

# Proteínas multifuncionales, enfermedades humanas y dianas farmacológicas: una actualización

**José Félix Rojas Cabeza**

Máster Universitario en Bioinformática y Bioestadística

Área 2. Subárea 8: Bioinformática y proteómica en virulencia y enfermedades humanas

**Marc Maceira Duch**

**Luis Franco Serrano**

16 de noviembre de 2020



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

**Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)**

**A) Creative Commons:**



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento [3.0 España de Creative Commons](#)

**B) GNU Free Documentation License (GNU FDL)**

Copyright © 2020 José Félix Rojas Cabeza.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free

Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

### **C) Copyright**

© José Félix Rojas Cabeza.

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Proteínas multifuncionales, enfermedades humanas y dianas farmacológicas: una actualización</i>
<b>Nombre del autor:</b>	<i>José Félix Rojas Cabeza</i>
<b>Nombre del consultor/a:</b>	<i>Luis Franco Serrano</i>
<b>Nombre del PRA:</b>	<i>Marc Maceira Duch</i>
<b>Fecha de entrega (mm/aaaa):</b>	01/2021
<b>Titulación:</b>	Máster en Bioinformática y Bioestadística
<b>Área del Trabajo Final:</b>	<i>Área 2. Subárea 8: Bioinformática y proteómica en virulencia y enfermedades humanas</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Proteómica (Proteomics), redes PPI (PPI networks), Proteínas multifuncionales (Moonlighting proteins)</i>
<b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i>	
<p>La multifuncionalidad de proteínas se refiere a su habilidad para llevar a cabo un cabo dos o más funciones biológicas. El objetivo de este trabajo fue examinar con herramientas bioinformáticas la relación entre proteínas multifuncionales y enfermedades humanas, tomando en cuenta los mecanismos moleculares de la enfermedad y su interacción con dianas farmacológicas.</p> <p>La aproximación tomada involucraba evaluar estadísticamente la asociación entre enfermedades humanas y proteínas multifuncionales, y revisar bases de datos de interacción de proteínas para analizar la interacción local entre ellas y un análisis funcional de las proteínas de estas redes.</p> <p>No se observaron suficientes evidencias para afirmar que el número de clases funcionales de proteínas moonlighting en humanos es distinto al resto de los organismos. Sí se obtuvo suficiente información para sugerir que las proteínas moonlighting son propensas a estar involucradas en enfermedades humanas.</p> <p>Además, se hicieron análisis de redes de interacción de proteínas moonlighting involucradas en enfermedades humanas, y se encontraron 7 grupos de interacción.</p> <p>También se hizo un análisis de términos GO de cada grupo asociado a enfermedades, para una caracterización de los procesos biológicos, funciones</p>	

moleculares y compartimientos celulares donde se encontraban.

Se encontraron redes de interacción proteína-proteína para los grupos de interacción asociados a enfermedades, y se hizo un acoplamiento molecular entre proteínas pertenecientes a uno de los grupos.

Se puede concluir que en las proteínas moonlighting humanas asociadas a enfermedades existen redes de interacción de tamaño considerable que tienen conexiones entre sí.

**Abstract (in English, 250 words or less):**

Protein multifunctionality refers to their ability to carry out two or more biological functions. The objective of this work was to examine with bioinformatics tools the relationship between multifunctional proteins and human diseases, considering the molecular mechanisms of the disease and its interaction with drug targets.

The approach taken involved statistically evaluating the association between human diseases and multifunctional proteins and reviewing protein interaction databases to analyze the local interaction between them and a functional analysis of the proteins in these networks.

Not enough evidence was observed to affirm that the number of functional classes of moonlighting proteins in humans is different from other organisms.

Enough information was obtained to suggest that moonlighting proteins are likely to be involved in human disease.

In addition, analysis of interaction networks of moonlighting proteins involved in human diseases were made, and 7 interaction groups were found.

An analysis of GO terms of each group associated with diseases was also done, for a characterization of the biological processes, molecular functions, and cellular compartments where they were found.

Protein-protein interaction networks were found for the disease-associated interaction groups, and a molecular coupling was made between proteins belonging to one of the groups.

It can be concluded that in human moonlighting proteins associated with diseases there are interaction networks of considerable size that have connections between them.

# Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	4
1.2.1 Objetivo General:.....	4
1.2.2 Objetivos Específicos:.....	4
1.3 Enfoque y método seguido.....	4
1.4 Planificación del Trabajo.....	5
1.5 Breve resumen de productos obtenidos.....	5
1.6 Breve descripción de los otros capítulos de la memoria.....	5
2. Resto de capítulos.....	7
<b>2.1 INTRODUCCIÓN</b> .....	7
2.1.1 Proteínas Moonlighting: Concepto e introducción.....	7
2.2.2 Proteínas Moonlighting: Historia.....	7
2.2.3 Proteínas Moonlighting: Características.....	9
Categorías funcionales.....	11
2.2.4 Proteínas Moonlighting: Ejemplos principales.....	12
2.2.5 Proteínas Moonlighting: Aspectos evolutivos.....	13
2.2.6 Proteínas Moonlighting: Identificación y predicción.....	15
2.2.7 Proteínas Moonlighting: Relación con virulencia.....	15
2.2.8 Proteínas Moonlighting: Relación con enfermedades humanas.....	16
2.2 Métodos.....	17
2.2.1 Bases de datos, servidores y herramientas bioinformáticas utilizadas.....	17
2.2.2 Selección de un listado proteínas moonlighting asociadas a enfermedades humanas.....	17
2.2.3 Odds Ratio (Razón de Probabilidades).....	17
2.2.4 Extracción de códigos GO.....	18
2.2.5 Análisis de funciones de proteínas moonlighting asociadas a enfermedades:.....	18
2.2.6 Redes de interacción.....	19
2.2.7 Docking con ClusPro 2.0.....	19
2.3 RESULTADOS.....	21
2.3.1 Selección de un listado proteínas moonlighting asociadas a enfermedades humanas.....	21
2.3.1.1 Interacción entre proteínas moonlighting humanas.....	21
2.3.1.2 Interacción entre proteínas moonlighting humanas asociadas a enfermedades.....	23
2.3.2 Odds ratio.....	23
2.3.3 Extracción de códigos GO de proteínas moonlighting.....	24
2.3.4 Análisis de funcionalidad de proteínas moonlighting.....	25
2.3.5 Redes de interacción.....	30
2.3.6 Docking con ClusPro 2.0.....	31
2.4 DISCUSIÓN.....	33
2.4.1 Interacción entre proteínas moonlighting asociadas a enfermedades humanas.....	33
2.4.3 Análisis de códigos GO, funcionalidad de proteínas moonlighting y redes de interacción.....	34
2.4.4 Análisis del Docking.....	38

2.4.7 Observaciones finales .....	39
3. Conclusiones.....	40
4. Glosario .....	41
5. Bibliografía .....	42
6. Anexos .....	49

## Lista de ilustraciones

Ilustración 1. <i>Factores que pueden afectar la función de una proteína multifuncional.</i> .....	3
Ilustración 2. <i>Distintas formas de multifuncionalidad en Ure2.</i> .....	9
Ilustración 3. <i>Ejemplos de mecanismos para cambio de función de una proteína multifuncional.</i> .....	10
Ilustración 4. <i>Proteínas moonlighting en la glicólisis.</i> .....	13
Ilustración 5. <i>Proteínas moonlighting en el ciclo de los ácidos carboxílicos.</i> ....	14
Ilustración 6. Encabezado de salida de GOTrapper con threshold = 4. ....	18
Ilustración 7. Código utilizado para asignar los códigos GO a cada gen.....	18
Ilustración 8. Código utilizado para generar el archivo que se colocó en STRING para generar la red del grupo enf_7.....	19
Ilustración 9. Grupos de interacción entre proteínas moonlighting humanas. ( <i>enlace</i> ) .....	21
Ilustración 10. Ejemplo de interacción entre ENO1 y GAPDH en la base de datos STRING.....	22
Ilustración 11. <i>Test F y two sample t test de Welch.</i> .....	23
Ilustración 12. <i>Ejemplo de los cinco primeros genes en la tabla.</i> .....	25
Ilustración 13. Red de interacción de proteínas moonlighting humanas ligadas a enfermedades. <i>123 proteínas (enlace)</i> .....	25
Ilustración 14. Procesos biológicos asociados a los genes del Grupo 1_ enf. ( <i>enlace</i> ) .....	27
Ilustración 15. Mecanismos moleculares asociados a los genes del Grupo 1_ enf. ( <i>enlace</i> ) .....	28
Ilustración 16. Compartimientos celulares asociados a los genes del Grupo 1_ enf. ( <i>enlace</i> ) .....	29
Ilustración 17. <i>Red de interacción de proteínas del grupo 2_ enf. 1029 ítems.</i> ( <i>enlace</i> ) .....	30
Ilustración 18. Ejemplos de Docking de TP53 Cadena A con MDM2 Cadena A, y ATF2 Cadena D modelado con ClusPro 2.0 en un solvente balanceado. ....	31
Ilustración 19. Procesos biológicos asociados a los genes del grupo G4_ enf .	36
Ilustración 20. Mecanismos moleculares asociados a los genes del grupo G4_ enf .....	36
Ilustración 21. Compartimientos celulares asociados a los genes del grupo G4_ enf .....	37
Ilustración 22. Red de interacción de proteínas del grupo 4_ enf. 1735 ítems. ( <i>enlace</i> ) .....	38

## Lista de Tablas

Tabla 1. Bases de datos, servidores y herramientas bioinformáticas a evaluarse o utilizarse.....	17
Tabla 2. Grupos de interacción entre proteínas moonlighting humanas. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de BioGrid .....	22
Tabla 3. Resultados de análisis estadístico odds ratio, que muestra el grado de significación, con un intervalo de confianza de 95%. .....	24
Tabla 4. Grupos de interacción entre proteínas moonlighting humanas relacionadas a enfermedades. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de BioGrid. ....	26
Tabla 5. Redes de interacciones de proteínas para los grupos de proteínas moonlighting humanas asociadas a enfermedades .....	30
Tabla 6. Características comunes entre proteínas multifuncionales humanas y proteínas multifuncionales humanas relacionadas a enfermedades .....	35

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

La proteómica ha cambiado de un enfoque en moléculas individuales a una perspectiva integrada, en la cual las proteínas emergen como actores dinámicos con múltiples funciones, en lugar de especialistas monofuncionales (1).

A Medida que se elucidan más secuencias del genoma, existe una necesidad creciente de información sobre las funciones de los millones de proteínas codificadas. La función de una proteína recién secuenciada a menudo se estima mediante el alineamiento de su secuencia con las secuencias de proteínas con funciones conocidas. Sin embargo, las superfamilias de proteínas pueden contener miembros que comparten una secuencia de aminoácidos significativa y una homología estructural alta, pero catalizan diferentes reacciones o actúan sobre diferentes sustratos(2).

Algunas proteínas homólogas se diferencian por tener una segunda o incluso una tercera función, llamadas proteínas multifuncionales. Las proteínas multifuncionales (moonlighting) exhiben más de una función bioquímica o biofísica fisiológicamente relevante dentro de una cadena polipeptídica(3).

Más recientemente, se encontró que la mayoría de las superfamilias de proteínas también incluyen pseudoenzimas, una proteína o un dominio dentro de una proteína, que tiene un pliegue tridimensional que se asemeja a una enzima catalíticamente activa convencional, pero no tiene actividad catalítica(2).

La anotación del repertorio funcional completo de proteínas ha impactado los campos de la bioquímica y la genética, y continuará influyendo en las cuestiones científicas básicas y aplicadas(1).

La ocurrencia de multifuncionalidad (moonlighting) es relevante en los siguientes aspectos de la investigación en bioquímica de proteínas(4):

- *Anotaciones genómicas funcionales*: Imprecisión, descripciones insuficientes.
- *Estudio de las rutas metabólicas*: Análisis y predicción
- *Knock-Outs y Knock-Downs*: Dificulta la interpretación de los efectos observados, debido a que no se puede saber cuál de las rutas bioquímicas afectadas causa el efecto observado (o la ausencia de este).
- *Análisis de redes de interactómica*: complica la interpretación de redes de interactómica.
- *Farmacología*: Selección de fármacos, efectos secundarios, toxicidad, inespecificidad no documentada.
- *Neomorphic Moonlighting*(5): una proteína multifuncional que adquiere una función que causa una enfermedad.

- *Resistencia bacteriana a antibióticos*: Un ejemplo es el caso de Glutamato Racemasa, una enzima de *Mycobacterium tuberculosis* que está involucrada en resistencia a ciprofloxacina(6).
- *Mecanismos de virulencia y adhesión al hospedero*: algunas proteínas de la glicólisis tienen una segunda función, como factor de virulencia, cuando se ubican en la membrana, debido a que facilitan la unión de las células a las del hospedero(7).

Este trabajo pretende mejorar los conocimientos actuales sobre el involucramiento de las proteínas multifuncionales en los mecanismos moleculares de enfermedades humanas. Específicamente en términos de identificar causas de posibles efectos secundarios o reutilizar medicamentos para fines no descritos anteriormente.

Para esto se requiere mejorar la comprensión actual de funciones bioquímicas y roles celulares de proteínas multifuncionales, para mejorar los programas que identifican funciones a partir de secuencia o estructura, y mejorar las anotaciones de las bases de datos de secuencia y estructurales. Así como la comprensión de la amplia diversidad de funciones de las proteínas (2).

Al ser las proteínas multifuncionales centrales para el papel de muchas enfermedades, el desarrollo de nuevos tratamientos se vería favorecido al abordar preguntas como el mecanismo por el que alcanzan distintas ubicaciones celulares (2,3,8) o cómo afectar una sola función como diana terapéutica(3,9).

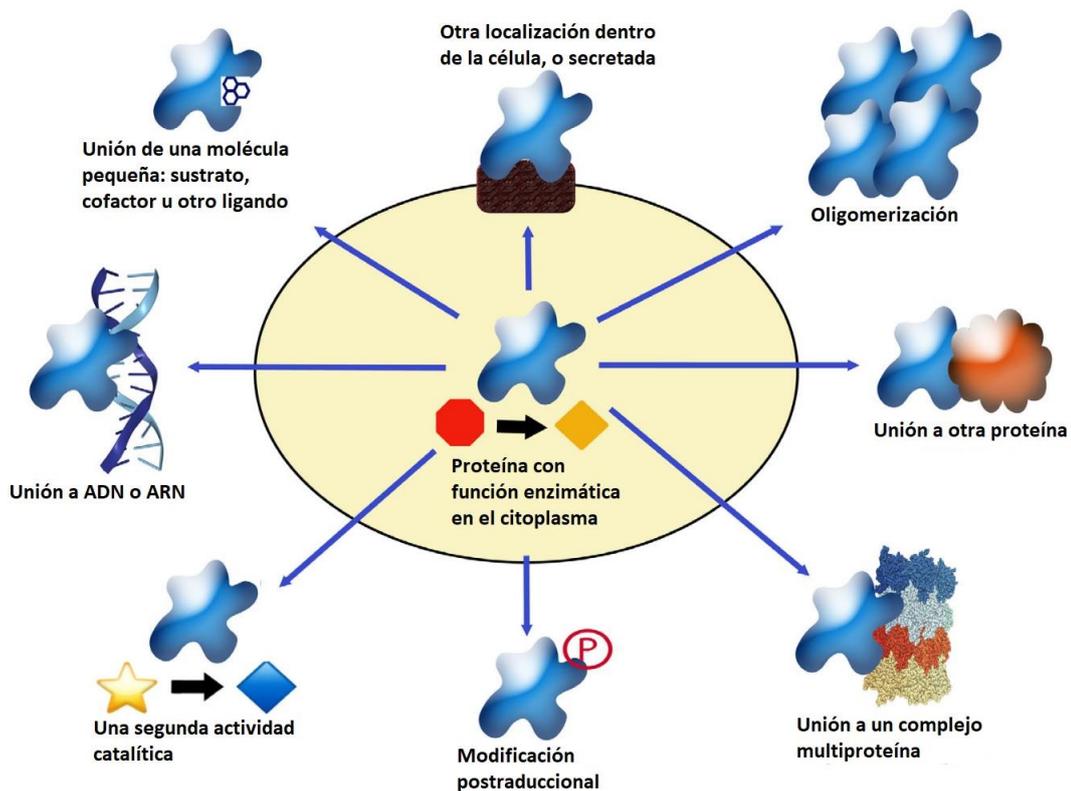
Si bien las proteínas multifuncionales pueden ser una de las causas de las reacciones adversas a medicamentos, también es posible especular que dichas proteínas también pueden ser dianas terapéuticas útiles(10). Especialmente considerando casos como peptidilprolil isomerasa de *Legionella pneumophila*, originalmente identificada como un potenciador de infectividad de macrófagos; es un factor de virulencia importante en ciertas bacterias intracelulares, y les ayuda a entrar al macrófago uniéndose a colágeno tipo IV(11). Posteriormente, se llegó al desarrollo de varios inhibidores sintéticos de la adhesividad de Mip, incluidos algunos compuestos de cicloheximida novedosos con actividad antimicrobiana contra *L. pneumophila*(12).

Luego, la ataxia telangiectasia mutada (ATM) quinasa se llama así porque las mutaciones en esta proteína causan ataxia telangiectasia, una enfermedad neurodegenerativa poco común. Esta quinasa es una proteína multifuncional y ahora se ve como un objetivo terapéutico(13).

A medida que aumenta el número de proteínas multifuncionales y comienza a identificarse el número de sitios activos multifuncionales, es más probable que las importantes acciones en que participan estas proteínas conviertan a algunas de ellas en objetivos relevantes para el desarrollo de nuevos agentes farmacéuticos(10).

Entre las limitaciones del estudio, se encuentra que la forma de demostrar que la proteína es en efecto multifuncional debe ser experimental(3,14), por lo que los resultados obtenidos deben esperar confirmación experimental. Además, resultados experimentales inesperados suelen ser descartados por los investigadores, debido a que en general, se considera que los datos experimentales de proteómica contienen muchos falsos positivos, estimados en alrededor del 20% (14,15).

Otro factor que complica el estudio de las proteínas multifuncionales es que existen muchos factores que pueden afectar la función que exhibe una proteína. En la célula, los niveles de expresión, los moduladores alostéricos, las modificaciones postraduccionales, el secuestro y otros factores pueden afectar el nivel de función de la proteína(8,16). Para las proteínas multifuncionales, factores celulares como estos también pueden afectar el tipo de función de la proteína. (ver ilustración 1)



**Ilustración 1. Factores que pueden afectar la función de una proteína multifuncional. (modificado de (17))**

Recientemente se han utilizado las siguientes aproximaciones que utilizan secuencias, estructuras, interactómica, algoritmos y programas para investigar multifuncionalidad (14): Búsquedas de homología remota utilizando Psi-Blast (18), Detección de motivos y dominios funcionales, Análisis de datos de bases de datos de interacción proteína-proteína (PPI)(14,19), Análisis de coincidencia de la consulta de la secuencia con una base de datos 3D (ej. algoritmos como PiSITE (20) o Phyre2 (21)), Análisis de correlación de mutaciones entre aminoácidos mediante algoritmos como MISTIC2 (22). En el trabajo se pretende retomar algunas de estas técnicas para estudiar la

interacción de proteínas moonlighting humanas asociadas a enfermedades con otras proteínas, buscando esclarecer los mecanismos moleculares de enfermedades (23–25).

