Microbiol Rev.* 2001;25(5): 531-52.

39. Sanderson-Smith ML, De Oliveira DMP, Ranson M, McArthur JD. Bacterial Plasminogen Receptors: Mediators of a Multifaceted Relationship. *J Biomed Biotechnol.* 2012;12(1):1-14.
40. Peetermans M, Vanassche T, Liesenborghs L, Lijnen RH, Verhamme P. Bacterial pathogens activate plasminogen to breach tissue barriers and escape from innate immunity. *Crit Rev Microbiol.* 2016;42(6):866-82.
41. Li W, Eng M, Chong SSN, Huang EY, Tuan T. Perspective Article Plasminogen activator / plasmin system : A major player in wound healing ? *Wound Repair Regen.* 2003.
42. Lähteenmäki K, Edelman S, Korhonen TK. Bacterial metastasis: The host plasminogen system in bacterial invasion. *Trends Microbiol.* 2005;13(2):79-85.
43. Bhattacharya S, Ploplis VA, Castellino FJ. Bacterial Plasminogen Receptors Utilize Host Plasminogen System for Effective Invasion and Dissemination. *J Biomed Biotechnol.* 2012;2012:1-19.
44. Fulde M, Steinert M, Bergmann S. Interaction of streptococcal plasminogen binding proteins with the host fibrinolytic system. *Front Cell Infect Microbiol.* 2013;3.
45. Degen JL, Bugge TH, Goguen JD. Fibrin and fibrinolysis in infection and host defense. *J Thromb Haemost.* 2007;5(1):24-31.
46. Ayón-Núñez DA, Fragoso G, Bobes RJ, Laclette JP. Plasminogen-binding proteins as an evasion mechanism of the host's innate immunity in infectious diseases. *Biosci Rep.* 2018;38(5).
47. Feltcher ME, Gibbons HS, Ligon LS, Braunstein M. Protein Export by the Mycobacterial SecA2 System Is Determined by the Preprotein Mature Domain. *J Bacteriol.* 2013;195(4):672-81.
48. Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, et al. Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nat Rev Microbiol.* 2015;13(6):343-59.
49. Kornblatt MJ, Kornblatt JA, Hancock MA. The interaction of canine plasminogen with streptococcus pyogenes enolase: They bind to one another but what is the nature of the structures involved? *PLoS One.* 2011;6(12).

6. Anexos

- 1) Material suplementario 0 (MS0): diagrama de la planificación del TFM.
- 2) Material suplementario 1 (MS1): listado completo de las 240 proteínas moonlighting escogidas para el estudio.
- 3) Material suplementario 2 (MS2): listado con la clasificación de las proteínas moonlighting del estudio en función de la proteína a la que se une como función moonlighting.
- 4) Material suplementario 3 (MS3): listado con las parejas de proteínas de unión que comparten las proteínas moonlighting estudiadas.
- 5) Material suplementario 4 (MS4): listado con la clasificación de las proteínas moonlighting del estudio en función del microorganismo al que pertenecen.
- 6) Material suplementario 5 (MS5): análisis completo de DREME y todos sus resultados.
- 7) Material suplementario 6 (MS6): código completo del programa desarrollado en Python para la búsqueda de motifs en proteínas.
- 8) Material suplementario 7 (MS7): listados de proteínas que contienen los motifs seleccionados, resultantes de la búsqueda en el programa de Python.
- 9) Material suplementario 8 (MS8): información detallada sobre las interacciones de proteínas moonlighting descubiertas en APID.
- 10) Material suplementario 9 (MS9): listado completo de aminoácidos del plasminógeno y aminoácidos de cada una de las proteínas moonlighting involucrados en esta interacción y que forman parte de la zona de unión modelada por ClusPro 2.0.