## 1.2 Objetivos del Trabajo

### 1.2.1 Objetivo General:

Examinar la relación entre proteínas multifuncionales y enfermedades humanas, tomando en cuenta los mecanismos moleculares de la enfermedad y su interacción con dianas farmacológicas.

### 1.2.2 Objetivos Específicos:

1.2.2.1 Evaluar mediante un odds ratio (razón de probabilidades) la asociación entre enfermedades humanas y proteínas multifuncionales.

1.2.2.2 Revisar si se pueden agrupar los mecanismos moleculares en que intervienen las proteínas multifuncionales asociadas con enfermedades.

1.2.2.3 Mapear y vincular enfermedades a funciones de proteínas multifuncionales.

1.2.2.4 Revisar bases de datos de interacción de proteínas para hacer análisis de interacción local entre las proteínas asociadas encontradas en el mapeado.

## 1.3 Enfoque y método seguido

Para cumplir con los objetivos, establecidos en 1.2, se utilizó una estrategia que consistió, grosso modo, en seleccionar un conjunto de proteínas multifuncionales, tomadas de la base de datos de proteínas multifuncionales MultitaskProtDB-II (26), y estudiar la interacción de proteínas de interés involucradas en enfermedades humanas con otras proteínas metabólicas de interés.

Se empezó por obtener un listado de proteínas moonlighting humanas implicadas en enfermedades humanas, y se hizo un análisis de mecanismos moleculares y su asociación con enfermedades humanas. Se verificaron las regiones de interacción de los mecanismos más relevantes, incluyendo interacciones de proteínas multifuncionales entre sí. Se realizaron búsquedas en bases de datos de interactómica, para hacer dockings estructurales. La base de datos STRING (27) y el software Cytoscape (28) se utilizaron para construir la interacción proteína-proteína (PPI). También se modeló la interacción proteína-proteína con

El análisis de interacción local entre proteínas asociadas encontradas con el mapado se realizará de la siguiente forma: Se efectuará un análisis manual del listado de proteínas a las que se unen las proteínas moonlighting, para examinar las principales proteínas interactoras que presentan, y si comparten algunas de ellas.

También se analizaron los datos en BioGrid (29)(30) y APID (31) para los interactores de algunas proteínas moonlighting. La idea será evaluar con cuáles proteínas interactúan estas proteínas moonlighting. Por ejemplo, se espera verificar interacciones de patógenos microbiológicos con plasminógeno(32,33), y casos de interacciones de proteínas moonlighting humanas que se relaciones con.

#### 1.4 Planificación del Trabajo

Los objetivos a lograr han sido divididos en una serie de tareas, cuya elaboración y cumplimiento se ha repartido en dos fases, correspondientes a las PEC2 y PEC3 (Ver anexo 1). La primera fase tiene que ver con la revisión bibliográfica y la determinación de viabilidad de diferentes objetivos. Se elaboró el listado de proteínas moonlighting. Además, se llevó a cabo la revisión de la interacción de proteínas entre sí mismas, utilizando STRING(27).

Durante La segunda parte del proyecto se realizarán los análisis correspondientes al estudio interactómico, modelado de dockings estructurales y modelado de dockings estructurales y análisis de interacciones entre proteínas. También se ha tomado en cuenta que estas actividades deberán realizarse en paralelo con la redacción de la memoria.

#### 1.5 Breve resumen de productos obtenidos

Los principales resultados que se obtuvieron durante este trabajo son:

- Listado y análisis de proteínas moonlighting asociadas a enfermedades humanas.
- Análisis de las funciones de estas proteínas, y clasificación de acuerdo con el tipo de enfermedad con que están involucradas
- Elaboración de un script en python para extraer los códigos GO involucrados en cada proteína
- Listado de un conjunto de procesos biológicos y funciones moleculares en enfermedades humanas
- Estudio de la interacción de proteínas moonlighting involucradas en enfermedades.
- Docking molecular de interacción de TP53 con ATF2, y TP53 con MDM2 hecho en ClusPro 2.0.
- Inventario de Procesos Biológicos, Funciones Moleculares, y Compartimientos Celulares de varios grupos de proteínas moonlighting relacionadas con enfermedades hecho con GONet.

#### 1.6 Breve descripción de los otros capítulos de la memoria

La primera parte corresponde a la introducción a las proteínas moonlighting, y una explicación de sus características más importantes. Se exponen los criterios principales de la definición y algunos ejemplos, relacionados con dos ámbitos de interés: enfermedades humanas, y limitaciones del estudio de este tipo de proteínas.

La segunda parte corresponde a los métodos que se emplearon para la obtención de los resultados, en términos de bases de datos, y herramientas utilizadas, así como el proceso seguido para realizar los análisis del proyecto.

La tercera parte muestra los resultados obtenidos, incluyendo gráficos, cálculos, ilustraciones y tablas de cada análisis realizado.

Finalmente, se realiza una discusión general, donde se hace la interpretación y comentario de los resultados del trabajo, en función de las conclusiones que se pueden obtener del proyecto, algunas limitaciones o mejoras que podrían hacerse sobre el trabajo, así como recomendaciones para estudios posteriores.

## 2. Resto de capítulos

### 2.1 INTRODUCCIÓN

#### 2.1.1 Proteínas Moonlighting: Concepto e introducción

La proteómica ha cambiado de un enfoque en moléculas individuales a una perspectiva integrada, en la cual las proteínas emergen como actores dinámicos con múltiples funciones, en lugar de especialistas monofuncionales (1). A medida que se elucidan más secuencias del genoma, existe una necesidad creciente de información sobre las funciones de los millones de proteínas codificadas.

La función de una proteína recién secuenciada a menudo se estima mediante el alineamiento de su secuencia con las secuencias de proteínas con funciones conocidas. Sin embargo, las superfamilias de proteínas pueden contener miembros que comparten una secuencia de aminoácidos significativa y una homología estructural alta, pero catalizan diferentes reacciones o actúan sobre diferentes sustratos (2).

Algunas proteínas homólogas se diferencian por tener una segunda o incluso una tercera función, llamadas proteínas multifuncionales: estas proteínas realizan más de una función biológica en diferentes localizaciones celulares (1,3–5,9,34,35). Las proteínas multifuncionales (moonlighting) exhiben más de una función bioquímica o biofísica fisiológicamente relevante dentro de una cadena polipeptídica (3); éstas funciones no se encuentran relacionadas (14,36–38).

Ambas funciones no se encuentran necesariamente separadas en dominios distintos (3). La multifuncionalidad (moonlighting) referida no es debida al splicing alternativo o a la fusión de genes, o fragmentos que expresen diferentes funciones luego de la proteólisis (4,37). Tal multifuncionalidad tampoco se refiere a proteínas homólogas con funciones diferentes, o a factores de transcripción que se unen a diferentes promotores, o enzimas con diversos sustratos (35,38).

Una proteína multifuncional (moonlighting) tiene una función canónica y una o varias funciones “moonlighting”(26). La función canónica es la primera función descrita y corresponde sólo al azar, porque no implica ninguna relevancia funcional especial, sino cronológica. Las funciones descritas luego de la primera serían “moonlighting”(39). La adquisición de una función no implica la pérdida de la función canónica(40).

#### 2.2.2 Proteínas Moonlighting: Historia

El cambio de perspectiva de la proteómica, así como la anotación del repertorio funcional completo de proteínas han impactado la bioquímica y la genética, y seguirán influyendo en las ciencias básicas y aplicadas, desde el problema del genotipo al fenotipo hasta nuestra comprensión de las patologías humanas y el diseño de fármacos(1).

La multifuncionalidad ha ido cambiando a medida que avanzan los conocimientos de biología. En los 40's se pensaba en un paradigma de "una enzima – una proteína"(41) y se creía que las proteínas eran moléculas altamente específicas y monofuncionales(1). Esta forma de pensar influyó tremendamente las anotaciones funcionales de los sistemas biológicos: una vez que se asigna una función a determinada proteína, usualmente las otras funciones a cualquier nivel son ignoradas (1,42). Sin embargo, se han descrito muchos mecanismos de multifuncionalidad(1).

La primera proteína moonlighting descrita fue aldehído deshidrogenasa, en 1987(2). Wistow y Pitagorsky mostraron que la proteína, además de tener una función estructural en la lente del ojo humano, también tenía actividad enzimática(2,14). Interesantemente, se han encontrado proteínas moonlighting en el cristalino de algunos organismos como proteínas estructurales, que también funcionan como enzimas citosólicas, como Alfa-enolasa, Arginosuccinato-liasa y Lactato-deshidrogenasa(2,38,43).

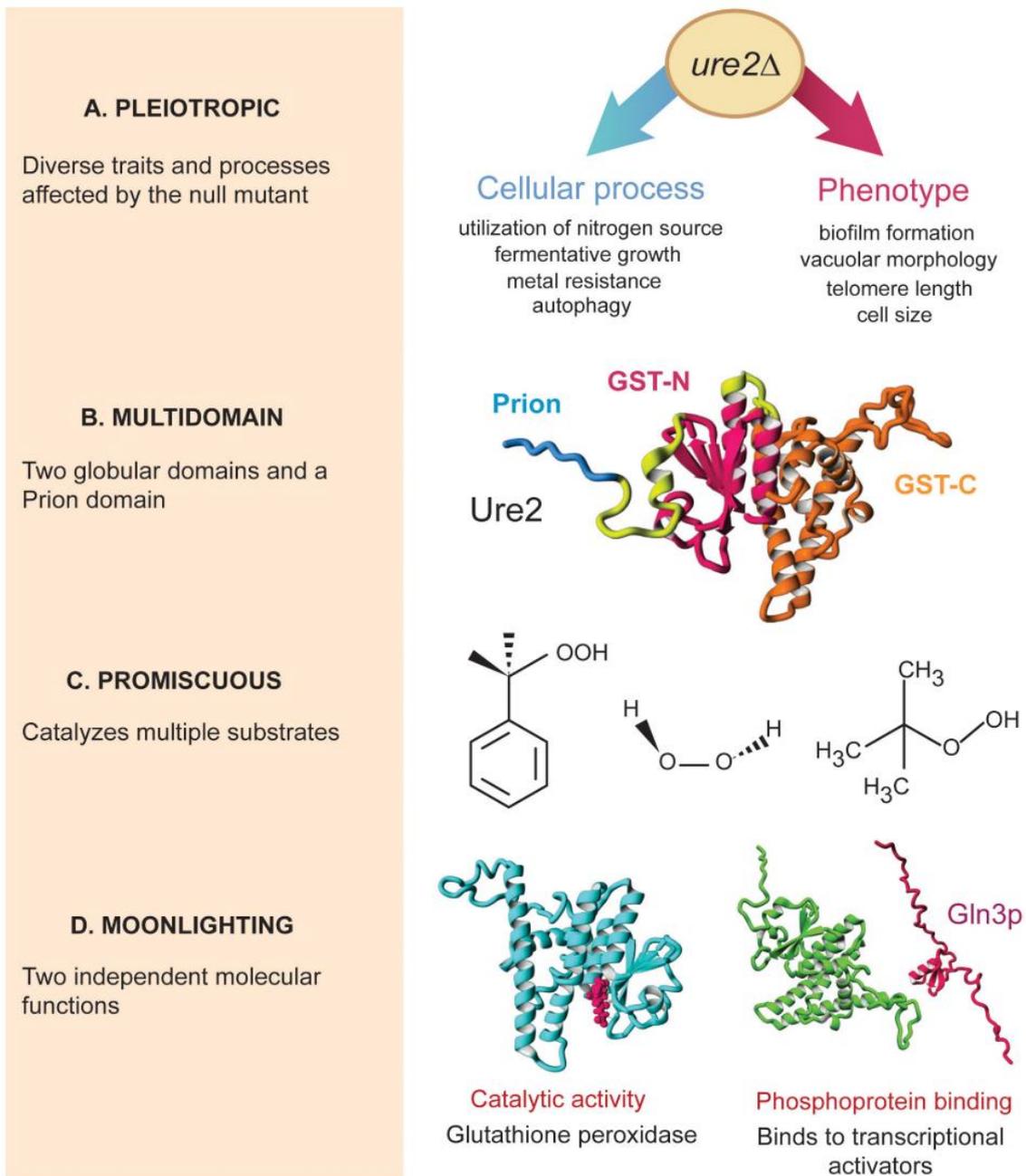
Las proteínas moonlighting han recibido diferentes nombres: un primer intento fue "gene sharing" (genes compartidos)(44), pero debido a que en el splicing alternativo también se comparten genes, no es un término adecuado. Moonlighting se refiere al caso en que sólo un polipéptido realiza dos funciones no relacionadas.

El término moonlighting fue introducido por Jeffery(8) aludiendo a que la palabra en inglés significa "tener un segundo trabajo". Se excluye la fusión génica como causa de la segunda función y que se presenten dos funciones moleculares en el mismo centro activo(14,45). Esto presenta la complicación de que determinar si ha ocurrido una función génica no es factible, porque la historia evolutiva de la proteína de interés no es necesariamente un dato disponible(46), lo que influyó en el uso del término de la forma como se propuso.

La denominación "proteínas promiscuas" se introdujo posteriormente (47). Este tipo de proteína realiza una función molecular que interviene en varias funciones biológicas. Por ejemplo: las proteínas GPCR interactúan de manera promiscua con la subunidad alfa de diferentes proteínas G (48).

Un trabajo más reciente indica que la revisión del concepto de proteína moonlighting ha continuado expandiéndose en una definición menos restrictiva, y hoy en día también se usa para nombrar proteínas que muestran roles no relacionados en los organismos, independientemente de su historia evolutiva (1). En la ilustración 2 se ilustra un ejemplo de los casos descritos anteriormente en el contexto de Ure2 (una proteína moonlighting de levadura).

A pesar de que algunos autores sugieren que las proteínas intrínsecamente desordenadas (IDPs o IUPs, por sus siglas en inglés) interactúan con partners distintos, y que dan lugar a varias funciones (14,49), otras evidencias (50) indican que la mayoría de las proteínas multifuncionales no pertenece a este tipo de proteínas.



**Ilustración 2. Distintas formas de multifuncionalidad en Ure2. (Tomado de (1))**

La proteína Ure2 de levadura es una peroxidasa que puede formar priones y está directamente involucrada en represión del catabolito de nitrógeno al unirse y secuestrar factores de transcripción fosforilados en el citoplasma, tales como Gln3.

A. Pleiotrópica: características diversas y procesos afectados por el mutante nulo (*ure2Δ*).

B. Multidominio: dos dominios globulares y un dominio de prion.

C. Promiscua: Cataliza múltiples sustratos.

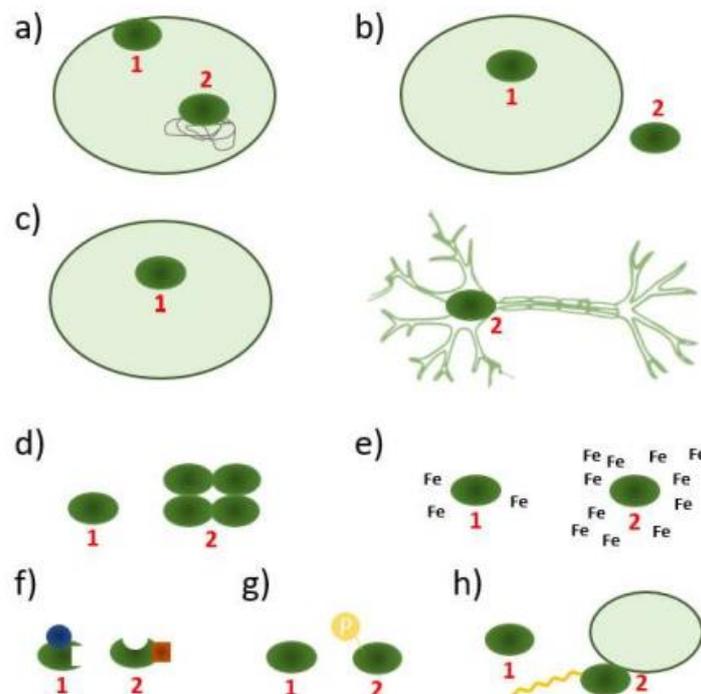
D. Moonlighting: Dos funciones moleculares independientes.

### 2.2.3 Proteínas Moonlighting: Características

Los aspectos que suelen asociarse a la multifuncionalidad son:

- Localización celular: Si la proteína se encuentra en más de un compartimiento celular, donde no realiza su función (9). Una proteína puede tener diferentes funciones en diferentes ubicaciones dentro de

- una célula (por ejemplo, cuando está unida a la membrana celular, o al DNA). Un caso (Ilustración 3 a).
- Participación en secreción: Si la proteína presenta alguna actividad cuando está en citosol y otra cuando es secretada (35). Una proteína puede tener actividad enzimática en el citoplasma, pero sirve como factor de crecimiento cuando es secretada (Ilustración 3 b).
  - Expresión diferencial: Cuando realiza distintas funciones, dependiendo del factor de activación o la célula que la exprese (4). Una proteína puede tener diferentes funciones cuando es expresada por diferentes tipos de células (por ejemplo, una célula endotelial en oposición a una neurona) (Ilustración 3 c).
  - Oligomerización: Si presenta un tipo de actividad cuando se encuentra en forma monomérica y otro cuando se encuentra polimerizada (50). Un multímero puede tener una actividad diferente a la de un monómero (Ilustración 3 d).
  - Concentración de un metabolito: La unión del sustrato, producto o cofactor puede provocar un cambio en la actividad (51) (Ilustración 3 e).
  - Presentar distintos sitios de unión: Algunas proteínas pueden tener diferentes sitios de unión para diferentes sustratos (52) (Ilustración 3 f).
  - Presentar partners de interacción inesperados: Si se encuentra una proteína interactuando como partner de una proteína inesperada. Por ejemplo, Acetil-glutamina-quinasa se une a varias regiones del DNA, por esto se descubrió que también actúa como factor de transcripción (53).
  - Modificación postraducciona: La proteína tiene una función, y al recibir una modificación postraducciona cambia de función. Ejemplo: Fosfoglucosa isomerasa, que fosforilada actúa como factor de motilidad autocrina (54).



**Ilustración 3. Ejemplos de mecanismos para cambio de función de una proteína multifuncional. (tomado de (55))**

También es conveniente aclarar el concepto de proteína moonlighting estableciendo cuáles casos no pueden considerarse como tales:

- Variantes protéicas producto del splicing alternativo (son diferentes proteínas)(37).
- Proteínas con diferentes funciones biológicas producto de la proteólisis de una proteína original(4).
- Factores de transcripción que se unen a distintos promotores(46).
- Enzimas que intervienen en varias rutas aportando la misma función molecular(14).

Como se mencionó en la sección de historia, algunos autores son más restrictivos al utilizar la definición de moonlighting, excluyendo las proteínas producto de la fusión de dos genes, y proteínas con dos actividades enzimáticas en el mismo centro activo.

Las proteínas moonlighting son muy relevantes para la bioquímica de proteínas y la investigación. Al estar involucradas en diversos procesos fisiológicos, es importante caracterizar sus funciones, para entender procesos fisiológicos (particularmente de enfermedades) y mecanismos de virulencia, entre otros.

#### Categorías funcionales

Las proteínas moonlighting se pueden agrupar sobre los pares de funciones que presentan. En la base de datos MultitaskProtDB-II (26) se observó que las clases funcionales de éstas proteínas suelen ser enzimas y proteínas de unión a ácidos nucleicos. La segunda combinación más frecuente es enzima y adhesión celular(4).

Las proteínas moonlighting también pueden presentar otras categorías funcionales. Por ejemplo: receptores, proteínas estructurales, chaperonas, etc. Otros factores de interés pueden ser la temporalidad de la expresión y la localización celular de las mismas, debido a que pueden utilizarse para determinar si la proteína es moonlighting(37,38).

Interesantemente, pocas proteínas moonlighting son de membrana. Posiblemente el requerimiento de aminoácidos no polares en los dominios transmembrana opera negativamente contra la emergencia de nuevas funciones. Sin embargo, en el caso de microorganismos patógenos, las proteínas citosólicas o nucleares se transportan a la membrana sin péptido señal, para participar en mecanismos de adhesión al plasminógeno del hospedero(4). Muchos microorganismos utilizan enolasa para este fin (46,50). El mecanismo por el que ocurre este proceso se desconoce, no parece ser la vía de secreción RE-Golgi para llegar a la membrana, y el mecanismo.

Una hipótesis sobre las proteínas funcionales(4) es que pueden ser estables debido a que contribuyen a coordinar diferentes actividades celulares mediante estos mecanismos:

- Facilitando “cambios” o conexiones entre rutas metabólicas.
- Mecanismos de regulación (ej. Realimentación)

Es importante considerar que la actividad de redes metabólicas y de interacción, teniendo “puntos de control” clave se puede afectar de forma más efectiva. Por ejemplo, si se tienen dos rutas en que una proteína moonlighting actúa, y se causa un cambio conformacional que afecta una actividad, la otra se vería favorecida.

#### 2.2.4 Proteínas Moonlighting: Ejemplos principales

La proteína moonlighting con el mayor número de funciones identificadas es GAPDH, que tiene 18 funciones, contando todas las funciones que cumple entre los organismos en los que se encuentra(4). Es de las proteínas mejor caracterizadas, junto con enolasa, que es de gran interés por su participación en mecanismos de virulencia.

Las funciones de GAPDH se enumeran a continuación(1,9,11,56):

- Señal de stress y apoptosis.
- Transporte vesicular.
- Receptor de plasminógeno.
- Lente del cristalino.
- Unión de RNA y dianosina tetrafosfato.
- Motilidad de esperma.
- Receptor de lactoferrina en macrófagos.
- Función flagelar.
- DNA glicosilasa de uracilo (Reparación de DNA).
- Actividad de ribosilación de NAD.
- Proceso de endocitosis.
- Unión a microtúbulos.
- Transferencia de grupos fosfato.
- Proceso de apoptosis.
- Síntesis de polisacáridos extracelulares.
- Inhibición de la traducción.

La enolasa tiene las siguientes funciones: actúa como receptor de plasminógeno en bacterias, unirse a células epiteliales y facilitar la evasión del sistema inmune del hospederos(46,50,56).

La proteína supresora de tumores p53 es otra proteína moonlighting, cuyas funciones son como factor de transcripción en el núcleo y como promotor de permeabilidad de membrana en las mitocondrias. Como podríamos esperar, tiene muchas interacciones con proteínas reguladoras(3).

Entre las proteínas moonlighting que tienen muchas funciones descritas se encuentran: aconitasa, dihidropoliamida deshidrogenasa (DLD), Hsp70, Hsp90, HMGB1(3,57). Al estar involucradas en mecanismos de regulación de expresión, no es sorprendente que Hsp70, Hsp90, HMGB1 y p53 estén involucradas en procesos oncogénicos. También se han descrito proteínas moonlighting virales, y también están involucradas en la evasión del sistema inmune del hospedero. Por ejemplo, la proteína E5 del virus de papiloma humano(58).

Otro grupo interesante corresponde a proteínas moonlighting, originalmente intracelulares, que son secretadas sin tener secuencias (motifs) para esto(55). Estas proteínas llevan a cabo diferentes funciones que intervienen en varias rutas metabólicas, y procesos como el de inflamación(59). Por ejemplo: Factor de crecimiento de fibroblastos (bFGF), N-acetil-glucosamida 1 (NAG1), gen estimulado por interferón 15 (ISG15), Matriz de metalopeptidasa 12 (MMP12), Timosina  $\beta$ -4 (T $\beta$ 4)(35).

### 2.2.5 Proteínas Moonlighting: Aspectos evolutivos

Se ha descrito que las proteínas involucradas en estos procesos son propensas a tener funciones moonlighting, posiblemente debido a su ancestralidad(4). La ubicuidad de las proteínas moonlighting en enzimas del metabolismo primario (TCA y Glicólisis. Ver ilustraciones 4 y 5) parece ser indicativa de que esta condición aporta una ventaja en proteínas que tienen una "presión selectiva" muy alta, por lo que la emergencia de nuevas funciones parecería estar desfavorecida. Particularmente considerando que la tasa evolutiva de una proteína está predominantemente influenciada por su nivel de expresión en lugar de su importancia funcional(60).

Las proteínas multifuncionales pueden contribuir a coordinar diferentes actividades celulares mediante los siguientes mecanismos: facilitando el "cambio" o las conexiones entre diferentes vías metabólicas, proporcionando mecanismos reguladores como la retroalimentación y también debido a que corresponden a centros clave en las redes metabólicas y de interacción(61).

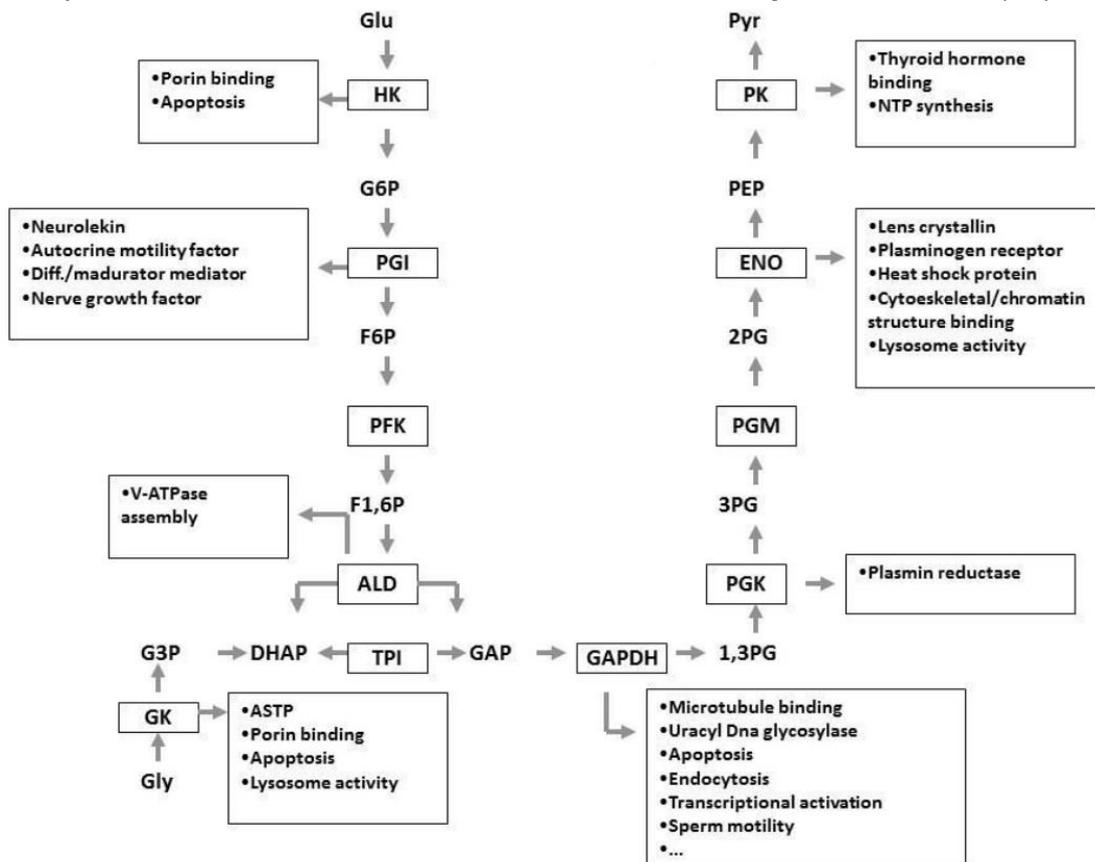


Ilustración 4. Proteínas moonlighting en la glicólisis.

(tomado de (4)). Se muestran las funciones adicionales, además de la función canónica en el metabolismo primario.

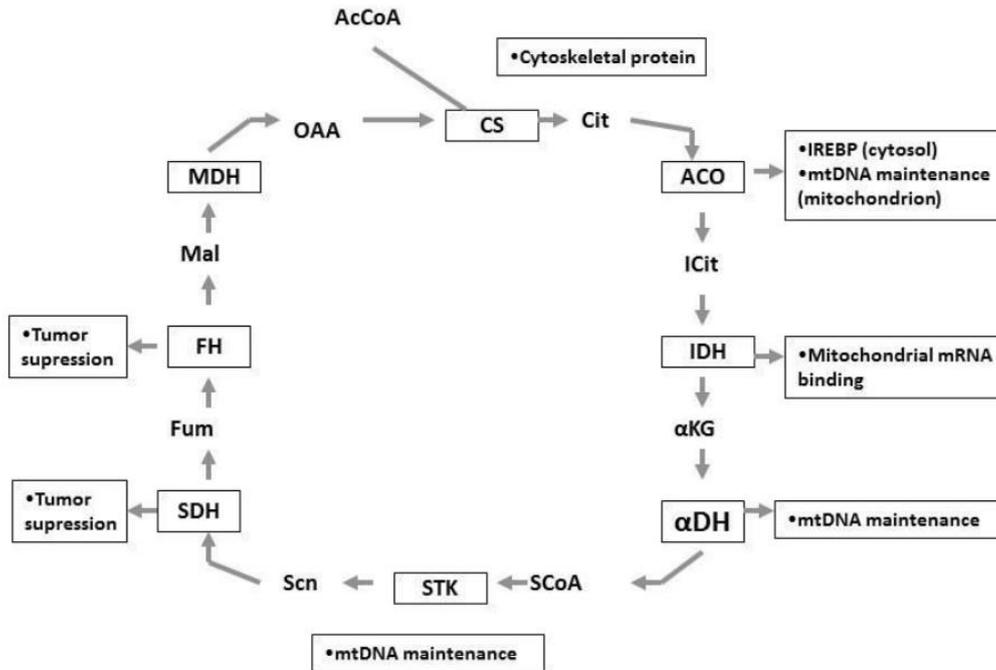


Ilustración 5. Proteínas moonlighting en el ciclo de los ácidos carboxílicos. (tomado de (4)). Se muestran las funciones adicionales, además de la función canónica en el metabolismo primario.

Parece ser que las secuencias y estructuras de proteínas, una vez “seleccionadas”, y dado que presentan condiciones de estabilidad y presencia en cierto ambiente celular, pueden utilizarse para otro fin sin tener que sufrir otro proceso de selección en tiempos evolutivos. Este proceso ocurriría en mucho menos tiempo, considerando que la adquisición de una segunda función puede deberse a cambios que no necesariamente involucran secuencia. Por ejemplo, un cambio de patrón en la expresión, como ocurre en GAPDH el cristalino del ojo humano(4,62). La función protéica también podría ser afectada por cualquiera de los procesos descritos en la ilustración 1.

Debido a la complejidad de la multifuncionalidad, su evolución es un fenómeno bastante desconocido(4). El desplazamiento de genes no ortólogos(63) y el reclutamiento enzimático(64) se han sugerido como posibles mecanismos para la emergencia de funciones moonlighting a partir de proteínas previas.

En el caso de GroEL, chaperona de *Enterobacter aerogenes*, se necesita un cambio de cuatro aminoácidos para que tenga actividad tóxica en insectos. La misma proteína en *E. coli* se diferencia de la de *E. aerógenes* en once aminoácidos, siendo solamente cuatro de estos clave para la segunda función (51).

Tomando en cuenta el caso de GroEL, un asunto interesante sería si se conserva la multifuncionalidad de proteínas entre especies. Está claro que identificar a una proteína como moonlighting no implica que esto ocurra en especies cercanas. En el caso de Piruvato carboxilasa de *S. cereviseae*, *H.*

*polymorpha* y *P. pastoris*, se observa una identidad de 80% entre las proteínas; sin embargo, la enzima de *S. cerevisiae* no es moonlighting, mientras que lo es en las otras dos especies(65). Por supuesto, la función moonlighting no necesariamente será la misma(63).

#### 2.2.6 Proteínas Moonlighting: Identificación y predicción

Existen una serie de indicios sobre la multifuncionalidad de una proteína que tienen mucho que ver con las características de estas:

- Si se encuentra una enzima en un compartimento celular inesperado, considerando su función descrita(3).
- Ubicación en la membrana de proteínas que no tengan señales para ser secretadas o exportadas a la membrana(14,35,43).
- Concentración más alta de lo esperado(3).
- Encontrar fenotipos inesperados en casos de delección(3).
- Interacción con partners inesperados al hacer análisis interactómico, descartando los casos de proteínas promiscuas o “pegostas” (muchas interacciones biológicamente irrelevantes) (38).

#### 2.2.7 Proteínas Moonlighting: Relación con virulencia

Ya se ha mencionado el papel de algunas proteínas del metabolismo primario de microorganismos patógenos (GAPDH y enolasa, principalmente) que tienen funciones de factores de virulencia, usualmente adhesión al hospedero (10,46,50).

También hay casos donde se fuerza una segunda función en proteínas del hospedero para forzarlas a una segunda función que facilita la actividad del patógeno. Por ejemplo, *E. coli* recluta actina del hospedero para facilitar la unión al epitelio intestinal del mismo(66). Esta situación, donde se fuerza una segunda función de una proteína del hospedero en detrimento del organismo que la sintetizó, es habitual en virus.

Distintas proteínas del metabolismo primario tienen la misma función moonlighting en relación con la virulencia. Esto implica que las enzimas que no tienen similitud de secuencia comparten alguna característica conformacional o motif no descrito(4).

Actualmente hay interés en la comunidad científica por identificar regiones de interacción con el plasminógeno humano u otras proteínas del hospedero. También se ha estudiado el proceso de invasión de tejidos para hacer metástasis.

En *Streptococcus* se han descrito algunos motifs ubicados en el extremo carboxiterminal: Lys-Lys-COOH y Lys-x-x-Lys-COOH. Nótese que x representa un aminoácido no polar. Estos motivos están involucrados en la unión al plasminógeno humano a través de interacción con los dominios de Kringle (32,33). Interesantemente, ninguno de estos motifs es exclusivo de microorganismos patógenos. Los motifs están presentes en muchos microorganismos no patógenos, que no se unen al plasminógeno. La mutación de los motivos reduce la actividad, pero no la elimina. Estos motifs no son

compartidos con otros patógenos o proteínas de virulencia (GAPDH, fosfoglucomutasa, fosfoglucoquinasa)(4).

Se han identificado aminoácidos cargados positivamente como sitios potenciales de unión a enolasa o a endopeptidasa neumocócica(4). Los aminoácidos involucrados son arginina e histidina están involucrados en la interacción proteína-proteína. A pesar de las diferencias en los motivos de unión al plasminógeno en distintas proteínas y microorganismos, la presencia de aminoácidos cargados positivamente en un entorno hidrofóbico parece ser determinante para la unión al plasminógeno del hospedero (32).

Nótese que estas proteínas moonlighting, que ordinariamente son citosólicas, son transportadas a la membrana sin señal de exportación. Como posibles explicaciones se han sugerido (4) mecanismos parecidos al complejo Sec de *Mycobacterium*, u otra ruta asistida por una chaperona, a la fecha no descrita, por la que se secretan proteínas como superóxido dismutasa (67,68). Es posible que las proteínas exportadas por SecA2 no necesiten el reconocimiento del péptido señal por parte del receptor.

## 2.2.8 Proteínas Moonlighting: Relación con enfermedades humanas

Las proteínas moonlighting también están relacionadas con procesos patológicos, generalmente debido a alguna mutación(61). Algunos casos son GAPDH (enfermedades neurodegenerativas), PGI (Anemia hemolítica, defectos neurológicos), XOR (Anomalías de lactancia), etc.(61). Hay otras proteínas moonlighting involucradas en procesos carcinógenos, como p53, TG2 y Hsp90 (3,5,50).

También hay muchos ejemplos donde la segunda función no se presenta en condiciones normales, y está relacionada con patologías, al ser la nueva función tóxica para el organismo. Jeffery lo ha llamado función neomórfica (5). Una proteína con una función neomórfica presenta la función moonlighting patológica solamente durante el proceso patológico.

Un análisis realizado con datos de MultitaskProtDB-II(26), OMIM(69), HGMD(70), TTD(71) y DrugBank(72) encontró que 78% de las proteínas moonlighting analizadas están involucradas en enfermedades humanas, un porcentaje mucho mayor al 17,9% que se observa en las proteínas humanas en general. Un análisis similar encontró que 48% de las proteínas analizadas son blancos terapéuticos, un porcentaje mucho mayor al 9,8% correspondiente al porcentaje considerando todas las proteínas humanas. Esto sugiere que la presencia de varias funciones es un fenómeno común en proteínas que causan enfermedades humanas, y que su estudio detallado puede explicar algunos efectos no deseados del uso de fármacos (9).

## 2.2 Métodos

### 2.2.1 Bases de datos, servidores y herramientas bioinformáticas utilizadas

Se utilizaron o evaluaron las siguientes bases de datos:

**Tabla 1. Bases de datos, servidores y herramientas bioinformáticas a evaluarse o utilizarse**

Nombre	Tipo	Referencia	Enlace
APID	Base de datos	(31)	<a href="http://cicblade.dep.usal.es:8080/APID/">http://cicblade.dep.usal.es:8080/APID/</a>
BioGRID	Base de datos	(29)(30)	<a href="https://thebiogrid.org/">https://thebiogrid.org/</a>
Drugbank	Base de datos	(72)(73)	<a href="https://go.drugbank.com/">https://go.drugbank.com/</a>
Gene Ontology	Base de datos	(74)(75)	<a href="http://geneontology.org/">http://geneontology.org/</a>
HGMD	Base de datos	(70)(76)	<a href="http://www.hgmd.cf.ac.uk/ac/index.php">http://www.hgmd.cf.ac.uk/ac/index.php</a>
IMEx	Base de datos	(77)	<a href="https://www.imexconsortium.org/">https://www.imexconsortium.org/</a>
InterPro	Base de datos	(78)(79)	<a href="https://www.ebi.ac.uk/interpro/">https://www.ebi.ac.uk/interpro/</a>
MultitaskProtDBII	Base de datos	(26)	<a href="http://wallace.uab.es/multitaskII/">http://wallace.uab.es/multitaskII/</a>
OMIM	Base de datos	(69)	<a href="https://omim.org/">https://omim.org/</a>
PDB	Base de datos	(80)	<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>
PiSITE	Base de datos	(20)	<a href="http://pisite.hgc.jp/cgi-bin/top.cgi">http://pisite.hgc.jp/cgi-bin/top.cgi</a>
TTD	Base de datos	(71)(81)	<a href="http://db.idrblab.net/ttd/">http://db.idrblab.net/ttd/</a>
STRING	Base de datos	(27)	<a href="https://string-db.org/">https://string-db.org/</a>
GOnet	Herramienta	(82)	<a href="https://tools.dice-database.org/GOnet/">https://tools.dice-database.org/GOnet/</a>
GOTrapper	Aplicación	(83)	<a href="https://github.com/BioGeneTools/GOTrapper">https://github.com/BioGeneTools/GOTrapper</a>
ClusPro 2.0	Servidor Docking PPI	(84)	<a href="https://cluspro.bu.edu/">https://cluspro.bu.edu/</a>

### 2.2.2 Selección de un listado proteínas moonlighting asociadas a enfermedades humanas

Para cumplir con los objetivos, establecidos en 1.2, se tomó una estrategia que consiste, grosso modo, en seleccionar un conjunto de proteínas multifuncionales, tomadas de la base de datos de proteínas multifuncionales MultitaskProtDB-II(26), y estudiar la interacción de proteínas de interés involucradas en enfermedades humanas con otras proteínas metabólicas de interés.

Al evaluarse los recursos para determinar la interacción proteína-proteína (PPI) se determinó que el más adecuado era STRING(27), basado en Cytoscape (28). STRING cuenta con 932,553,897 interacciones(85), por lo que supera las otras opciones para hacer los análisis.

El análisis de interacción local entre proteínas asociadas encontradas con el mapado se realizó de la siguiente forma: Se efectuó un análisis manual del listado de proteínas a las que se unen las proteínas moonlighting, según BioGRID(29)(30) para examinar las principales proteínas interactoras que presentan, y si comparten algunas de ellas.

### 2.2.3 Odds Ratio (Razón de Probabilidades)

Este tipo de test, también llamado “razón de probabilidades”, es una estadística que cuantifica la fuerza de la asociación entre dos eventos, A y B (86,87). El procedimiento seguido involucra definir una matriz de confusión, mostrada a continuación, y obtener la información asociada de la misma.

	Test	Proteína Moonlighting	Proteína No Moonlighting
Grupo Asociada a Enfermedad		127	44
Grupo No Asociada a Enfermedad		5732	70045

Los resultados se pueden verificar en este [enlace](#) de la universidad de Hong Kong.

## 2.2.4 Extracción de códigos GO

Una vez que se obtuvieron las proteínas humanas asociadas a enfermedades, se incluyeron en GOTrapper(83) (ver ilustración 6), y se utilizó la opción de “within” para encontrar términos asociados en dentro de una sola lista de genes. También se utilizó threshold = 4, para obtener los códigos GO más generales. Nótese que la salida del archivo colocaba los códigos GO como únicos y unía los genes que tenían el código GO específico.

id	GO_id	GO_category	#_Background_Genes	Genes	Gene ID	Score	GO_term	
0	1	GO:0042803	MF	906	BST2, CDH1, CRYAB, TYMP, ENO1, HLCS, HES1, HSP...	684, 999, 1410, 1890, 2023, 3141, 3280, 3320, ...	8.8234	protein homodimerization activity
1	2	GO:0043312	BP	490	ALAD, ANPEP, ANXA2, ARG1, BST2, B4GALT1, GPI, ...	210, 290, 302, 383, 684, 2683, 2821, 2934, 314...	7.9366	neutrophil degranulation
2	3	GO:0005524	MF	1512	ACTB, ADK, ASS1, CFTR, EGFR, ERCC2, GK, GSK3B,...	60, 132, 445, 1080, 1956, 2068, 2710, 2932, 30...	9.5622	ATP binding
3	4	GO:0051301	BP	684	RCC1, ERCC2, FGF1, FGF2, PLK1, PPP1CC, RAN, RP...	1104, 2068, 2246, 2247, 5347, 5501, 5901, 6188...	8.4179	cell division
4	5	GO:0019901	MF	727	ACTB, CDKN1A, ATF2, CTNNB1, DNM2, EGFR, GSK3B,...	60, 1026, 1386, 1499, 1785, 1956, 2932, 3320, ...	8.5058	protein kinase binding

Ilustración 6. Encabezado de salida de GOTrapper con threshold = 4.

Posteriormente se utilizó un script en Python para generar un diccionario que le asignaba los códigos GO correspondientes a cada gen (ver Ilustración 7):

```
GO_DIC = {}
for search_gene in tabla_1["Gene Symbol"]:
    temporary_list = []
    for i in range(len(GO_tool["Genes"])):
        if search_gene in GO_tool["Genes"][i]:
            temporary_list.append(GO_tool["GO_id"][i])
    GO_DIC[search_gene] = temporary_list
```

Ilustración 7. Código utilizado para asignar los códigos GO a cada gen

Luego las funciones biológicas se anotaron manualmente consultando QuickGO(88). Esto quedó registrado en un archivo adjunto en el anexo.

## 2.2.5 Análisis de funciones de proteínas moonlighting asociadas a enfermedades:

Se asignaron categorías a cada enfermedad manualmente, consultando OMIM(69) y MedGen(89). Las categorías utilizadas fueron: "anemia", "atherosclerosis", "autoimmune", "blood", "cancer", "circulatory", "dermatitis", "genetic", "Hypertension", "immune", "Keloid", "metabolic", "myopathy", "necrosis", "neoplasia", "nervous", "bone", "retinal degradation", "virulence". Se decidió no tomar estos grupos para examinarlos precisamente, por razones de tiempo. Sin embargo, los genes de cada grupo quedan registrados en un archivo adjunto en el anexo.

También se identificaron en MultitaskProtDBII los genes humanos involucrados con enfermedades, se colocaron los 127 genes en string,

tomando en cuenta textmining, databases e interacciones experimentales, y se identificaron 7 grupos de interacción ([enlace](#)). Luego los 7 grupos se examinaron con GONet(82), tomando la opción de “GO Term annotation”, sin los términos asociados a los genes, para interpretar más claramente la función de cada grupo.

Es interesante mencionar que GONet también puede utilizarse para análisis de enriquecimiento, en caso de que se desee revisar los niveles de expresión de los genes en cierto tipo específico de tejido.

#### 2.2.6 Redes de interacción

El análisis se realizó de la siguiente manera:

1. Se identificaron en MultitaskProtDBII los genes humanos involucrados con enfermedades,
2. Se colocaron los 127 genes en string, tomando en cuenta textmining, databases e interacciones experimentales,
3. Se identificaron 7 grupos de interacción ([enlace](#)),
4. Se descargaron de BioGRID los datos de las proteínas de interés de cada grupo,
5. Se convirtieron en .csv,
6. Se incorporaron a RStudio,
7. Se tomaron las columnas 8 y 9 del .csv, que contienen los genes que interactúan con la proteína de interés. Nótese que las columnas 8 y 9 corresponden al Gene Symbol de los interactores.
8. Las columnas 8 y 9 del .csv se colocaron en un vector de interacciones,
9. Se seleccionaron los valores únicos de ese vector de interacciones,
10. Finalmente, se exportaron que un archivo tipo .csv,
11. Se utilizó STRING para generar las redes de interacción. Las redes se crearon mostrando la red completa, solamente con interacción experimental, y con el nivel más alto de confianza.

**Ilustración 8. Código utilizado para generar el archivo que se colocó en STRING para generar la red del grupo enf\_7.**

#### 2.2.7 Docking con ClusPro 2.0

Luego de que se identificaron las proteínas asociadas a enfermedades humanas en MultitaskProtDBII, se escogió el grupo 4\_enf (ATF2, MDM2, TP53) para realizar un docking con ClusPro 2.0.

El docking se realizó sin hacer una identificación previa de motifs. Solamente se utilizó el código de identificación de estructura tridimensional de PDB(80) para cada uno de los genes.

Se utilizaron los siguientes códigos para cada proteína:

<b>Gen</b>	<b>Código PDB</b>
ATF2:	3d06 (Cadena A)
MDM2:	2lzg (Cadena A)
TP53:	1t2k (Cadena D)

Las interacciones modeladas fueron de TP53 con ATF2 (trabajo número 580168) y TP53 con MDM2 (trabajo número 580167). Los datos de cada interacción están como archivos adjuntos.



**Tabla 2. Grupos de interacción entre proteínas moonlighting humanas. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de BioGrid**

G1		G2		G4		G8		G15	
HMGB1	312	NUP50	130	EPRS	286	CRYAB	127	NACA2	21
PIAS1	129	NUP153	149	QARS	295	HSPB4	70	CALR	235
TP53	1,379	RANBP2	215	AIMP1	147	G9		G16	
MDM2*	601	RANGAP1	200	LARS	245	HK1	168	CTNNB1	730
RPL11	370	RAN	229	KARS	224	Hk2	80	CDH1	659
EEF1A1	510	RCC1	137	G5		G10		G17	
RPS3	406	SMAD4	266	HSPE1	206	NUBP1	59	AP2M1	345
RPS7**	267	SMAD1	168	HSPD1	719	NUBP2	92	RALBP1	148
RPS13	321	SMAD7	103	HSP90AB1	527	G11			
RPS14	337	SMAD3	380	HSP90AA1	995	MAPK1	364		
RPL5	320	SMAD2	323	G6		EGFR	1,360		
RPL7	333	SMAD6	51	NME1	151	G12			
RPL10	458	G3		NQO1	65	FGF1	39		
RPL13A	220	GPX4	52	ADK	26	FGF2	39		
RPL22	296	ENO2	110	G7		G13			
RPL23	359	ENO1	433	PPP1CC	665	NUP98	140		
RPL26	179	TPI1	188	YWHAZ	625	NUP62	220		
RPLP0	291	GAPDH	331	CFTR	859	G14			
		PGK1	182			TMSB4X	40		
		TPPP	22			ACTB	497		

\* Q00987 se cambió a A7UKY0 (en uniprot\_id\_1, en R). Ambas corresponden a MDM2. La primera está mejor descrita, pero la segunda está en MultitaskProtDBII.

\*\* P62081 en la base de datos está como "40S ribosomal protein S7", el nombre utilizado es RPS7.

En el grupo 3 Destacan GAPDH, y 2 Enolasas. Interesantemente, las interacciones reportadas incluyen interacción entre ENO1 y GAPDH. Estas interacciones se calculan tomando referencias de ensayos bioquímicos como cross linking, inmunoprecipitación y purificación por afinidad en tándem, pero no se encontraron reportes que mencionaran involucramiento de la membrana plasmática (ver ilustración 10).

**Single Interaction Record**

protein-protein interaction link out: [GRID](#)

*Homo sapiens*: ● ENO1 ● GAPDH

Detected by psi-mi:"MI:0401"(biochemical) assay

*Comment*: BioGRID interaction database, imported October 2018. Large-scale datasets were scored individually, all other interactions globally. This interaction has been curated based on the following abstract:

PMID:22939629: A census of human soluble protein complexes.  
Havugimana PC, Hart GT, Neupus T, Yang H, Tarinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Beznin A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM, Emili A  
Cell. 150(5):1068-81 2012.

**Abstract**:  
Cellular processes often depend on stable physical associations between proteins. Despite recent progress, knowledge of the composition of human protein complexes remains limited. To close this gap, we applied an integrative global proteomic profiling approach, based on chromatographic separation of cultured human cell extracts into more than one thousand biochemical fractions that were subsequently analyzed by quantitative tandem mass spectrometry, to systematically identify a network of 13,993 high-confidence physical interactions among 3,006 stably associated soluble human proteins. Most of the 622 putative protein complexes we report are linked to core biological processes and encompass both candidate disease genes and unannotated proteins to inform on mechanism. Strikingly, whereas larger multiprotein assemblies tend to be more extensively annotated and evolutionarily conserved, human protein complexes with five or fewer subunits are far more likely to be functionally unannotated or restricted to vertebrates, suggesting more recent functional innovations.

[BACK](#)

**Ilustración 10. Ejemplo de interacción entre ENO1 y GAPDH en la base de datos STRING.**

### 2.3.1.2 Interacción entre proteínas moonlighting humanas asociadas a enfermedades

Posteriormente se han seleccionado de la base de datos utilizada(26) proteínas humanas con enfermedades descritas, obteniéndose 127 proteínas.

Se hizo un análisis preliminar de las cantidades de funciones por proteína, mediante un test F, para confirmar si las varianzas son diferentes. Luego un two sample t test indica que las cantidades de funciones de proteínas moonlighting no son mayores en *Homo sapiens* que en todo el conjunto. Estos datos están en el archivo de R utilizado para análisis de datos, adjunto como anexo.

```
{r}
var.test(fcn_go_count, fcnh_go_count, alternative="two.sided")
t.test(fcn_go_count, fcnh_go_count, var.equal = F)
```

```
F test to compare two variances

data: fcn_go_count and fcnh_go_count
F = 0.34379, num df = 2775, denom df = 683, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3047263 0.3862018
sample estimates:
ratio of variances
 0.3437924

Welch Two Sample t-test

data: fcn_go_count and fcnh_go_count
t = -5.5646, df = 802.2, p-value = 3.582e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.942732 -2.843410
sample estimates:
mean of x mean of y
 5.911023 10.304094
```

Ilustración 11. Test F y two sample t test de Welch.

*fcn\_go\_count* se refiere al conteo de funciones GO en cada proteína moonlighting (16409), y *fcn\_h\_go\_count* se refiere al conteo de funciones GO en cada proteína moonlighting humana (7048).

### 2.3.2 Odds ratio

Ya se ha descrito anteriormente que hay una gran proporción de proteínas moonlighting relacionadas con enfermedades(9). En este trabajo se realizó un odds ratio con los datos más recientes. Se observó que 74% de las proteínas moonlighting humanas están relacionadas con enfermedades.

Interesantemente, las cifras difieren de la bibliografía consultada debido a que se han reportado nuevas proteínas moonlighting, y se han incluido muchas proteínas nuevas tanto a OMIM como a UniProt. Los cálculos fueron realizados con epiR(90) y a mano(87,91).

**Tabla 3. Resultados de análisis estadístico odds ratio, que muestra el grado de significación, con un intervalo de confianza de 95%.**

	n	proporción
Número de proteínas moonlighting involucradas en enfermedades	127	74.269
Número de proteínas moonlighting No involucradas en enfermedades	44	25.731
Total	171	100.000
Número de proteínas humanas en UniProt y OMIM	5732	7.564
Número de proteínas humanas en UniProt y no OMIM	70045	92.436
Número de proteínas humanas en UniProt*	75777	100.000

Estimadores de riesgo

Estimadores de Riesgo	Estimador	IC 95%	
		(Inferior,	Superior)
Riesgo absoluto en el grupo tratamiento	0.74	0.65	0.83
Riesgo absoluto en el grupo control	0.08	-0.02	0.17
Reducción absoluta de riesgo (ARR)	-0.67	-0.76	-0.58
Riesgo Relativo (RR)	9.82	8.96	10.76
Reducción del Riesgo Relativo (RRR)	-8.82	-9.76	-7.95
<b>Odds ratio</b>			
Probabilidades en el grupo tratamiento	2.11	1.73	2.48
Probabilidades en el grupo control	0.15	-0.12	0.42
Odds ratio	35.27	25.01	49.75

Esto muestra que las proteínas moonlighting son más propensas a estar involucradas en enfermedades humanas que el grupo general de proteínas.

### 2.3.3 Extracción de códigos GO de proteínas moonlighting

Una vez obtenido el listado, se observó que había una cantidad importante de códigos GO(74,75) asociados a la función canónica y a la principal de cada proteína moonlighting asociada a una enfermedad. Debido a que la cantidad de códigos GO asociados a cada proteína era bastante elevada (ver extracto de la tabla a continuación, ilustración 12), se utilizó GOTrapper(83) para filtrar automáticamente los GO, con un threshold de 4, para obtener los más generales de la jerarquía, descartando los específicos. Esto que disminuyó la cantidad de códigos GO asociados a cada gen (ver anexos) de 1100 a 359. Sin embargo, hubo 5 genes que no fueron reconocidos: NACA2, XOR, SMN1 e IDH. En estos casos se colocaron todos los códigos GO de la base de datos de proteínas moonlighting. Como ejemplo: ANPEP, disminuyó de 25 códigos GO a 6 (ilustración 12).

Debido a que la organización del resultado colocaba varios genes en cada código GO. Se decidió extraerlos de forma automática, para lo que se diseñó un script en Python (ver anexos).

Uniprot	Gene Symbol	Disease	GO_ID	Biological Process	GO_ID	Molecular function
P15144	ANPEP	[1] "Urinary Bladder Neoplasms, Dermatitis, Allergic Contact"\n"	GO:0046718	viral entry into host cell	GO:0008233	peptidase activity
			GO:0043312	neutrophil degranulation	GO:0038023	signaling receptor activity
					GO:0008270	zinc ion binding
P18754	RCC1	[2] "Raynaud disease"\n"	GO:0007088	regulation of mitotic nuclear division	GO:0042277	peptide binding
			GO:0007052	mitotic spindle organization	GO:0046982	protein heterodimerization activity
			GO:0051301	cell division	GO:0051020	GTPase binding
			GO:0051262	protein tetramerization	GO:0003682	chromatin binding
			GO:0000082	G1/S transition of mitotic cell cycle		
P00558	PGK1	[3] "Phosphoglycerate kinase deficiency, hepatocellular carcinoma, Renal cell carcinoma"\n"	GO:0061621	canonical glycolysis	GO:0016651	oxidoreductase activity, acting on NAD(P)H
			GO:0001666	response to hypoxia	GO:0005524	ATP binding
			GO:0043467	regulation of generation of precursor metabolites and energy		
			GO:0016525	negative regulation of angiogenesis		
			GO:0006094	gluconeogenesis		
			GO:0062013	positive regulation of small molecule metabolic process		
			GO:0007088	regulation of mitotic nuclear division		
A1A528	ZW10	[4] "Osteosarcoma"\n"	GO:0007093	mitotic cell cycle checkpoint		
			GO:2001251	negative regulation of chromosome organization		
			GO:0010256	endomembrane system organization		
			GO:0033045	regulation of sister chromatid segregation		
			GO:1901991	negative regulation of mitotic cell cycle phase transition		
			GO:0051301	cell division		
			GO:0051656	establishment of organelle localization		
			GO:0000070	mitotic sister chromatid segregation		
			GO:0030010	establishment of cell polarity		
			GO:0071840	cellular component organization or biogenesis	GO:0005624	nucleus
GO:0022613	ribonucleoprotein complex biogenesis	GO:0005732	cytoplasm			
GO:0044085	cellular component biogenesis					
GO:0008150	biological process					
GO:0042254	ribosome biogenesis					
						Note: This gene had no output on GOTrappor The strikethrough genes are GO cellular components

Ilustración 12. Ejemplo de los cinco primeros genes en la tabla.

Nótese que NACA2 tiene un color diferente porque fue uno de los 5 genes para los que se colocaron todos los GO. Los GO tachados corresponden a la categoría de compartimiento celular, que no se encontró en ningún otro caso.

El procedimiento empleado fue el siguiente: Se importó la tabla donde se iban a incorporar los GO, se importó la tabla resultante de aplicar GOTrappor con threshold de 4, y se realizó un diccionario que contenía el Gene Symbol de cada proteína moonlighting con sus funciones GO, que después se agregaron a la tabla con los GO.

### 2.3.4 Análisis de funcionalidad de proteínas moonlighting

Posteriormente se realizó un análisis de interacción utilizando STRING, para determinar las interacciones con las 127 proteínas moonlighting relacionadas con enfermedades humanas. En este caso el análisis se realizó con el nivel más alto de confianza (0.900), y tomando en cuenta sólo bases de datos (amarillo), datos minados de texto (azul) y experimentales (morado).

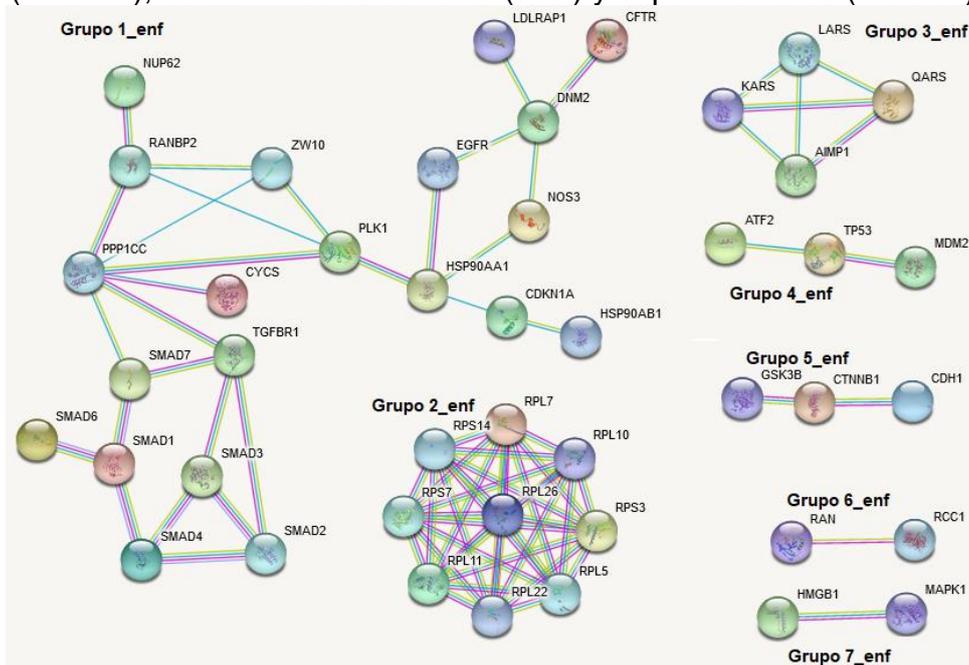


Ilustración 13. Red de interacción de proteínas moonlighting humanas ligadas a enfermedades. 123 proteínas ([enlace](#))

**Tabla 4. Grupos de interacción entre proteínas moonlighting humanas relacionadas a enfermedades. La columna de la izquierda contiene la notación abreviada de la proteína. La de la derecha contiene el número de interactores a la fecha, de acuerdo con una revisión manual de BioGrid.**

G1_enf		G2_enf		G4_enf	
CDKN1A* <sup>!</sup>	335	RPL5	320	ATF2* <sup>!!!</sup>	227
CFTR	859	RPL7	333	MDM2	601
CYCS*	86	RPL10	458	TP53	1,379
DNM2*	173	RPL11	370		
EGFR	1,360	RPL22	296	G5_enf	
HSP90AA1	995	RPL26	179	CDH1	659
HSP90AB1	527	RPS3	406	CTNNB1	730
LDLRAP1*	30	RPS7	267	GSK3B*	577
NOS3* <sup>!!</sup>	88	RPS14	337		
NUP62	220			G6_enf	
PLK1	374	G3_enf		RAN	229
PPP1CC	665	AIMP1	147	RCC1	137
RANBP2	215	KARS	224		
SMAD1	168	LARS	245	G7_enf	
SMAD2	323	QARS	295	HMGB1	312
SMAD3	380			MAPK1	364
SMAD4	266				
SMAD6	51				
SMAD7	103				
TGFBR1*	211				
ZW10*	73				

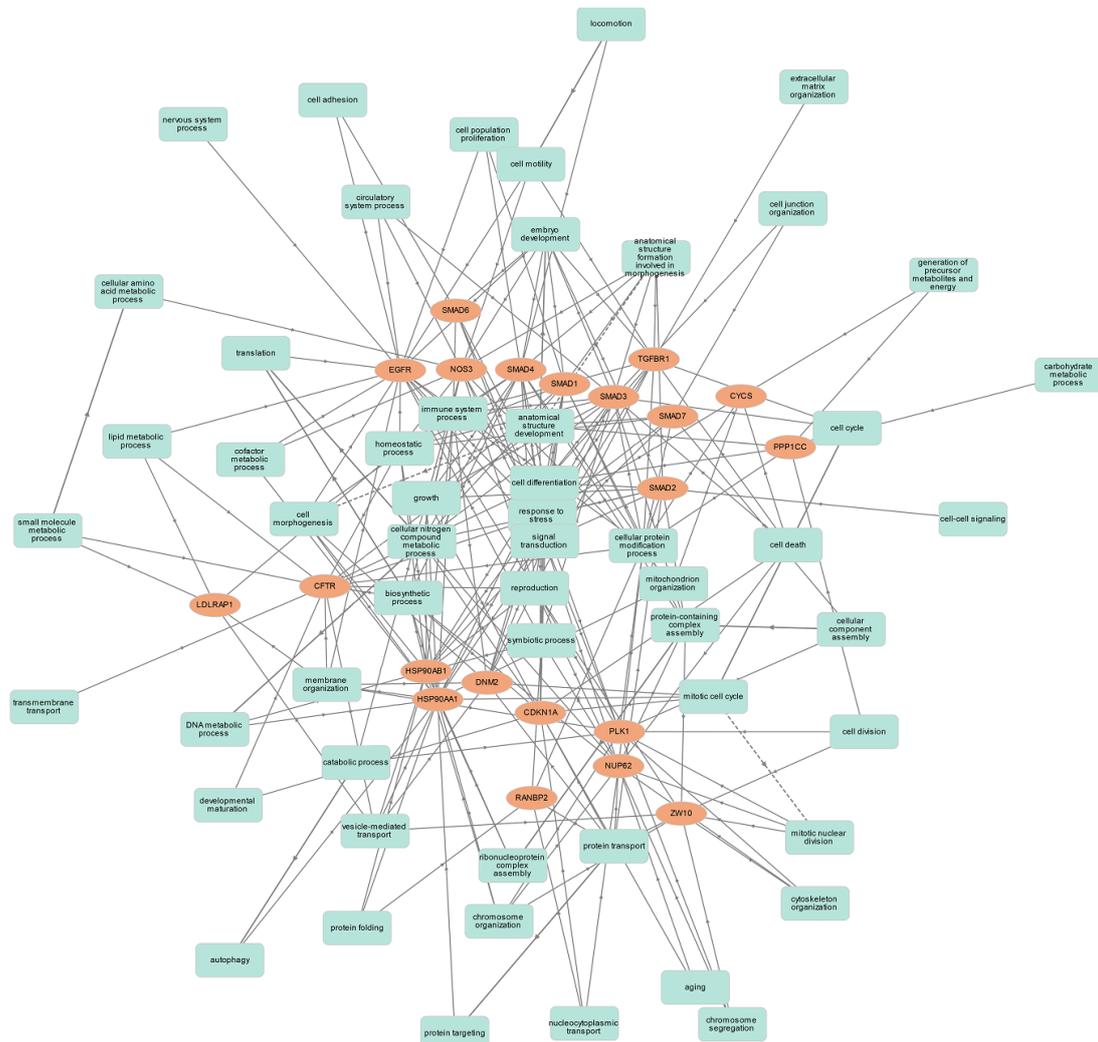
\* No aparecen en la red de interacción entre proteínas moonlighting humanas.

<sup>!</sup> Proteína sinónima con PAK3. <sup>!!</sup> Proteína sinónima con NANOS3. <sup>!!!</sup> Proteína sinónima con GDNF.

Posteriormente, se analizaron las funciones GO de cada grupo en GOnet(82). Los códigos GO pueden corresponder a procesos biológicos, funciones moleculares o compartimientos celulares, y todos dan cierta información. Para mayor claridad en las imágenes, se muestran sin términos no conectados a los genes.

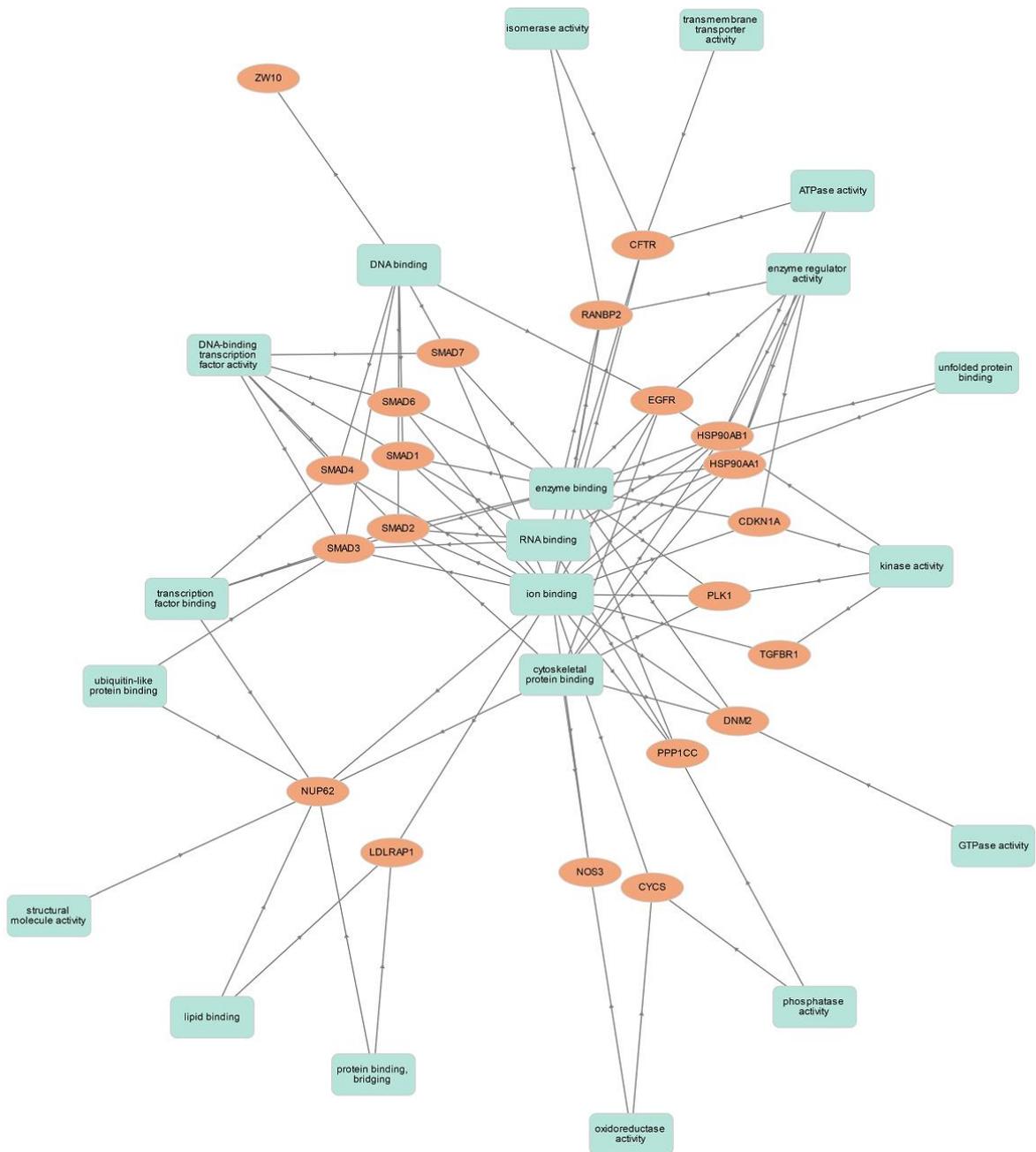
A continuación, se muestra el análisis para el primer grupo obtenido en el análisis de asociación (ilustración 14). GOnet permite la visualización de la jerarquía de términos, así como las relaciones del gen con cada término GO, por lo que se escogió esta opción sobre otras para el análisis. El modo de uso fue análisis de anotación de términos GO (mapeado de entradas proporcionadas a subconjuntos GO).

En la ilustración a continuación se observan los genes parte del grupo 1\_enf. Los genes (figuras ovaladas, color albaricoque) se conectan con los procesos biológicos con los que están involucrados (figuras cuadradas, color verde claro). Por ejemplo: LDLRAP1 (centro izquierda) está involucrado con transporte mediado por vesículas, organización de membrana, morfogénesis celular, procesos metabólicos de lípidos y procesos metabólicos de moléculas pequeñas.



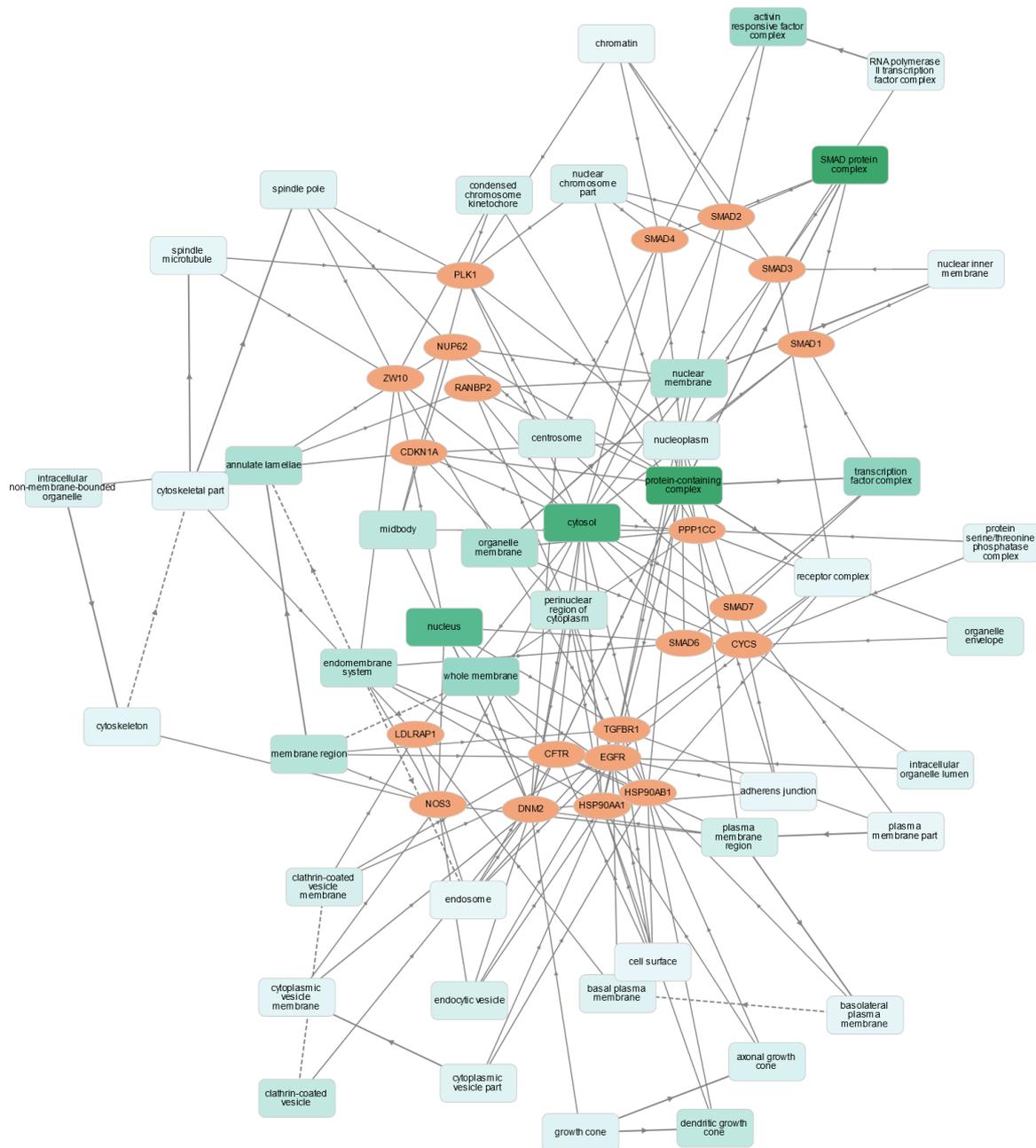
**Ilustración 14. Procesos biológicos asociados a los genes del Grupo 1\_inf. ([enlace](#))**

Los procesos más comunes en cada gen se observan cerca del centro (Ilustración 14): Procesos del sistema inmune, procesos homeostáticos, crecimiento, procesos biosintéticos, procesos metabólicos de compuestos de nitrógeno celular, desarrollo de estructuras anatómicas, diferenciación celular, respuesta al estrés, transducción de señal, reproducción, procesos simbióticos, procesos de modificación celular de proteínas, organización de la mitocondria, ensamblado de complejos que contienen proteína. Otros procesos tienen menos conexiones, como locomoción (locomotion, arriba, centro) y envejecimiento (aging, abajo, centro), que se relacionan con solo dos genes. De igual manera se pueden revisar las funciones moleculares (ilustración 15), o las localizaciones celulares (ilustración 16).



**Ilustración 15. Mecanismos moleculares asociados a los genes del Grupo 1\_inf. ([enlace](#))**

En esta ilustración, los procesos más comunes son de unión: a enzimas, a RNA, a iones y a proteínas citoesqueléticas. También es interesante notar que el grupo de SMAD (izquierda, arriba) muestra actividad de factor de transcripción de unión a DNA, y unión a DNA, junto con ZW10. Las otras actividades enzimáticas tienen que ver con metabolismo energético: oxidorreductasa, fosfatasa, GTPasa, cinasas (derecha). La mayoría de los mecanismos restantes también son de unión a DNA, a factores de transcripción, a proteínas similares a ubiquitina, y a lípidos. El resto de las funciones son: actividad de molécula estructural, actividad isomerasa y transporte transmembrana.



**Ilustración 16. Compartimentos celulares asociados a los genes del Grupo 1\_enf. ([enlace](#))**

En esta ilustración se observa dónde están ubicadas las proteínas expresadas por los genes del grupo 1\_enf. Mientras más oscuro es el color verde, menor el p valor: así podemos estar seguros que las proteínas SMAD se asocian a su complejo; y que la mayoría de las proteínas están en el citosol; NOS3, ZW10 y SMAD6 están en núcleo. También se observan muchas interacciones en zonas como nucleoplasma, membrana nuclear, centrosoma, membranas de orgánulos y nuclear.

La consideración de las tres imágenes anteriores nos permite describir los productos genéticos de los genes involucrados en el Grupo\_1enf en detalle.

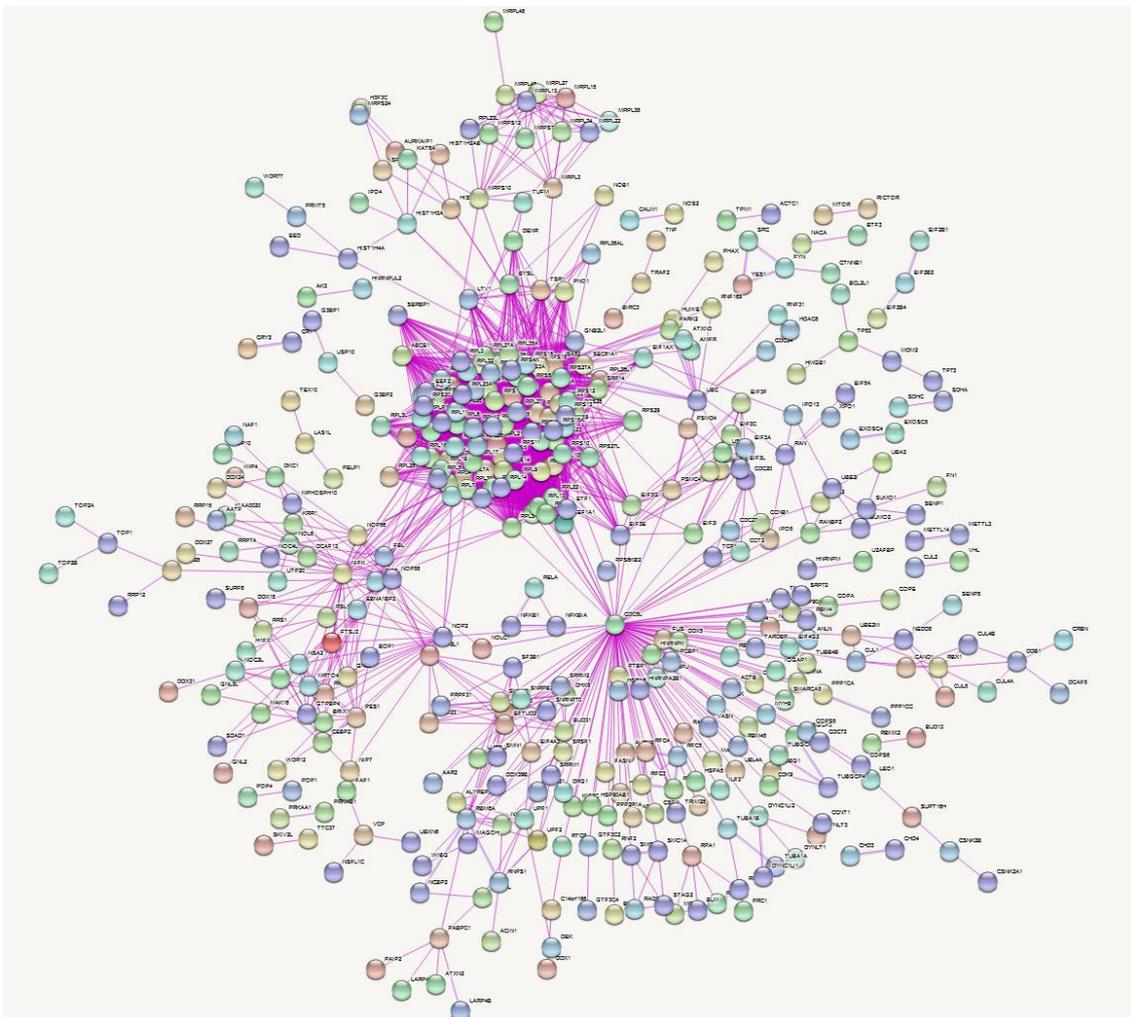
### 2.3.5 Redes de interacción

Luego de seguir el procedimiento indicado para los 7 grupos indicados en la ilustración 13 y la tabla 4, se llegó a las siguientes redes de interacción, mostradas a continuación en la tabla 5:

**Tabla 5. Redes de interacciones de proteínas para los grupos de proteínas moonlighting humanas asociadas a enfermedades**

Grupo	Enlace	Interacciones	Genes
G1_enf	*	-	4612
G2_enf	<a href="#">enlace</a>	1029	1063
G3_enf	<a href="#">enlace</a>	478	551
G4_enf	<a href="#">enlace</a>	1735	1790
G5_enf	<a href="#">enlace</a>	1612	1612
G6_enf	<a href="#">enlace</a>	297	318
G7_enf	<a href="#">enlace</a>	624	636

\* Para el momento de la redacción del trabajo, STRING soportaba máximo 2000 interacciones, lo que impidió la realización de la red de interacciones para el grupo 1.



**Ilustración 17. Red de interacción de proteínas del grupo 2\_enf. 1029 ítems. ([enlace](#))**

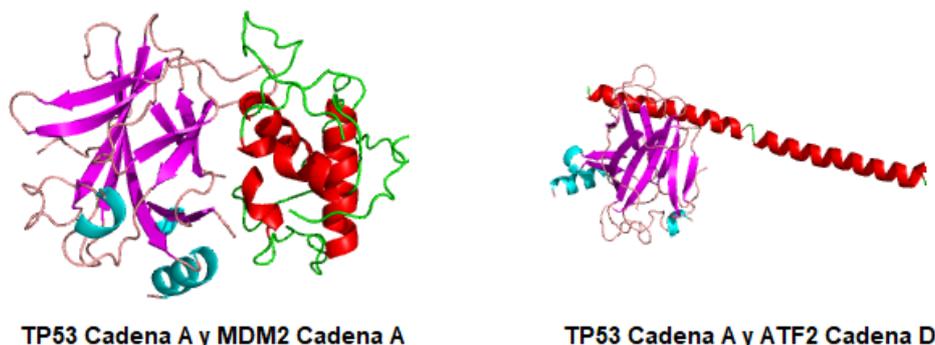
Este grupo correspondía a proteínas ribosómicas de tipo L y S, que corresponden al grupo principal (centro, arriba-medio) de la ilustración. Se pueden notar algunos nodos que concentran muchas interacciones que comunican la red de proteínas ribosómicas con otros elementos: destaca CDC5L (Cell division cycle 5 like protein) en el centro de la imagen, que une al

grupo con el resto de interactores de la parte inferior. Arriba al centro, MRPL3 (Mithochondrial Ribosomal Protein L3) comunica al resto de las proteínas ribosomales mitocondriales con el grupo principal. Las otras interacciones ocurren por cúmulos de varios interactores, como el formado por FBL (Fibrilarina) y dos proteínas nucleolares (56 y 58), en la parte inferior izquierda, y cuatro ubiquitina-ligasas (AMFR, ATXN3, PARK2, HUWE1), que junto con UBC unen al grupo principal con el resto de las proteínas ubicadas en la parte inferior.

Es importante notar que los procesos biológicos en los que está involucrados más componente de este grupo son biogénesis ribosomal, traducción, targeting de proteínas, procesos simbióticos y procesos catabólicos que contienen nucleobases. Los compartimientos celulares en los que más componentes se encuentran involucrados son: núcleo, nucleoplasma, ribosoma, nucleolo, y citosol. Las funciones moleculares más relevantes son unión a RNA mensajero y ribosomal, y como constituyente estructural del ribosoma y unión a enzimas (RPL11, RPL5, RPS7 y RPS3).

### 2.3.6 Docking con ClusPro 2.0

ClusPro 2.0(84) genera varios modelos para cada pareja de receptor y ligando, según la formación de sitios de unión o interacción con baja energía. Las interacciones modeladas fueron de TP53 (Morado y azul) con ATF2 (Rojo y Verde) (trabajo número 580168) y TP53 (Morado y azul) con MDM2 (Rojo y Verde) (trabajo número 580167). El trabajo de TP53 y ATF2 dio 22 modelos de docking, mientras que el de TP53 y MDM2 dio 30 resultados. Los datos de cada interacción están como archivos adjuntos. Abajo se observan los 2 modelos que muestran la interacción con menor energía (Ilustración 18).



**Ilustración 18. Ejemplos de Docking de TP53 Cadena A con MDM2 Cadena A, y ATF2 Cadena D modelado con ClusPro 2.0 en un solvente balanceado.**

La interacción de TP53 y de MDM2 se ha documentado anteriormente en 11 ocasiones en Homo sapiens (PMIDs: [24207125](#), [21857681](#), [19255450](#), [19234109](#), [18316739](#), [18172499](#), [17470788](#), [17438265](#), [17080083](#), [16432196](#), [926727](#)) y en 2 rutas curadas. Se puede verificar haciendo click en la línea morada entre TP53 y MDM2 en el siguiente [enlace](#), correspondiente a la Ilustración 13. En la misma consulta de STRING mencionada anteriormente no se encontraron interacciones de TP53 y ATF2 en humanos, aunque hay numerosos reportajes en otras especies.

Una revisión manual de la interacción de TP53 y ATF2 da 8 resultados (92)(93)(94)(95)(96)(97)(98)(99). Algunas fuentes no fueron relevantes(92)(98), pero otras aportan ciertas ideas sobre los mecanismos de acción en que se involucra la interacción de estas dos proteínas: Implicación en "Angiosarcoma cutáneo"(92); activación de RAS-GTPasa(93); proteínas que interactúan con DNA o mRNA(95); genes de "resistencia a los fármacos" en el carcinoma hepatocelular (96); papel de ATF2 en el aumento independiente de TP53 en la expresión de KIN (proteína de unión a RNA y DNA) después de la irradiación gamma(97); activación de p38 MAPK y el aumento de las actividades de unión al DNA de ATF-2 y p53 median la sobreexpresión de ciclooxigenasa-2 en la senescencia prematura en fibroblastos pulmonares IMR-90(99).

## 2.4 DISCUSIÓN

Como se mencionó anteriormente, la proteómica cambió de un enfoque en moléculas específicas a una perspectiva integrada, donde las proteínas son actores dinámicos con varias funciones(1). Esto trae varios retos: obtener información utilizable sobre la gran cantidad de secuencias e interrelaciones accesibles actualmente, y manejar otros factores que limitan la usabilidad de la información disponible (resumidos en la ilustración 1).

### 2.4.1 Interacción entre proteínas moonlighting asociadas a enfermedades humanas

Al hacer la revisión de literatura, se encontró que las proteínas moonlighting de *E. coli* tienen más interacciones físicas con un número mayor de clases funcionales distintas (38). Además, la misma fuente (38) menciona que tienden a interactuar entre sí, por lo tanto se decidió hacer una revisión de la interacción entre proteínas moonlighting humanas.

Aparentemente, hay una tendencia en las proteínas moonlighting de humanos a interactuar entre sí (Ilustración 9, Tabla 2). Una posible explicación yace en que la inversión metabólica para producir una proteína es alta, lo que daría una ventaja selectiva a especies que puedan utilizar una enzima para más de una función.

Un caso bien conocido de proteínas moonlighting humanas es el siguiente: alrededor de una docena de enzimas ubicuas fueron adoptadas para jugar un segundo papel en las células especializadas de la lente del ojo (17). Además, hallazgos recientes(17) que indican que factores celulares que afectan la función de proteínas moonlighting soportan esta hipótesis de “economía metabólica”.

Es decir, la reutilización de una estructura preexistente, con plegamiento, estructura y estabilidad seleccionados evolutivamente, puede cambiar de función con alguno de los cambios indicados en la ilustración 1 (unión a compuestos pequeños: sustratos, productos, cofactores u otros ligandos; otras proteínas, o complejos multiprotéicos; unión a RNA o DNA, modificaciones post-traduccionales; oligomerización) para permitir la selección de una proteína que cumpla más de una función.

La posibilidad de modular distintas funciones con cambios que requieren poca energía es una ventaja adaptativa considerable, lo que favorecería a un organismo capaz de utilizar proteínas multifuncionales. Otra posible ventaja de la multifuncionalidad, además de la reutilización de secuencias previamente seleccionadas, tiene que ver con un requerimiento de menor cantidad de genes (i. e. el organismo podría ejecutar más funciones con la misma cantidad de genes que otro que no presente multifuncionalidad de proteínas).

Otro elemento a considerar es que la tasa evolutiva de una proteína está dictada por su nivel de expresión (60), lo que deja a proteínas expresadas

constitutivamente en una posición favorable en términos de ser seleccionada como proteína multifuncional.

#### 2.4.2 Razón de probabilidades

No se observaron suficientes evidencias para afirmar que el número de clases funcionales de proteínas moonlighting en humanos es distinto al del resto de los organismos (Ilustración 11).

Dos eventos son independientes si y solo si el odds ratio es igual a 1, es decir, las probabilidades de un evento son las mismas en presencia o ausencia del otro evento. Si el odds ratio es mayor que 1, entonces A (la proteína es moonlighting) y B (la proteína está vinculada a una enfermedad) están asociados (correlacionados) en el sentido de que, en comparación con la ausencia de B, la presencia de B aumenta las probabilidades de A, y simétricamente la presencia de A aumenta las probabilidades de B. Por el contrario, si el OR es menor que 1, entonces A y B están correlacionados negativamente y la presencia de un evento reduce las probabilidades del otro evento (86,87).

Es posible que, una alteración que impida las funciones de una proteína multifuncional perturbe varios sistemas, y tenga consecuencias más serias que la pérdida de función de una proteína con una sola función. Aunado esto a que la mayoría de las proteínas moonlighting son parte de procesos constitutivos (lo que resalta su importancia fisiológica), una perturbación en este tipo de proceso puede ser más seria que una perturbación que involucre una sola función.

Observando la tabla 3, se puede notar que el riesgo absoluto en el grupo tratamiento (0.74) y control (0.08) es menor que 1, lo que implica que es significativo. Asimismo, se puede notar que la reducción absoluta de riesgo no contiene 1 en el intervalo de confianza, lo que implica que el valor es significativo. El riesgo relativo (9.82) corresponde al número de veces que ambos grupos difieren. La reducción del riesgo relativo (-8.82) corresponde a la cantidad de veces que una proteína normal está relacionada con una enfermedad, comparada con las proteínas moonlighting.

Los resultados obtenidos sugieren que las proteínas moonlighting son propensas a estar involucradas en enfermedades humanas (Tabla 3). Esto coincide con la sugerencia de que el fenómeno de multifuncionalidad no es un fenómeno extraño, y que su estudio detallado puede explicar mecanismos moleculares de enfermedades y efectos colaterales de algunos medicamentos(9).

#### 2.4.3 Análisis de códigos GO, funcionalidad de proteínas moonlighting y redes de interacción

Los códigos GO obtenidos a través de la herramienta GONet(82) se utilizaron para analizar las anotaciones de los distintos grupos examinados (grupo 1\_enf: Ilustraciones 14, 15, 16. Resto de grupos: [enlace](#)).

También se examinaron las proteínas multifuncionales presentes tanto en la red de proteínas multifuncionales humanas tenían ciertas características en común, que se presentan en la tabla a continuación:

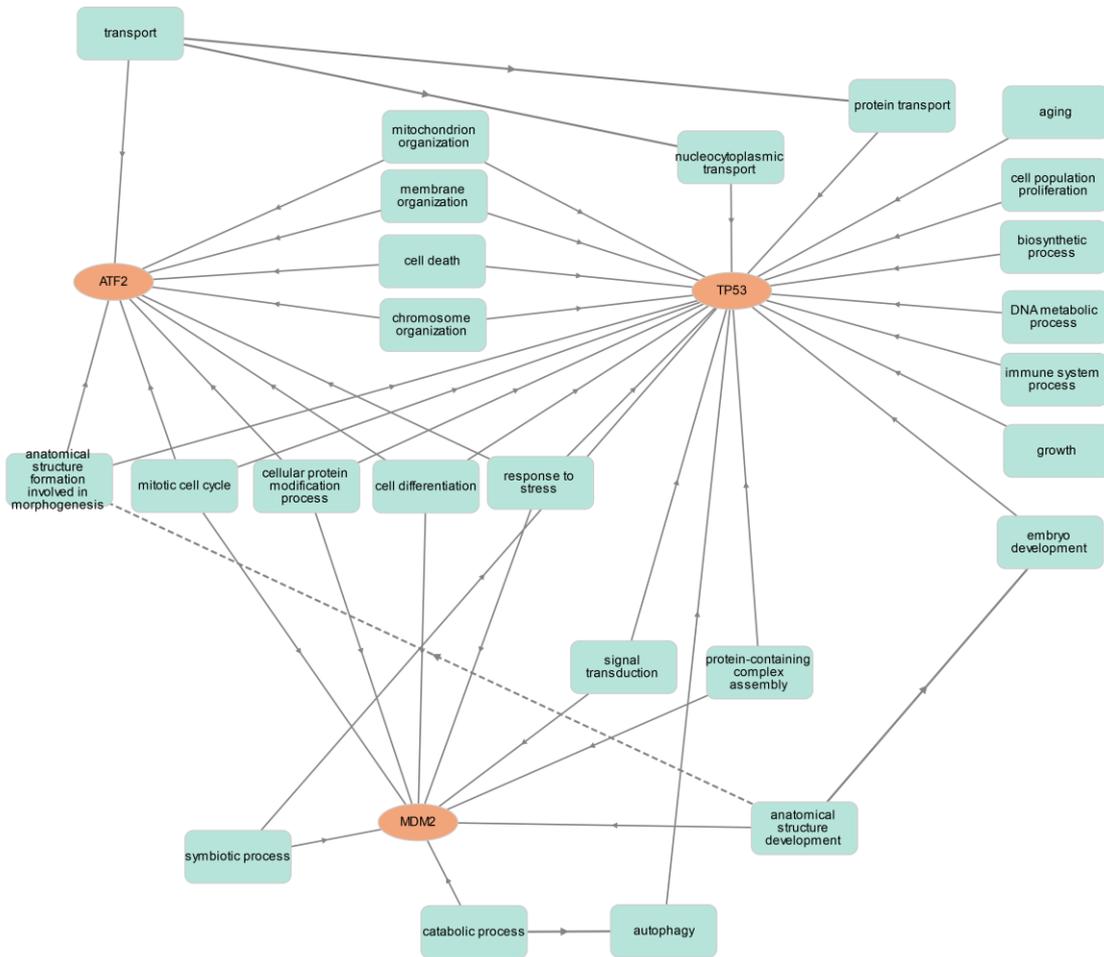
**Tabla 6. Características comunes entre proteínas multifuncionales humanas y proteínas multifuncionales humanas relacionadas a enfermedades**

Grupo	Proteínas en común	Procesos biológicos compartidos	Funciones moleculares compartidas	Compartimientos celulares compartidos
1_enf	CFTR, EGFR, HSP90AA1, HSP90AB1, NUP62, NUP98, PPP1CC, RANBP2, SMAD1, SMAD2, SMAD3, SMAD4, SMAD6, SMAD7	-	-	Complejos que contienen proteínas (GO:0032991)
2_enf	RPL10, RPL11, RPL22, RPL26, RPL5, SRPL7, RPS14, RPS3, RPS7	Procesos catabólicos compuestos que contienen nucleobases (GO:0034655) Procesos simbióticos (GO:0044403) Targeting de proteínas (GO:0006605)	Constituyente estructural del ribosoma (GO:0003735)	Citosol (GO:0005829) Ribosoma (GO:0005840)
3_enf	AIMP1, KARS, LARS, QARS	Traducción (GO:0006412) Proceso metabólico de aminoácidos celulares (GO:0006520) Proceso metabólico del tRNA (GO:0006399)	-	Citosol (GO:0005829) Complejos que contienen proteínas (GO:0032991)
4_enf	MDM2, TP53	Proceso de modificación de proteínas celulares (GO:0006464) Ciclo celular mitótico (GO:000278) Autofagia (GO:0006914) Diferenciación celular (GO:0030154) Ensamblamiento de complejos que contienen proteínas (GO:0065003) Procesos simbióticos (GO:0044403) Respuesta al estrés (GO:0006950)	Unión a enzimas (GO:0019899) Unión a iones (GO:0043167)	Citosol (GO:0005829) Complejos que contienen proteínas (GO:0032991) Nucleolo (GO:0005730) Nucleoplasma (GO:0005654)
5_enf	CDH1, CTNBN1	Adhesión celular (GO:0007155) morfogénesis celular (GO:000902) Diferenciación celular (GO:0030154) Organización de uniones celulares (GO:0034330) Procesos simbióticos (GO:0044403)	-	Membrana plasmática (GO:0005886) Complejos que contienen proteínas (GO:0032991) Espacio extracelular (GO:0005615)
6_enf	RAN, RCC1	Segregación de cromosomas (GO:0007059) Organización del citoesqueleto (GO:0007010) Procesos simbióticos (GO:0044403) División celular (GO:0051301)	Unión a iones (GO:0043167)	Complejos que contienen proteínas (GO:0032991) Envoltura nuclear (GO:0005635) Nucleoplasma (GO:0005654)
7_enf	HMGB1, MAPK1	Diferenciación celular (GO:0030154) Muerte celular (GO:0008219) Procesos homeostáticos (GO:0042592) Procesos del sistema inmune (GO:0002376) Respuesta al estrés (GO:0006950) Transducción de señal (GO:0007165) Transporte mediado por vesículas (GO:0016192)	Unión a enzimas (GO:0019899) Unión a iones (GO:0043167) Unión a factores de transcripción (GO:0008134) Unión a DNA (GO:0003677)	Complejos que contienen proteínas (GO:0032991) Endosoma (GO:0005768) Membrana plasmática (GO:0005886) Nucleoplasma (GO:0005654)

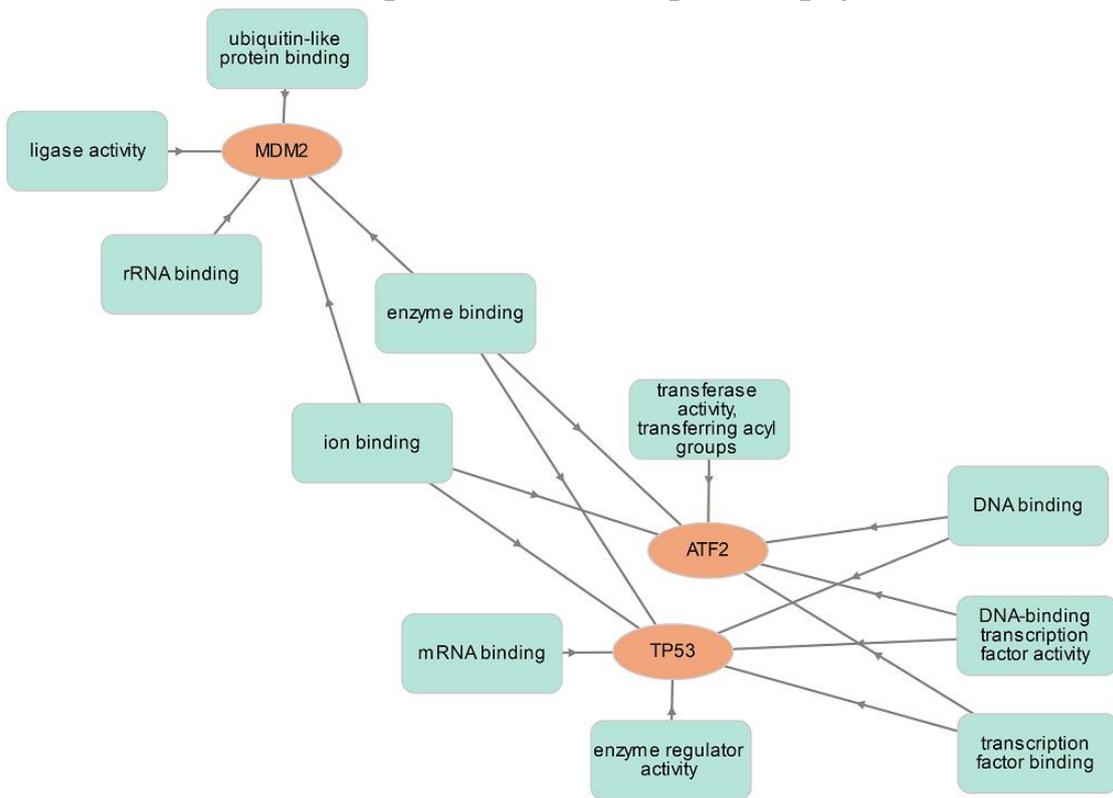
- representa que no hubo elementos GO comunes para ese grupo, en esos casos se coloca un enlace a una ilustración de los códigos.  
[Enlace](#) a todas las ilustraciones.

Es importante notar que este cuadro solamente marca los elementos comunes de los genes comunes entre los dos grupos mencionados, buscando facilitar la comprensión de los procesos comunes entre las proteínas moonlighting humanas que interactúan entre sí, y las proteínas moonlighting humanas relacionadas con enfermedades.

Podemos ilustrar la limitación del cuadro considerando el grupo 4\_enf (ATF2, MDM2, TP53): En ilustración 19, se puede notar que existen 4 procesos biológicos comunes a los tres genes del grupo: ciclo celular mitótico, procesos de modificación de proteínas celulares, diferenciación celular, respuesta al estrés. Además, hay muchos procesos biológicos en los que están involucrados dos de los tres genes. El mismo examen puede hacerse para las

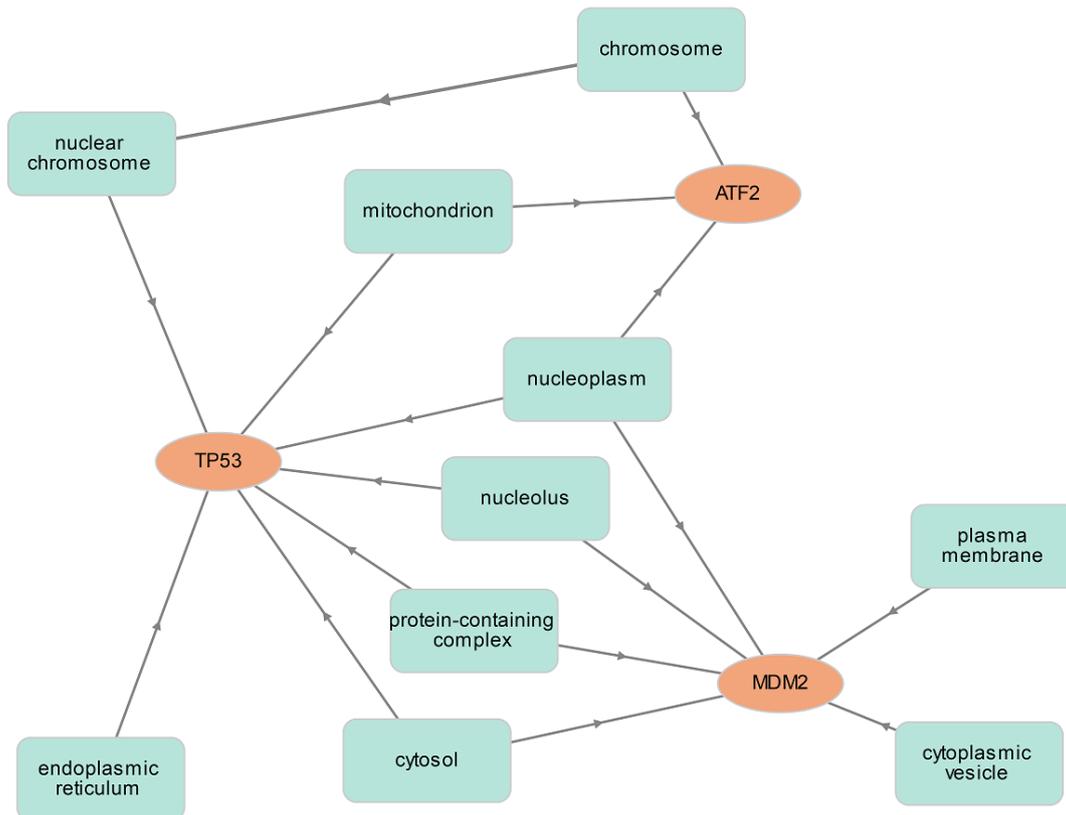


**Ilustración 19. Procesos biológicos asociados a los genes del grupo G4\_enf**



**Ilustración 20. Mecanismos moleculares asociados a los genes del grupo G4\_enf**

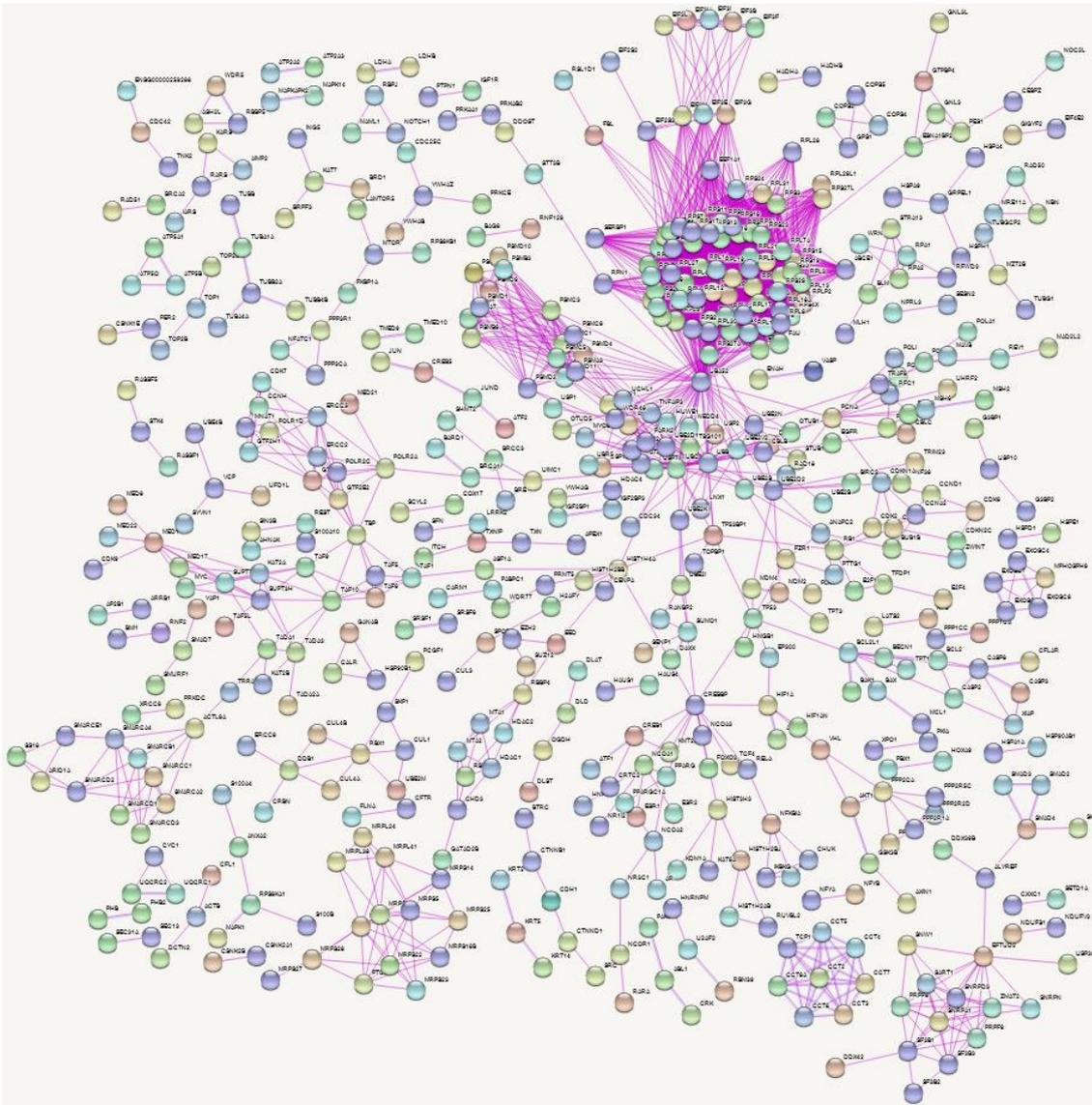
funciones moleculares (Ilustración 20, donde se observan solamente dos funciones comunes: unión a iones y a enzimas) o compartimientos celulares (Ilustración 21, donde se observa que sólo comparten localización en el nucleoplasma).



**Ilustración 21. Compartimientos celulares asociados a los genes del grupo G4\_enf**

Finalmente, se puede revisar la red de proteínas con la que interactúan las proteínas del grupo 4\_enf (Ilustración 22). Se puede notar que esta red de interacción también llega a las proteínas estructurales ribosómicas del grupo 2\_enf (Centro, hacia arriba), se ven las proteínas del grupo 5\_enf (abajo, centro), y algunas SMAD del grupo 1\_enf (abajo, derecha). Interesantemente, MAPK1 del grupo 7\_enf (abajo, izquierda), aparece separada de HGMB1, que a su vez aparece interactuando con TP53 (centro, hacia la derecha). No se observan proteínas de grupo 3\_enf, ni del grupo 6\_enf.

A pesar de que se observan separaciones en la red de interacción, debe tomarse en cuenta que sólo se muestran las conexiones con score de interacción por encima de 0.9. Esta decisión se tomó para disminuir la cantidad de nodos que se mostraban en la figura y hacerla más comprensible. A pesar de eso, sin tener una familiaridad con las proteínas que aparecen en la red, difícilmente puede hacerse una interpretación.



**Ilustración 22. Red de interacción de proteínas del grupo 4\_enf. 1735 ítems. ([enlace](#))**

Uno de los temas por completar sería hacer "enrichment analysis" de los genes de interés en un caso específico de enfermedad, utilizando los datos de GONet o datos propios generados de un experimento. También observar perturbaciones en el interactoma de proteínas durante procesos de enfermedad.

Si se mejora la forma de generar los grupos de enfermedades para ser más específicos se pueden utilizar las redes de interacción para generar "enrichment analysis" de tejidos específicos y observar perfiles de expresión para la red de interés.

#### 2.4.4 Análisis del Docking

En este caso no se contaba con motifs específicos de interacción, y se utilizó ClusPro 2.0 como "prueba de concepto" para encontrar interacciones documentadas con anterioridad (MDM2-TP53 y ATF2-TP53). La parte valiosa del uso de este recurso fue el desarrollo de familiaridad y aprendizaje involucrados.

Uno de los temas por completar que sería de interés en cuanto a la interacción entre proteínas moonlighting es realizar el análisis de todos los modelos propuestos por ClusPro 2.0 para pares de proteínas de interés e identificar los aminoácidos de interacción, con la finalidad de revisar si coinciden con reportes de variantes asociadas a enfermedades, como la sección de Pathology/Biotech de UniProt.

#### 2.4.7 Observaciones finales

Tomando en cuenta los resultados observados (Ilustraciones 9, 13, 17 y Tabla 5) se pueden ver esbozos de una amplia red de interacciones, que empieza con proteínas moonlighting humanas, y de estas algunas se relacionan con enfermedades. Así, las proteínas moonlighting parecen parte de una gran red de interacciones de procesos fisiológicos de gran importancia, que suelen ser constitutivos.

La gran cantidad de interacciones que muestran las proteínas moonlighting humanas sugiere que son “encrucijadas” metabólicas, donde confluyen muchos procesos fisiológicos.

La anotación de un repertorio funcional de proteínas humanas y de otras especies ha impactado en los campos de la bioquímica y la genética, y posiblemente seguirá influyendo en preguntas científicas básicas y aplicadas, desde el problema del genotipo al fenotipo hasta nuestra comprensión de las patologías humanas y el diseño de fármacos.

### 3. Conclusiones

- Las proteínas multifuncionales humanas tienen redes metabólicas con muchas interacciones entre sí, incluyendo las proteínas multifuncionales relacionadas con enfermedades humanas.
- Las proteínas multifuncionales son más propensas a estar involucradas en enfermedades humanas que otras proteínas.
- La cantidad de información disponible en redes de interacción de proteínas y de códigos GO requiere una estrategia elaborada para hacer la información comprensible. Sin embargo, es de mucha utilidad para la caracterización de grupos de proteínas.
- La interacción proteína-proteína juega un papel clave en la predicción de funciones protéicas de interés, así como en la capacidad farmacológica de las moléculas.
- Los modelos de docking molecular contribuyen a la comprensión de los detalles de interacción de proteínas y facilitan la formulación hipótesis de mecanismos moleculares.
- Se recomienda que otros trabajos inviertan tiempo en elaborar "enrichment analysis" de tejidos específicos y observar perfiles de expresión para la red de interés.
- Para profundizar en el análisis de docking se puede hacer una búsqueda bibliográfica de motivos específicos de interés para alguna enfermedad, o utilizar la información de Pathology/Biotech de UniProt.
- Las proteínas multifuncionales siguen siendo un tema de gran interés para la determinación de mecanismos moleculares de procesos fisiológicos, y la sofisticación de exámenes de redes de interacción de proteínas sólo puede favorecer el estudio de este tema.

## 4. Glosario

GO	Ontología de genes (Geneontology.org)
PPI	Interacción proteína-proteína
csv	Valores separados por coma
RNA	Ácido Ribonucleico
DNA	Ácido Deoxirribonucleico
gut	Gene Unconnected Terms (proviene de GONet)
MultitaskProtDB-II	Base de datos de proteínas multitarea, versión II (Multitask Protein Database, version II)
STRING	Herramienta de búsqueda para la recuperación de genes / proteínas que interactúan (Search Tool for the Retrieval of Interacting Genes/Proteins)
TP53	Tumor protein P53
ATF2	Activating Transcription Factor 2
MDM2	Mouse double minute 2 homolog (proto-oncogén)
Ure2	Transcriptional regulator URE2
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Lys	Lisina
OMIM	Online Mendelian Inheritance in Man
HGMD	Human Gene Mutation Database
TTD	Therapeutic Target Database

## 5. Bibliografía

1. Espinosa-Cantú A, Cruz-Bonilla E, Noda-Garcia L, DeLuna A. Multiple Forms of Multifunctional Proteins in Health and Disease. *Front Cell Dev Biol.* 2020;8(June):1–10.
2. Jeffery CJ. Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS J.* 2020;1–9.
3. Jeffery CJ. What is Protein Moonlighting and Why is it Important? Moonlighting Proteins Nov Virulence Factors *Bact Infect.* 2016;1–19.
4. Franco-Serrano L. Proteínas “moonlighting”: Identificación y relación por microorganismos patógenos y clínica humana [Internet]. Universitat Autònoma de Barcelona; 2018. Available from: <https://ddd.uab.cat/record/203802>
5. Jeffery CJ. Proteins with neomorphic moonlighting functions in disease. *IUBMB Life.* 2011;63(7):489–94.
6. Gupta MN, Pandey S, Ehtesham NZ, Hasnain SE. Medical Implications of Protein Moonlighting. *Indian J Med Res* [Internet]. 2019;322–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23144490>
7. Jeffery CJ. Intracellular proteins moonlighting as bacterial adhesion factors. *AIMS Microbiol.* 2018;4(2):362–76.
8. Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci.* 1999;24(1):8–11.
9. Franco-Serrano L, Huerta M, Hernández S, Cedano J, Perez-Pons JA, Piñol J, et al. Multifunctional Proteins: Involvement in Human Diseases and Targets of Current Drugs. *Protein J* [Internet]. 2018;37(5):444–53. Available from: <http://dx.doi.org/10.1007/s10930-018-9790-x>
10. Henderson B, Fares MA, Martin ACR. Protein Moonlighting in Biology and Medicine. *Protein Moonlighting in Biology and Medicine.* 2016. 1–317 p.
11. Unal CM, Steinert M. Microbial Peptidyl-Prolyl cis/trans Isomerases (PPIases): Virulence Factors and Potential Alternative Drug Targets. *Microbiol Mol Biol Rev.* 2014;78(3):544–71.
12. Rasch J, Theuerkorn M, Ünal C, Heinsohn N, Tran S, Fischer G, et al. Novel cycloheximide derivatives targeting the moonlighting protein Mip exhibit specific antimicrobial activity against *Legionella pneumophila*. *Front Bioeng Biotechnol.* 2015;3(MAR):1–8.
13. Guleria A, Chandna S. ATM kinase: Much more than a DNA damage responsive protein. *DNA Repair (Amst)* [Internet]. 2016;39:1–20. Available from: <http://dx.doi.org/10.1016/j.dnarep.2015.12.009>
14. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I, et al. Bioinformatics and moonlighting proteins. *Front Bioeng Biotechnol.* 2015;3(JUN).
15. Lievens S, Eyckerman S, Lemmens I, Tavernier J. Large-scale protein interactome mapping: Strategies and opportunities. *Expert Rev Proteomics.* 2010;7(5):679–90.
16. Copley SD. Moonlighting is mainstream: Paradigm adjustment required. *BioEssays.* 2012;34(7):578–88.
17. Jeffery CJ. An enzyme in the test tube, and a transcription factor in the cell: Moonlighting proteins and cellular factors that affect their behavior. *Protein Sci.* 2019;28(7):1233–8.
18. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al.

- Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
19. Gómez A, Hernández S, Amela I, Piñol J, Cedano J, Querol E. Do protein-protein interaction databases identify moonlighting proteins? *Mol Biosyst.* 2011;7(8):2379–82.
  20. Higurashi M, Ishida T, Kinoshita K. PiSite: A database of protein interaction sites using multiple binding states in the PDB. *Nucleic Acids Res.* 2009;37(SUPPL. 1):360–4.
  21. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc [Internet].* 2015;10(6):845–58. Available from: <http://dx.doi.org/10.1038/nprot.2015-053>
  22. Colell EA, Iserle JA, Simonetti FL, Marino-Buslje C. MISTIC2: Comprehensive server to study coevolution in protein families. *Nucleic Acids Res.* 2018;46(W1):W323–8.
  23. Caldera M, Buphamalai P, Müller F, Menche J. Interactome-based approaches to human disease. *Curr Opin Syst Biol [Internet].* 2017;3:88–94. Available from: <http://dx.doi.org/10.1016/j.coisb.2017.04.015>
  24. Gonzalez MW, Kann MG. Chapter 4: Protein Interactions and Disease. *PLoS Comput Biol.* 2012;8(12).
  25. Kuzmanov U, Emili A. Protein-protein interaction networks: Probing disease mechanisms using model systems. *Genome Med.* 2013;5(4):1–12.
  26. Franco-Serrano L, Hernández S, Calvo A, Severi MA, Ferragut G, Pérez-Pons JA, et al. MultitaskProtDB-II: An update of a database of multitasking/moonlighting proteins. *Nucleic Acids Res.* 2018;46(D1):D645–8.
  27. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res.* 2019;18(2):623–32.
  28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res [Internet].* 2003 Nov;13(11):2498–504. Available from: <http://ci.nii.ac.jp/naid/110001910481/>
  29. Breitkreutz B-J, Stark C, Tyers M. The GRID: The General Repository for Interaction Datasets. *Genome Biol [Internet].* 2002;3(12):PREPRINT0013. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12537553>
  30. Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D529–41.
  31. Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, et al. APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database.* 2019;2019(i):1–8.
  32. Fulde M, Steinert M, Bergmann S. Interaction of streptococcal plasminogen binding proteins with the host fibrinolytic system. *Front Cell Infect Microbiol [Internet].* 2013;3. Available from: <http://journal.frontiersin.org/article/10.3389/fcimb.2013.00085/abstract>
  33. Derbise A, Song YP, Parikh S, Fischetti VA, Pancholi V. Role of the C-

- Terminal Lysine Residues of Streptococcal Surface Enolase in Glu- and Lys-Plasminogen-Binding Activities of Group A Streptococci. *Infect Immun* [Internet]. 2004 Jan;72(1):94–105. Available from: <https://iai.asm.org/content/72/1/94>
34. Jeffery CJ. Moonlighting proteins: Complications and implications for proteomics research. *Drug Discov Today TARGETS*. 2004;3(2):71–8.
  35. Yoon JH, Ryu J, Baek SJ. Moonlighting activity of secreted inflammation-regulatory proteins. *Yonsei Med J*. 2018;59(4):463–9.
  36. Chapple CE, Brun C. Redefining protein moonlighting. *Oncotarget* [Internet]. 2015 Jul 10;6(19):16812–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26220711>
  37. Jeffery CJ. Why study moonlighting proteins? *Front Genet*. 2015;6(JUN).
  38. 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(1).
  39. Hernández S, Ferragut G, Amela I, Perez-Pons J, Piñol J, Mozo-Villarias A, et al. MultitaskProtDB: A database of multitasking proteins. *Nucleic Acids Res*. 2014;42(D1):517–20.
  40. Piatigorsky J. *Gene Sharing and Evolution: The Diversity of Protein Functions*. Harvard University Press; 2007. 337 p.
  41. Beadle GW, Tatum EL. Genetic Control of Biochemical Reactions in *Neurospora*. *Proc Natl Acad Sci* [Internet]. 1941 Nov 15;27(11):499–506. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.27.11.499>
  42. McKay S, Ashley B. *Gene Sharing and Evolution: The Diversity of Protein Functions*. By Joram Piatigorsky. *Q Rev Biol* [Internet]. 2008 Mar;83(1):95–109. Available from: <https://doi.org/10.1086/586921>
  43. Amblee V, Jeffery CJ. Physical features of intracellular proteins that moonlight on the cell surface. *PLoS One*. 2015;10(6):1–16.
  44. Piatigorsky J, Wistow GJ. Enzyme/crystallins: Gene sharing as an evolutionary strategy. *Cell*. 1989;57(2):197–9.
  45. Chapple CE, Brun C. Redefining protein moonlighting. *Oncotarget*. 2015 Jul;6(19):16812–3.
  46. Henderson B, Martin A. Bacterial Moonlighting Proteins and Bacterial Virulence. In: Dobrindt U, Hacker JH, Svanborg C, editors. *Between Pathogenicity and Commensalism* [Internet]. 1st ed. Springer; 2013. p. 155–213. Available from: <https://link.springer.com/book/10.1007/978-3-642-36560-7#about>
  47. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009;27(2):157–67.
  48. Wong SKF. G protein selectivity is regulated by multiple intracellular regions of GPCRs. *NeuroSignals*. 2003;12(1):1–12.
  49. Tompa P, Szász C, Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci* [Internet]. 2005 Sep;30(9):484–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0968000405002185>
  50. 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.
  51. Santosh Kumar CM, Mande SC. Protein chaperones and non-protein substrates: On substrate promiscuity of GroEL. *Curr Sci*.

- 2011;100(11):1646–53.
52. Weisberg RA. Transcription by Moonlight: Structural Basis of an Extraribosomal Activity of Ribosomal Protein S10. *Mol Cell*. 2008;32(6):747–8.
  53. Lu Z, Hunter T. Metabolic Kinases Moonlighting as Protein Kinases. *Trends Biochem Sci* [Internet]. 2018;43(4):301–10. Available from: <http://dx.doi.org/10.1016/j.tibs.2018.01.006>
  54. Jeffery CJ. Multifunctional proteins: Examples of gene sharing. *Ann Med*. 2003;35(1):28–35.
  55. Nájjar García A. Análisis de la virulencia de proteínas multifuncionales mediante bioinformática [Internet]. Universitat Oberta de Catalunya; 2019. Available from: <http://openaccess.uoc.edu/webapps/o2/handle/10609/98607>
  56. Kainulainen V, Korhonen T. Dancing to Another Tune—Adhesive Moonlighting Proteins in Bacteria. *Biology (Basel)* [Internet]. 2014 Mar 10;3(1):178–204. Available from: <http://www.mdpi.com/2079-7737/3/1/178>
  57. Min KW, Lee SH, Baek SJ. Moonlighting proteins in cancer. *Cancer Lett* [Internet]. 2016;370(1):108–16. Available from: <http://dx.doi.org/10.1016/j.canlet.2015.09.022>
  58. Venuti A, Paolini F, Nasir L, Corteggio A, Roperto S, Campo MS, et al. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer* [Internet]. 2011;10(1):140. Available from: <http://molecular-cancer.biomedcentral.com/articles/10.1186/1476-4598-10-140>
  59. Yoshida N, Oeda K, Watanabe E, Mikami T, Fukita Y, Nishimura K, et al. Chaperonin turned insect toxin. *Nature*. 2001;411(6833):44.
  60. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. *Nat Rev Genet* [Internet]. 2015 Jul 9;16(7):409–20. Available from: <http://www.nature.com/articles/nrg3950>
  61. 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(6):911–24.
  62. Cvekl A, Piatigorsky J. Lens development and crystallin gene expression: many roles for Pax-6. *BioEssays* [Internet]. 1996 Aug;18(8):621–30. Available from: <http://doi.wiley.com/10.1002/bies.950180805>
  63. Espinosa-Cantú A, Ascencio D, Barona-Gómez F, DeLuna A. Gene duplication and the evolution of moonlighting proteins. *Front Genet* [Internet]. 2015 Jul 7;6. Available from: <http://journal.frontiersin.org/Article/10.3389/fgene.2015.00227/abstract>
  64. Hernández S, Calvo A, Ferragut G, Franco L, Hermoso A, Amela I, et al. Can bioinformatics help in the identification of moonlighting proteins? *Biochem Soc Trans*. 2014;42(6):1692–7.
  65. Ozimek P, Kötter P, Veenhuis M, van der Klei IJ. Hansenula polymorpha and Saccharomyces cerevisiae Pex5p's recognize different, independent peroxisomal targeting signals in alcohol oxidase. *FEBS Lett* [Internet]. 2006 Jan 9;580(1):46–50. Available from: <http://doi.wiley.com/10.1016/j.febslet.2005.11.045>
  66. Backert S, Feller SM, Wessler S. Emerging roles of Abl family tyrosine kinases in microbial pathogenesis. *Trends Biochem Sci* [Internet]. 2008 Feb;33(2):80–90. Available from:

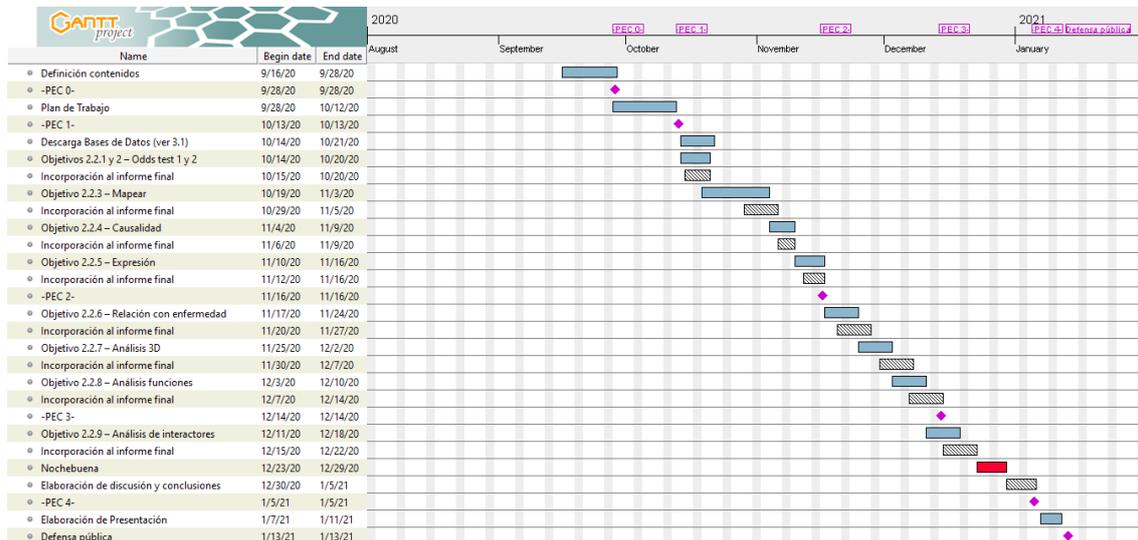
- <https://linkinghub.elsevier.com/retrieve/pii/S0968000407002927>
67. Feltcher ME, Sullivan JT, Braunstein M. Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiol* [Internet]. 2010 Oct;5(10):1581–97. Available from: <https://www.futuremedicine.com/doi/10.2217/fmb.10.112>
  68. 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* [Internet]. 2015;13(6):343–59. Available from: <http://dx.doi.org/10.1038/nrmicro3456>
  69. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(DATABASE ISS.):514–7.
  70. Cooper DN, Ball E V., Krawczak M. The human gene mutation database. *Nucleic Acids Res*. 1998;26(1):285–7.
  71. 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(D1):1118–23.
  72. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008;36(SUPPL. 1):901–6.
  73. 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(D1):D1074–82.
  74. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology David. *Nat Genet*. 2000;25(1):25–9.
  75. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–8.
  76. Stenson PD, Mort M, Ball E V., Chapman M, Evans K, Azevedo L, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet* [Internet]. 2020;139(10):1197–207. Available from: <https://doi.org/10.1007/s00439-020-02199-3>
  77. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, et al. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat Methods*. 2012;9(4):345–50.
  78. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. *Nucleic Acids Res*. 2009;37(SUPPL. 1):211–5.
  79. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47(D1):D351–60.
  80. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallogr Sect D Biol Crystallogr*. 2002;58(6 I):899–907.
  81. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, et al. Therapeutic target database 2020: Enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res*.

- 2020;48(D1):D1031–41.
82. Pomaznoy M, Ha B, Peters B. GONet: A tool for interactive Gene Ontology analysis. *BMC Bioinformatics*. 2018;19(1):1–8.
  83. Hassan H, Shanak S. GOTrapper: A tool to navigate through branches of gene ontology hierarchy. *BMC Bioinformatics*. 2019;20(1):1–6.
  84. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc*. 2017;12(2):255–78.
  85. Ding Z, Kihara D. Computational Methods for Predicting Protein-Protein Interactions Using Various Protein Features. *Curr Protoc protein Sci*. 2018;93(1):e62.
  86. Stewart A. *Basic Statistics and Epidemiology: A Practical Guide*: [Internet]. Health Education Research. Oxford: Radcliffe Medical Press; 2002. 151 p. Available from: <https://academic.oup.com/her/article-lookup/doi/10.1093/her/17.6.774>
  87. Pita-Fernández S, Vila-Alonso M, Carpena Montero J. Determinación de factores de riesgo. *Cad Aten Primaria*. 1997;4:75–8.
  88. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* [Internet]. 2009 Nov 15;25(22):3045–6. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp536>
  89. National Center for Biotechnology Information. MedGen [Internet] [Internet]. National Library of Medicine (US). 2004 [cited 2020 Nov 24]. Available from: <https://www.ncbi.nlm.nih.gov/medgen/>
  90. Stevenson M, Nunes T, Sanchez J, Thornton R, Reiczigel J, Robison-Cox J, et al. epiR: An R package for the analysis of epidemiological data. 2013.
  91. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* [Internet]. 1986 Mar 15;292(6522):746–50. Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj.292.6522.746>
  92. Chen S-Y, Takeuchi S, Urabe K, Hayashida S, Kido M, Tomoeda H, et al. Overexpression of phosphorylated-ATF2 and STAT3 in cutaneous angiosarcoma and pyogenic granuloma. *J Cutan Pathol* [Internet]. 2008 Aug;35(8):722–30. Available from: <http://doi.wiley.com/10.1111/j.1600-0560.2007.00887.x>
  93. von Elsner L, Hagemann S, Just I, Rohrbeck A. C3 exoenzyme impairs cell proliferation and apoptosis by altering the activity of transcription factors. *Naunyn Schmiedebergs Arch Pharmacol*. 2016;389(9):1021–31.
  94. Elam MB, Majumdar G, Mozhui K, Gerling IC, Vera SR, Fish-Trotter H, et al. Patients experiencing statin-induced myalgia exhibit a unique program of skeletal muscle gene expression following statin re-challenge. *Artero R*, editor. *PLoS One* [Internet]. 2017 Aug 3;12(8):e0181308. Available from: <https://dx.plos.org/10.1371/journal.pone.0181308>
  95. Taguchi Y-H. Tensor decomposition-based unsupervised feature extraction identifies candidate genes that induce post-traumatic stress disorder-mediated heart diseases. *BMC Med Genomics* [Internet]. 2017 Dec 21;10(S4):67. Available from: <https://bmcmmedgenomics.biomedcentral.com/articles/10.1186/s12920->

- 017-0302-1
96. Ding S. Methylation profile of the promoter CpG islands of 14 “drug-resistance” genes in hepatocellular carcinoma. *World J Gastroenterol* [Internet]. 2004;10(23):3433. Available from: <http://www.wjgnet.com/1007-9327/full/v10/i23/3433.htm>
  97. Masson C, Mena F, Pinon-Lataillade G, Frobert Y, Radicella JP, Angulo JF. Identification of KIN (KIN17), a human gene encoding a nuclear DNA-binding protein, as a novel component of the TP53-independent response to ionizing radiation. *Radiat Res* [Internet]. 2001;156(535):544. Available from: <https://pubmed.ncbi.nlm.nih.gov/11604067/>
  98. Pratheeshkumar P, Sheeja K, Kuttan G. Andrographolide induces apoptosis in B16F-10 melanoma cells by inhibiting NF- $\kappa$ B-mediated bcl-2 activation and modulating p53-induced caspase-3 gene expression. *Immunopharmacol Immunotoxicol* [Internet]. 2012 Feb 20;34(1):143–51. Available from: <http://www.tandfonline.com/doi/full/10.3109/08923973.2011.588233>
  99. Zdanov S, Toussaint O, Debacq-Chainiaux F. p53 and ATF-2 partly mediate the overexpression of COX-2 in H<sub>2</sub>O<sub>2</sub>-induced premature senescence of human fibroblasts. *Biogerontology* [Internet]. 2009 Jun 11;10(3):291–8. Available from: <http://link.springer.com/10.1007/s10522-008-9204-0>

## 6. Anexos

### Anexo 1. Cronograma de actividades



También se adjunta como archivo (TFM.gan)

Enlaces para descarga de archivos:

[TFM.gan](#)

[Lista de enfermedades humanas asociadas a proteínas moonlighting](#)

[Programa para crear diccionario con códigos GO asociados a cada proteína moonlighting](#)

[Anotación códigos GO](#)

[Resultados de GOTrapper para eliminar los GO específicos](#)

[Imágenes de códigos GO de los grupos de genes asociados a enfermedades humanas obtenidos con STRING \(G\\_enf\)](#)

[Carpeta con el proyecto de R donde se encuentra el .RMD donde se realizan los análisis en R \(lista de proteínas, análisis estadísticos, etc\).](#)

[Carpeta con imágenes GO de los grupos de enfermedades](#)

[Carpeta con imágenes GO de las proteínas comunes entre grupos](#)

[Docking molecular](#)